ASYMPTOTIC ANALYSIS OF NETWORK STRUCTURES:

DEGREE-DEGREE CORRELATIONS AND DIRECTED PATHS

Pim van der Hoorn

**ASYMPTOTIC ANALYSIS OF NETWORK
STRUCTURES: DEGREE-DEGREE CORRELATIONS
AND DIRECTED PATHS**

Pim van der Hoorn

Dissertation committee

Chairman & secretary: Prof. dr. P. M. G. Apers
University of Twente

Promotor: Prof. dr. R. J. Boucherie
University of Twente

Co-promotor: Dr. N. Litvak
University of Twente

Members: Prof. dr. J. L. Hurink
University of Twente
Prof. dr. ir. M. R. van Steen
University of Twente
Prof. dr. R. W. van der Hofstad
Technische Universiteit Eindhoven
Dr. M. Olvera-Cravioto
Columbia University
Dr. M. Boguñá
Universitat de Barcelona
Prof. dr. P. Boldi
Università degli Studi di Milano

CTIT Ph.D. Thesis Series No. 16-402
Center for Telematics and Information Technology
University of Twente
P.O. Box 217, 7500 AE, Enschede, The Netherlands

ISBN: 978-90-365-4179-4

ISSN: 1381-3617 (CTIT Ph.D. thesis Series No. 16-402)

DOI: 10.3990/1.9789036541794

<http://dx.doi.org/10.3990/1.9789036541794>

Cover design: Ilona de Jong en Elwin Levels

Latin text: Siebe van der Horst en Myrthe van Rijn.

This research was supported by EU-FET Open grant NADINE (288956).

Copyright © 2016, Pim van der Hoorn, Enschede, the Netherlands.

All rights reserved. No part of this publication may be reproduced without the prior written permission of the author.

**ASYMPTOTIC ANALYSIS OF NETWORK
STRUCTURES: DEGREE-DEGREE CORRELATIONS
AND DIRECTED PATHS**

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
Prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 7 oktober 2016 om 14:45 uur

door

Wilhelmus Lucas Franciscus van der Hoorn

geboren op 18 augustus 1985
te Leiderdorp, Nederland

Dit proefschrift is goedgekeurd door:
Prof. dr. R.J. Boucherie (promotor)
Dr. N. Litvak (co-promotor)

Opgedragen aan mijn neefje Wouter:
*Iter tibi prosperitate scientiaque obseratur
et aetas amore sapientiaque plena sit.*

Voorwoord

Geen enkele uitzonderlijke prestatie is het werk van slechts één persoon. En hoewel ik, met betrekking tot dit proefschrift, geen uitspraak zal doen over het eerste deel van dit statement, is het tweede deel absoluut waar. Ik wil dan ook graag iedereen bedanken die hieraan op een of andere manier heeft bijgedragen. Sommige mensen wil ik in het bijzonder noemen en daarvoor zal ik deze pagina's gebruiken.

Toen ik, iets meer dan vier jaar geleden, in de zomer van 2012 besloot om een PhD te doen, was ik niet helemaal zeker van mij zaak, gezien het feit dat ik al een jaar geen wiskunde meer bedreven had. Voor mijn interview in Enschede was ik dan ook erg zenuwachtig. Helemaal omdat het hier zelfs over een onderwerp ging waar ik nog nooit iets mee gedaan had gedurende mijn tijd als student. Richard, dank je dat je mij deze kans hebt gegeven, voor je vertrouwen in mij als onderzoeker en het opnemen van mij in je groep. Ik zal altijd met veel plezier terugdenken aan mijn tijd in Twente.

Voor mijn dagelijkse begeleiding kon ik terecht bij Nelly. Ik weet nog goed dat ze mij tijdens ons eerste overleg meteen vertelde dat ze hele hoge verwachtingen van me had. Ik kan alleen maar hopen dat ik daar ergens in de buurt van ben gekomen. Maar Nelly, jij bent mijn verwachtingen als begeleider meer dan overstegen. Het voelt ergens vreemd om dit in het Nederlands te schrijven. Ondanks het feit dat je deze taal goed beheerst communiceren wij namelijk altijd met elkaar in het Engels. De stewardess van de KLM begreep er in ieder geval niets van. Jij was altijd heel duidelijk in wat je van me verwachtte en had altijd een plan klaar, hoewel dit laatste vaak erg flexibel was. Je gaf me de vrijheid om mijn onderzoek in te richten en liet me vaak mijn gang gaan, waar je hulp bood als ik dat nodig had. Maar je was ook streng, direct en ondubbelzinnig op de momenten dat dit moest. Naast een meer dan uitstekende begeleider ben je een ook een hele goede vriendin van me geworden. Ik heb ontzettend genoten van de conferenties die wij samen hebben bezocht en alle avonturen die wij hebben beleefd. Ik hoop dan ook dat wij nog heel lang vrienden zullen blijven. Maar nu zal ik verder gaan, voordat je me weer vertelt dat ik als een meisje schrijf.

Gedurende mijn PhD heb ik de mogelijkheid gehad om twee onderzoeksstages in het buitenland te doen en daar ben ik heel dankbaar voor.

Mariana, thank you for inviting me to Columbia. It was my first time collaborating with someone other than my supervisor and I more than appreciated the opportunity you gave me. When writing this thesis, especially when dealing with some technical proofs, I realized how much I learned from you in only these

two weeks. I also want to thank you for being part of my defense committee and reading my thesis. I hope to continue our collaboration in the future and learn even more from you.

For my second research internship I was able to spend two months in Moscow. I want to thank my host Andrei Raigorodskii for inviting me and arranging everything. I also want to thank the staff of MIPT and Dima for their help with all my daily problems, such as doing the laundry. During these two months I was able to further develop myself as an independent researcher and I am very grateful for this experience. Egor and Luida thank you for finding the time in your busy schedule to discuss research. I think that the work we did is really nice and hope to continue our collaboration.

Remco, ik wil je graag bedanken voor je gastvrijheid gedurende verschillende workshops in Eindhoven, voor alle constructieve discussies en voor de samenwerking die we recentelijk zijn begonnen. Ik hoop nog vele jaren samen met jou het veld van random graphs en complexe netwerken te mogen verkennen.

I also want to thank my other committee members Johann, Maarten, Marián and Paolo, for taking the time to read my thesis and providing valuable feedback, as well as for participating in my defense. The field of complex network analysis is multi-disciplinary and I am very happy to have such a diverse and experienced committee.

Op de universiteit had ik het genoeg om deel uit te maken van een erg leuke groep.

Maartje, jij was altijd mijn partner in crime als het aankwam op het maken van stukjes voor alle andere promovendi van onze groep. Ik heb heel erg genoten van onze avonden gevuld met Meisjes met IJsjes, Ja Zuster Nee Zuster en de vraag waarom een krokodil zijn tong niet kan uitsteken. Daarnaast was je ook niet te beroerd om af en toe in Utrecht een paar pasjes te doen en dat kan ik als notoire Derrick-ganger erg waarderen.

Het grootste deel van mijn tijd heb ik mijn kamer mogen delen met twee heel erg gezellige kamergenoten. Michaela en Yantin, dank jullie wel voor deze leuke tijd. Hopelijk is jullie Nederlands nog steeds goed genoeg om dit te kunnen lezen.

Jasper, Kamiel en Ruben. Heel erg bedankt voor de avonden Magic the Gathering en de, respectievelijk, speeddate, voetbal en bierdrink support.

Natuurlijk wil ik ook alle ander leden van de SOR en DMMP groep heel erg bedanken voor de afgelopen vier jaar. Dankzij jullie heb ik me, ondanks dat ik uit Utrecht kwam, altijd thuis gevoeld in Twente.

Een goede basis is essentieel voor bijna alles en zeker het behalen van een PhD. In mijn geval bestaat deze basis voor het grootste deel uit vriendschappen die stuk voor stuk onmisbaar zijn.

Maarten, Ilona, Irene, Siebe, Myrthe, Dennis en Kirsten. Dank jullie voor alle Nieuwjaarsdinners, Bikkell weekenden, boshuis bezoeken, pretpark tripjes, ellenlange bordspelsessie, kaasfondues, Lente-uitjes in september, Sinterklaas vieringen met veel te lange gedichten en alle andere dingen die we samen gedaan hebben.

Maarten, helaas, jij ontkomt niet aan je eigen stukje. Eindegeheid maakt bijzonder, over alles moet je een grap kunnen maken en alles moet gevierd worden. Slechts een kleine greep van de openbaringen die zijn ontstaan tijdens onze vaste maandagavond aan de bar van de Olivier. Er zijn slechts weinig mensen die be-

grijpen wat je op maandag in een café doet. En zelfs zij kunnen niet begrijpen wat deze 8 jaar lange traditie echt betekent, laat staan waarom we er een gezamenlijke rekening voor hebben geopend. Ik zou het dan ook meer dan tekort doen door het hier in woorden proberen samen te vatten. Voor onze vriendschap geldt hetzelfde. Jij stond altijd en onvoorwaardelijk voor me klaar, met een goed advies, noodzakelijk relativerend commentaar of goede drank. Daarnaast ben je de enige andere persoon die gek genoeg is om op een zaterdag op en neer naar Épernay te rijden om even twaalf dozen champagne te scoren of naar Westvleteren voor twee kratten bier. Ook wil ik je nog bedanken voor je hulp asl paranimf. Het verdedigen van je proefschrift is een belangrijke gebeurtenis en ik ben ontzettend blij dat jij hier deel van uitmaakt.

Elwin, jouw droge humor, affiniteit met cafés, uitzonderlijke gave voor het vinden van foute films en onweerlegbare LAN party skills waren een uitstekende afleiding tijdens het produceren van al deze quatsch. Daarnaast was de voorkant van dit proefschrift niet zo een episch kunstwerk geworden zonder jouw kundige hulp.

Marcel, Ronald en Tristan. Ik heb heel erg genoten van onze NERD lunches, inclusief alle woordgrappen. Ook onze avonden gevuld met Battle Blobs en Mother ships of Trouble Markers waren altijd weer een groot feest. Marcel ook jij heel erg bedankt dat jij mijn paranimf wil zijn. Wij hebben het grootste deel van onze opleiding samen gedaan en ik ben dan ook heel blij dat je ook onderdeel van deze afsluiting bent.

Dan wil ik nog alle andere NERDS, Bart, Jerfey, Harry, Marco, Martijn en Vincent, bedanken voor de briljante NGL weekenden, saboteur avonden en epische dansmomenten in Utrecht.

Ik wil ook graag mijn ouders bedanken. Mam en Pap, jullie hebben me altijd vrijgelaten om datgene na te jagen wat ik wilde en mij daarin gesteund. Of het nu een oorbel was of meterslang haar. Dankzij jullie opvoeding heb ik geleerd om kritische te zijn op mijzelf en altijd het positieve in dingen te zien. Ook mij zusjes, Tessa en Marloes, en Levon wil ik bedanken voor hun steun gedurende de afgelopen periode.

Tot slot wil ik graag mijn vriendin bedanken. Manon, jij bent, denk ik, de enige ter wereld die met een jongen op date gaat die er voor uitkomt een masochistische teddybeer met een alcoholprobleem te bezitten. Ik beloof ervoor te zorgen dat je hier geen spijt van krijgt. Ik wil je bedanken voor je steun en voor het aanhoren van al mijn gezwam omtrent mijn eigen en jouw onderzoek. De afgelopen tijd met jouw was ongelooflijk fijn en ik kan niet wachten om te beginnen aan ons volgende avontuur, aan de andere kant van de grote plas.

Pim van der Hoorn
Utrecht, 2016

Contents

1	Introduction	1
1.1	Complex networks	2
1.1.1	Power-law degrees and scale-free networks	3
1.1.2	Distances in complex networks	5
1.1.3	Analysis using random graphs	6
1.2	Degree-degree correlations	7
1.2.1	Difference between undirected and directed networks	7
1.2.2	Influence on network properties and processes	8
1.2.3	Measuring degree-degree correlations in directed networks	9
1.2.4	Generating graphs with specific degree-degree correlations	10
1.3	Directed configuration model	11
1.3.1	Simple graphs and removed edges	12
1.3.2	Neutral mixing of degrees	13
1.3.3	Length of directed paths	14
1.4	Methodology	14
1.4.1	Convergence statements	14
1.4.2	Proof strategy and typical arguments	15
1.4.3	Algorithms and experiments	17
1.5	Outline of the thesis	18
I	Preliminaries	21
2	Notations and definitions	23
2.1	Probabilistic tools	23
2.1.1	Convergence of probability mass functions	23
2.1.2	Regularly-varying random variables	24
2.1.3	Scaling of regularly-varying random variables	26
2.1.4	Continuization	30
2.2	Graphs and degree sequences	32
2.2.1	A different characterization of graphs	32
2.2.2	Degrees and degree sequences	32
2.3	Random graphs and regularity assumptions	35

3	Generating degree sequences	39
3.1	Undirected graphs	39
3.2	Directed graphs	42
3.2.1	The directed IID algorithm	42
3.2.2	Asymptotic results for bi-degree sequences generated by the IID algorithm	43
3.2.3	IID algorithm with finite covariance of out- and in-degrees	48
3.2.4	Scaling results for out- and in-degrees IID algorithm	56
II	Degree-degree correlations in directed random graphs	59
4	Measures for degree-degree correlations in directed graphs	61
4.1	Introduction	62
4.2	Pearson's correlation coefficient	62
4.3	Spearman's rho and related rank-correlation measures	65
4.3.1	Definition of Spearman's rho	65
4.3.2	Spearman's rho with uniform tie resolution	66
4.3.3	Spearman's rho with average ranking	68
4.4	Kendall's tau	69
4.5	Measured degree-degree correlation in real-world networks	70
5	Convergence of correlation measures in directed random graphs	73
5.1	Introduction	74
5.2	Non-negativity of Pearson's correlation coefficient	75
5.3	Convergence of rank-correlation measures in random graphs	80
5.3.1	Rank-correlation measures for integer valued random variables	80
5.3.2	Main result	81
5.4	Road map for the proof of Theorem 5.3	83
5.5	Proof of the main result	85
5.5.1	Proof of Theorem 5.3 i)	86
5.5.2	Proof of Theorem 5.3 ii)	92
5.5.3	Proof of Theorem 5.3 iii)	95
5.6	Convergence of rank-correlation measures under weaker assumptions	97
6	Directed graphs with neutral mixing of degrees and given degree distributions	99
6.1	Introduction	100
6.2	Generating graphs with prescribed bi-degree sequence	100
6.2.1	Neutral mixing of degrees in configuration graphs	101
6.2.2	Proof of Proposition 6.2	103
6.3	Repeated configuration model	108
6.4	Erased configuration model	111
6.4.1	Altered empirical distributions and the average number of removed edges	111
6.4.2	Convergence of rank-correlations in the erased configuration model	112

6.5	Numerical evaluations of rank-correlation measures	114
6.5.1	Setup of the numerical experiments	114
6.5.2	Numerical results	115
7	Maximal disassortative undirected graphs, with given degree distribution	119
7.1	Introduction	120
7.2	Spearman in undirected graphs	120
7.3	The disassortative graph algorithm (DGA)	122
7.4	Joint degree distribution of graphs generated by the DGA	124
7.4.1	Partitioned representation of the DGA	124
7.4.2	Limiting joint degree distribution	127
7.4.3	Proof of Theorem 7.4	127
7.5	Properties of the Disassortative Graph Algorithm	134
7.5.1	Minimizing Spearman's rho	134
7.5.2	Simplicity of DGA graphs	136
7.6	A lower bound for Spearman's rho.	139
7.7	Spearman's rho and the tail of the degree distribution	141
7.8	Spearman's rho on maximal disassortative graphs.	145
7.8.1	Regularly-varying degree distribution	146
7.8.2	Poisson degree distribution	148
7.8.3	Important observations and insights	148
III	Structural properties of the directed configuration model	151
8	Scaling of the number of removed edges in the erased configuration model	153
8.1	Introduction	154
8.2	Upper bounds for the number of removed edges in the undirected model	155
8.2.1	The upper bounds $n^{\frac{4}{\gamma}-3}$ and n^{-1}	157
8.2.2	The upper bound $n^{\frac{1}{\gamma}-1}$	157
8.3	Upper bounds for the number of removed edges in the directed model	160
8.3.1	Transition to i.i.d. degrees	161
8.3.2	The first set of upper bounds	164
8.3.3	An upper bound that scales as $n^{\frac{1}{\gamma^*}-1}$	165
8.4	A better scaling for average number of removed edges	168
9	Scaling of rank-correlations measures in the erased configuration model	173
9.1	Introduction	174
9.2	Structural negative Out-In correlations in the erased configuration model	175
9.3	Empirical analysis of the scaling of ρ_+^- in the erased configuration model	178

9.3.1	Methodology of the analysis	178
9.3.2	Scaling terms for Spearman's rho	179
9.3.3	Numerical results	181
9.4	Scaling of the In-Out, Out-Out and In-In degree-degree correlation types	181
10	Distances in the directed configuration model	187
10.1	Introduction	188
10.2	Coupling the exploration of a graph with a branching process . . .	189
10.2.1	Regularity assumptions on the bi-degree sequence	189
10.2.2	Exploration of new nodes in a graph	190
10.2.3	Construction of the coupling	191
10.2.4	Coupling results	193
10.2.5	Road map for the proof of Theorem 10.1	194
10.2.6	Some results for delayed branching processes	195
10.2.7	Proving the coupling theorem	199
10.3	Distribution of the hopcount	210
10.3.1	The main result	210
10.3.2	Ideas and heuristic explanation of the proof	212
10.3.3	Proof of the main result	213
10.4	Numerical examples	225
10.4.1	Computing the hopcount distribution	225
10.4.2	Results for different bi-degree sequences	226
11	Concluding remarks	229
	Glossary	241
	Summary	243
	Samenvatting	245
	About the author	248

Chapter 1

Introduction

The world around us is filled with complex systems, whether we are traveling by car, checking the latest status updates on social-media or reading a book.

Every day roads are filled with cars that drive at different, even changing, speeds and that have to navigate through many intersections to reach their destination. When two cars are driving too close behind each other, a small decrease in the speed of the car in the front can cause the car behind to break slightly harder. This process then cascades through the network of roads and finally we find ourselves stuck in a traffic jam, miles away from the initial origin and potentially even on an entirely different road.

When we log in to our favorite social media platform, sophisticated algorithms are checking updates to find those that we might be interested in. Based on our preferences and social connections, updates from our friends and possibly the friends of our friends, we might even get a suggestion for a new friendship with someone we have never seen before.

While looking at a page full of writing, electric signals travel via the optic nerves to the cerebral cortex. Here an intricate system of connected neurons fire, creating a cascading effect that expands to different regions of the brain and enables us to identify the words on this page and assign meaning to them.

Understanding such large and complicated systems is vital for determining, how to prevent traffic jams, which people, or groups, will form social relations or what impact a degenerative disease in the brain can have on our cognitive functions. For this, one must be able to analyze the structure of these systems as well as the influence this structure has on the processes associated with them.

This thesis is concerned with the analysis of large and complex systems, from the perspective of networks, by combining the mathematical theory of probability and graphs with large-scale numerical experiments. We derive measures for structural properties of networks and analyze their behavior as we let the networks grow in size. This approach leads to mathematical expressions that can help to quantify the structure of networks of finite size. In addition we design models that generate networks that have specific structures and characteristics. Such models are an essential tool for analyzing the impact of these structures and characteristics on processes in real-world networks.

1.1 Complex networks

Many systems can be represented as a network, a collection of nodes and relations between them called edges. These edges can be undirected, representing symmetric relations, or have a direction, in which case the relations are asymmetric. Note that symmetric relations between nodes can also exist in a directed network, in which case these are represented by two edges pointing in opposite directions. The mathematical terminology for a network is a graph and in the literature these two are used interchangeably. We use the term graph for a general collections of nodes and edges, while the term network refers to a real-world system represented as a graph. In general, most networks are *simple* graphs. They have no self-loops, edges that point to the same node from which they originate, and there is no more than one edge in a specific direction between two nodes.

Figure 1.1: Example of a simple undirected and directed graph.

Examples of undirected networks include the Internet, where nodes are routers and the edges are the physical connections between the routers, or social networks consisting of people and friendships between them. For directed networks one can think of Wikipedia (www.wikipedia.org), where nodes are the Wikipedia pages and the hyperlinks between the pages are the directed edges, or a citation network, which has papers as nodes and there is a link from paper *A* to paper *B*, if *A* cites *B*.

In the literature, the term *complex networks* refers to networks that have two distinctive properties. First of all, they have a large number of nodes. What "large" means depends on who you ask. Some consider the representation of the connectome (the human brain network), using *regions of interest*, ranging from 500 to 4000 nodes, see [45], large. Others start from 10^6 nodes and go up into the billions, cf. [22]. Still it should be clear that these networks contain more nodes than we would feel comfortable drawing on paper ourselves.

The second property of complex networks relates to their structure, the collection of edges between the nodes. This structure is *complex*. There are no general rules describing how nodes connect to one another and a graphical representation of the network structure gives almost no insights into its structural properties. Often, the process by which the connections in complex networks are created is unknown, and one of the main driving forces behind the development of models for networks is to understand how local rules can create certain global structures. For an example of a complex network see Figure 1.2.

The important consequence of the large size and complex rules that govern the forming of edges in complex networks is that it prohibits us from giving a global description of the network. Therefore we have to turn to local properties, for analyzing and characterizing networks, such as the number of nodes or the

Figure 1.2: High-resolution connectome map, showing the average anatomical connectivity network of the human brain across 500 healthy subjects from the Human Connectome Project. Courtesy of Marcel de Reus.

distribution of the number of links per node, see Section 1.1.1.

The list of interesting structural properties of networks is long. It includes, among others, (i) clustering coefficient, the proportion of triangles, (ii) communities, the tendency of nodes to form groups that are more connected with each other than with other nodes, and (iii) connected components, sets of nodes that can reach each other by following a path of edges and that cannot reach nodes outside this set. We refer the interested reader to [1, 61] for an overview of the field of complex networks. A mathematical treatment of models for studying complex networks can be found in [79].

There are two aspects of complex networks that are important for this thesis. These are scale-free behavior and distances, and we briefly describe both.

1.1.1 Power-law degrees and scale-free networks

A feature that many complex networks seem to share is related to the degrees of nodes, the number of connections that nodes have. Let us denote by D_i the number of links that node i has. Then, for an undirected network with n nodes, we have a sequence

$$\mathbf{D}_n = \{D_1, D_2, \dots, D_n\}$$

consisting of the degrees of each node. This is called the *degree sequence* of the network. In directed networks we make the distinction between the number of outgoing edges, *out-degree*, and incoming edges, *in-degree*. If we denote the out-degree of node i by D_i^+ and the in-degree by D_i^- , then a directed network gives rise to a joint sequence

$$(\mathbf{D}_n^+, \mathbf{D}_n^-) = \{(D_1^+, D_1^-), (D_2^+, D_2^-), \dots, (D_n^+, D_n^-)\}$$

Figure 1.3: Plot of the distribution of the in-degrees in the English Wikipedia network, which consists of pages in the English Wikipedia and the hyperlinks between these pages. The values for p_k and P_k are plotted on the y -axis against the degree k on the x -axis.

which we call the *bi-degree sequence*.

Often, many nodes will have a small number of links while a few nodes have very large degree. As an example, think of the Twitter network, where nodes are users and there is a directed edge from user A to user B if B follows A . Many users will have a small (≤ 100) number of followers. On the other hand, there are some users whose number of followers exceed 8×10^7 , such as Katy Perry or Justin Bieber¹.

This skewed distribution of the degree of nodes can be observed if we plot the fraction of nodes with a given degree, as a function of the degree. For this let p_k denote the proportion of nodes with degree k . Then a plot of p_k for complex networks, on log-log scale, typically looks as in Figure 1.3a. Another way to see this behavior of the degrees is to consider the tail of the cumulative distribution,

$$P_k = \sum_{t>k} p_t,$$

which we plotted on a log-log scale in Figure 1.3b.

From Figure 1.3a we observe that many pages in the English Wikipedia have no more than 100 hyperlinks pointing to them. However, from Figure 1.3b we see that a fraction of Wikipedia pages of the order 10^{-6} have more than 10^5 incoming hyperlinks from other pages. In addition, we observe that both p_k and P_k seem to decrease linearly, on a log-log scale. These observations show that there is a high variability in the degrees of nodes in the network, which is a feature that most complex networks share.

A well established methodology to describe this behavior is to write

$$p_k \approx C_1 k^{-(\gamma+1)}, \quad (1.1)$$

for some constant $C_1 < \infty$ and $\gamma > 0$. In that case we have

$$\log(p_k) \approx \log(C_1) - (\gamma + 1) \log(k),$$

¹As reported by Twitter Counter (twittercounter.com) on July 2nd, 2016

which describes a line with slope $-(\gamma + 1)$ on log-log scale. When p_k behaves as (1.1) then for P_k we have

$$P_k \approx C_2 k^{-\gamma}, \quad (1.2)$$

for some other constant $C_2 < \infty$. Hence, a plot of P_k on log-log scale shows a linear decrease with slope $-\gamma$, see for example Figure 1.3b. In the literature, this type of distribution is referred to as a *power law* and networks with such p_k , or equivalently P_k , are called *scale free*.

Distributions that model high variability are called heavy-tailed distributions. These are distributions for which the tail is not exponentially bounded, in contrast to, for instance, a Poisson distribution. This large class of distributions includes for instance subexponential distributions. In this thesis we use *regularly-varying distributions*, which are heavy-tailed distributions, to model degree distributions that behave as in (1.2). We give a mathematical description of these distributions in Section 2.1.2.

Determining whether a network has a power-law degree distribution and, if so, finding the exponent γ , is a non-trivial task which involves more than just looking at the plot of p_k or P_k . It is also important to note that for most networks the degree density p_k does not obey a power law for all k . Often this power law behavior is present for all $k > k_*$, for some threshold k_* and in some cases the tail of degree-distribution has an exponential cut-off at the end, see for instance Figure 1.3b. A widely used methodology to test for power-law behavior is introduced in [30]. Here, first the threshold value k_* and exponent γ which fit the data the best are determined, using maximum likelihood ratios. Then a goodness-of-fit test is applied to see if the data can be seen as being drawn from the determined power-law. For a full mathematical treatment of the statistics of heavy-tailed distributions and related phenomena we refer to the book by Resnick [71].

Most of the literature considers the probability mass function p_k and refers to $\gamma + 1$ as the exponent of the degree distribution. We, however, use the cumulative tail and hence we use the term *exponent of the distribution* to refer to γ . This exponent determines which moments of the distribution exists, since

$$\sum_{k=1}^{\infty} k^p k^{-(\gamma+1)} = \sum_{k=1}^{\infty} k^{p-\gamma-1}$$

is only finite when $\gamma > p$. In particular, the literature on scale-free complex networks often distinguishes between the cases $\gamma < 1$, where the degrees have infinite mean, $1 < \gamma < 2$, in which case the degrees have finite mean but infinite variance, and $\gamma > 2$, where both the mean and variance are finite (for certain properties of networks, higher moments might be required). Actually, many real-world networks have been reported to be scale-free with exponent $\gamma \in (1, 3)$, see for instance [1, Table II] and [61, Table II].

1.1.2 Distances in complex networks

Distances are an important structural property related to the connectivity of networks. Given an undirected graph, the distance between two connected nodes, also called the hopcount, is the smallest number of edges that form a path between

the nodes. When two nodes are not connected the distance is set to be ∞ . When the graph is directed the distance from node i to node j is defined as the number of edges in the smallest directed path, starting in i and ending in j . Again if there is no directed path from i to j the distance is ∞ . In this thesis we are interested in the *hopcount of a graph*, which is defined as the distance between two nodes selected uniformly at random.

In many complex networks the path lengths between nodes are small with respect to the size of the network, see for instance [1, Table II] and [61, Table II]. This property is referred to as the small-world phenomenon, after the model in [92] which showed how to create short paths in large networks. The model places nodes on a circle and connects nodes to their closest neighbors. Then a fraction of edges is selected, uniformly at random, and these are rewired to create connections between nodes that are at opposite ends of the circle. This creates a few long-range connections between certain nodes and hence introduces short paths in the graph which use these shortcuts. This model was adapted in [63] by replacing the rewiring step with adding a small fraction of shortcuts between randomly chosen nodes. A mathematical analysis of small-world models was done in [7] where it is shown that distances scale logarithmically with the graph size.

In a recent study of the Facebook network (facebook.com) [5] it was shown that the average distance between a randomly select pair of nodes is 4.74. Since this network consists of more than 7×10^6 nodes this result strongly displays the small-world phenomenon in complex networks.

1.1.3 Analysis using random graphs

Given a network one can compute a large variety of quantities, although, depending on the size of the network, computations might take a long time. The result, however, does not give any insight into how the structure of the network was formed, nor can you predict what might happen in the future, as the network grows and its structure changes. To approach such questions much research is dedicated to designing models that generate graphs that mimic certain aspects of the structure of networks. Often such models are probabilistic in nature, since no deterministic rules are known for the creation of the connections in a network. Therefore, we call the graphs they produce *random graphs* and refer to the models as *random graph models*. We already discussed an example of such a model in the previous section, the small-world model.

Random graph models are useful for analyzing complex networks in many ways. Firstly, they can be analyzed mathematically to gain insights into the driving forces behind certain structures in complex networks. In addition, they can be used to generate large ensembles of graphs that grow in size, which allows for the analysis of the size dependence of properties of graphs and establish limit results that can be used to approximate the behavior on finite-size networks.

We remark that the term random graph is often used in the literature to refer to the Erdős-Rényi random graph [42], which is seen as the start of the research field of random graphs. This model takes n nodes and then connects any pair of nodes, independently, with a fixed probability p , which can be a function of n , for instance $p = \lambda/n$, for some $\lambda > 0$. However, such random graphs have a Poisson degree distribution with parameter λ and hence are not scale free.

A well-known model for generating graphs with a given degree distribution, which can be scale free, is the configuration model. This is the main model of interest in this thesis and we discuss it in Section 1.3 and later in more detail in Chapter 6.

Another example of a random graph model that is scale free is the so-called *preferential attachment model*. This model was considered in [6] to understand the emergence of the scale-free structure of networks and has many different forms. The general idea behind the preferential attachment model is to start with some initial graph and then add nodes one by one, to simulate a growing network. The term *preferential attachment* comes from the fact that when a new node is added, it will connect to a certain number of nodes, already present in the graph, based on some local preference rules. For instance, in the classical version, the probability that a new node connects to node i is proportional to the degree of i . It can be shown that the classical model has a power-law with exponent $\gamma = 2$ [24], although there are currently many extensions that have different exponents. We refer the reader to [11] for a mathematical analysis of different preferential attachment models.

1.2 Degree-degree correlations

Although many networks are scale-free and have small distances, their structure can be quite different. For instance, in some networks, nodes with large degree will often be connected to other nodes with similar large degree, while in other networks, nodes with large degree will actually have a preference to connect to nodes with very small degree. This lead to the introduction of the structural property of networks that describes the correlation between the degrees of connected nodes, which is called degree-degree correlation, or network assortativity [59]. A network is said to have *assortative mixing* (of degrees) if nodes of a certain degree are mostly connected to nodes of similar degree, e.g. nodes of large degree connect to nodes of large degree. When, on the other hand, nodes of large degree are connected to nodes of small degree, the network is said to have *disassortative mixing* (of degrees). We say that a network has *neutral mixing* (of degrees) if nodes have no preference, in terms of degrees, to which nodes they connect.

Currently, degree-degree correlations are part of the standard set of properties used to characterize the structure of networks. See [64] for a survey of the work on network assortativity. Interestingly, cf. [61, Table II], many networks involving human interaction, such as co-authorship and collaboration networks, seem to have assortative mixing, while technological and biological networks, for instance the Internet or the network of protein interactions, exhibit disassortative mixing.

1.2.1 Difference between undirected and directed networks

For undirected networks, the degree-degree correlation is defined as the correlation between the degrees on both sides of a randomly selected edge. When the network is directed, we can consider the correlation between any combination of out- and in-degree for the source and target node of the edge. This gives rise to four different degree-degree correlations types, which we refer to as Out-In,

Figure 1.4: The four directed degree-degree correlation types

In-Out, Out-Out and In-In. See Figure 1.4 for an illustration of the four different types. With this notation, for example, the Out-Out degree-degree correlation is the correlation between the out-degrees of the source and target of a randomly selected edge.

1.2.2 Influence on network properties and processes

One area where the effect of degree-degree correlations has been extensively studied deals with the dynamics of diseases on networks, cf. [8, 16, 18, 19]. For instance, disassortative networks are easier to immunize and a disease takes longer to spread in assortative networks [32].

In the field of neuroscience degree-degree correlations are studied in the context of information spread in the brain. For instance, in [73], it is shown that assortative brain networks are better suited for signal processing, while assortative neural networks are shown to be more robust to random noise in [33].

Another case where the assortativity of a network plays an important role is in the robustness of networks under attack, where either edges or nodes are removed. It turns out that assortative networks are more resilient to such attacks than disassortative networks [60, 89]. On the other hand, when different networks interact assortativity actually decreases the robustness of the whole system [98].

There are many more properties and processes on complex networks that are influenced by degree-degree correlations. Therefore it is important to have a sound methodology and theoretical framework for analyzing them. In this thesis we contribute to this goal by defining proper measures for degree-degree correlations in directed networks and proving that they are consistent under very general assumptions. In addition we establish a null-model and use experiments to analyze the finite-size effects, and establish the scaling of the fluctuation of the measures in this model.

1.2.3 Measuring degree-degree correlations in directed networks

A measure for degree-degree correlations was first given for undirected networks in [59], which corresponds to Pearson's correlation coefficient on the joint vector of degrees at the both ends of a random edge in the network, see Chapter 4. This measure assigns to a network a real number in between -1 and 1 , that classifies the assortativity of the network. Here, a (negative)positive value means that the network has (dis)assortative mixing. A similar definition for directed networks was introduced in [60] and later adopted for analysis of directed complex networks in [44] and [69]. Currently these measures are still the default way to compute degree-degree correlations in networks.

However, it turns out that the Pearson's correlation coefficient does not behave properly on scale-free graphs, which are the main objects of interest in the study of complex networks.

A first indication for this was [38] where it was shown that Pearson's correlation coefficient converges to zero on strongly correlated trees, as the network grows in size. For undirected networks of which the degrees follow a regularly-varying distribution with exponent $1 < \gamma < 3$, which includes the finite second moment case, it was shown in [53, 82] that Pearson's correlation coefficient scales with the network size, converging to a non-negative number in the infinite network size limit. In [56] upper and lower bounds are derived for the size dependence of Pearson's correlation coefficient, using a rewiring algorithm that makes a network more (dis)assortative [95]. Under the assumption that the distribution of the degrees satisfies a power law with infinite second moment, these bounds vanish as the size of the graph increases. The results in [56] are extended in [97], where a asymptotic lower bound for Pearson's correlation coefficient is obtained for scale-free networks. In particular for $1 < \gamma \leq 3$, it is shown that the limit of this measure is zero.

In light of this evidence against Pearson's correlation coefficients, alternative measures for assortativity in undirected networks were introduced in [53, 70, 82]. These measures, referred to as rank-correlation measures, are related to the statistical correlation measures Spearman's rho [75] and Kendall's tau [49].

In Part II of this thesis we define Pearson's correlation coefficient, and introduce rank-correlation measures, for degree-degree correlations in directed networks. We analyze their asymptotic behavior as the size of the network tends to infinity. The results are different for the four directed degree-degree correlation types. Although, for all four types, Pearson's correlation coefficient converges to a non-negative number when the out- and in-degrees follow regularly-varying distributions, the conditions on the degree distributions differ per type. For instance, for the Out-In degree-degree correlation type the result holds whenever the third moment of both the out- and in-degree distribution is infinite, while for the In-Out degree-degree correlation type this only holds when the distributions have infinite second moment. In contrast, we prove that rank-correlation measures are consistent estimators for the correlations between the degrees of connected nodes in directed networks, for all four degree-degree correlation types.

1.2.4 Generating graphs with specific degree-degree correlations

An important methodology used to analyze degree-degree correlations of a specific network, is to measure it and compare the outcome to those on similar graphs with neutral mixing.

One approach for this is to sample from graphs with the same degree sequence as the given network, but with neutral mixing. A widely accepted methodology for such sampling is through the local rewiring model [48, 54], which takes the original network and randomly swaps edges until a randomized version is attained. The disadvantage of these methods is that no theoretical results on the mixing times are given and hence we have no performance guarantees. There are sequential algorithms for sampling graphs with prescribed degree sequence, for which the performance is known, see [15, 35] for undirected and [50] for directed graphs. However the complexity of these models is $O(nL)$, where L denotes the number of edges, which make them less suited for the analysis of large networks.

In this thesis we consider the well-known configuration model [23, 58, 62] for generating graphs with neutral mixing. This model has the advantage that it is simple to implement and has been widely studied in the literature. The configuration model is described in more detail in Section 1.3 and Chapter 6. In this chapter we rigorously prove the intuitive result, that the degrees on both ends of a randomly sampled edge in directed graphs generated by the configuration model are asymptotically independent, and hence that the rank-correlation measures converge to zero. Therefore, the directed configuration model can be used as a null model for the analysis of degree-degree correlations in directed networks, when using rank-correlation measures.

If we are interested in analyzing the general impact of assortativity on other network properties and processes, we could create graphs with specific degree-degree structures and analyze their structural properties, as well as the outcome of processes on them. The joint degree structure of a graph is determined by the joint degree matrix J , where an entry $J_{k\ell}$ denotes the number of edges between nodes of degrees k and ℓ . Note that this matrix also determines the degree distribution p_k . The problem of generating graphs with a given joint degree structure has been addressed in [60]. Here degrees are sampled from the degree distribution, corresponding to the joint degree matrix, and nodes are connected at random. Then a rewiring approach is used to generate a graph with the given joint degree structure. These graphs will however, in general, not be simple.

Recently, in [9] and [76], algorithms are introduced for constructing and sampling *simple* graphs with a given joint degree matrix J . Unfortunately the algorithms apply to graphs of fixed size and no asymptotic results are yet known. An algorithm for generating random graphs whose joint degree distribution converges to a given limiting distribution is given in [37] and [47] under the assumption that the degrees are uniformly bounded in the size of the graph. This is however not a realistic setting since the maximal degree of complex networks with power-law degrees scales with the size of the network.

Another way to analyze the impact of degree-degree correlations is to create graphs that have an extreme degree-degree correlation structure, either maximally assortative or disassortative mixing, and analyze structural properties of,

and processes on, such graphs. For this purpose, one can adopt the rewiring algorithm, see [48, 54]. These papers propose to start from an initial graph, usually with neutral mixing, and in each step two edges are sampled and switched based on some rule, in order to obtain a maximally (dis)assortative graph. In [95, 96] this algorithm is used to analyze several properties of networks, with these maximal correlated structures. For instance, they conclude that the average path length increases when the network becomes either more disassortative or more assortative. In contrast, the clustering coefficient is only larger for networks with assortative mixing, compared to neutral mixing.

One of the problems with the current analysis of graphs with extreme degree-degree correlation structure is the use of Pearson's correlation coefficient as a measure for assortativity. In Chapter 7, where we introduce a model for generating undirected graphs with prescribed degree sequence, for which the Spearman's rho rank-correlation measure is minimal. The algorithm gives insights into the joint degree structure of maximally disassortative graphs and we use this to derive the limiting joint distribution of the degrees on both sides of a random edge.

1.3 Directed configuration model

The undirected configuration model [23, 58, 94] takes a degree sequence and creates an undirected graph with this specific degree sequence. This is done by assigning nodes a number of half-edges, or stubs, according to their degree. Then two stubs are selected, uniformly at random amongst the available stubs, and paired to form an edge. This procedure continues until all stubs have been paired.

For the analysis of complex networks it is important to be able to generate graphs such that the empirical degree distribution converges to a prescribed limit distribution. This can be accomplished by applying the configuration model to a degree sequence where the degrees are sampled in an i.i.d. fashion from the given distribution. We refer the reader to [79] for a thorough treatment of the undirected configuration model.

In this thesis we consider the directed version of the configuration model for the purpose of generating simple graphs with prescribed limiting out- and in-degree distributions, as described and analyzed in [28].

The extension of the undirected configuration model to the directed case for degree sequences is straightforward. Given a bi-degree sequence, we assign to each node a number of outbound and inbound stubs, according to its out- and in-degree, respectively. Then we select an outbound and inbound stub, uniformly at random amongst the available number of outbound and inbound stubs, respectively, and pair these to form a directed edge. Here the edge points from the node, to which the selected outbound stub belongs, to the node whose inbound stub was selected.

The difficulty with the directed version arises when we want to use the directed configuration model to generate graphs such that the empirical degree distributions converge to some desired limits. Since for a bi-degree sequence the sum of the out-degrees should equal the sum of the in-degrees, one cannot sample these independently. In [28] an algorithm for generating bi-degree sequences

is given, where the degrees are still close to being i.i.d. samples from the given distributions. We consider this algorithm in a slightly more general setting in Chapter 3. For bi-degree sequences generated with this algorithm, by sampling from prescribed distributions, the empirical out- and in-degree distributions in the directed configuration model converges to these distributions as the graph grows in size.

1.3.1 Simple graphs and removed edges

Since most networks are simple it is important that random graph models used for the analysis of networks also give simple graphs. The configuration model will unfortunately, in general, give *multi-graphs* (graphs that are not simple). This happens, for instance, when we pair the first stub of node i to node j and then pair the second stub of i again to a stub of node j , which has a positive probability if the degree of both nodes is larger than one. In order to generate simple graphs there are two refined versions of the configuration model.

The first model simply repeats the pairing procedure until a simple graph is produced. To avoid an infinite loop, one can first check whether there exists a simple graph for the given bi-degree sequence, or include a loop counter and stop after a certain number of attempts. This model is called the *repeated configuration model*. The repeated model works well when the out- and in-degree distributions both have finite variance as well as finite covariance, in which case the probability of generating a simple graph converges to a positive number as the number of nodes increases, cf. [28]. However, as the variability of the degrees grows, the probability of generating a simple graph decreases. Therefore, in practice, numerical experiments with the repeated configuration model might take a long time.

When either the out- or in-degree distribution has infinite variance however, the probability of generating a simple graph converges to zero with increasing number of nodes and hence the repeated configuration model does not work anymore. In this case a simple graph is created by removing self-loops and replacing multiple edges, in the same direction, between two nodes by one such edge. This procedure effectively removes edges from the graph and hence this model is called the *erased configuration model*.

Although this model changes the bi-degree sequence, it is shown in [28] that the proportion of erased edges converges in distribution to zero. Consequently, the average number of erased edges converges to zero and as a result, when the bi-degree sequence is generated using their algorithm, the empirical out- and in-degree distribution still converge to the distributions from which the degrees are sampled.

There are many results for the configuration model when the out- and in-degree distribution have finite variance. In this case the number of self-loops and multiple edges converge to Poisson random variables, see [79, Proposition 7.13] for the undirected case and [28] for the directed case. Moreover, in [3] a central limit theorem for the self-loops is proven for both the undirected and directed model. However, until now neither the speed at which the proportion of erased edges goes to zero nor the dependence of the number of erased edges on the size of the graph and the degree distributions has been known. In this thesis we provide

several results on the scaling of the number of erased edges and argue, based on these results, that even when the degree distributions have finite variance, the erased model should be preferred over the repeated model.

1.3.2 Neutral mixing of degrees

The configuration model plays an important role in analyzing degree-degree correlations in complex networks.

When the degree distribution has finite variance it is shown in [79] that the repeated configuration model generates graphs that are sampled uniformly at random amongst all graphs with the given degree sequence. Now consider an undirected network for which we have measured the degree-degree correlations. Then we can ask whether the measured value is typical for the specific structure of the network. To test this we can take the degree sequence of the network, use the configuration model to generate graphs and measure the degree-degree correlation on each of these. We can then take an average and compare this to the measured value on the original network. This will then give an indication how much the degree-degree correlation in the network differs from a randomly sampled graph with the same degree sequence.

A different application of the configuration model is for analyzing the impact of degree-degree correlations on other network properties and processes. In [82] it is proved that the undirected configuration model has asymptotic neutral mixing. Suppose now that we have a network on which we run a certain process, for instance the spread of a disease. Then, if we want to understand to which extent the outcome of this process is determined by the assortativity of the network, we can generate graphs with the configuration model, using the degree distribution of the network as input, and run the process on these graphs. Then we can compare the average of all outcomes with those for the original network to see the significance of the impact of the degree-degree correlations.

In practice, since the degree distributions of many complex networks fall into the infinite variance regime, the erased configuration model is used to generate graphs with neutral mixing. However, for this model, it is observed [26, 55, 66] that the measured value of the degree-degree correlation in such graphs is always negative. As a result an alteration of the configuration model is introduced in [26] for generating graphs with neutral mixing. The problem here is that this model imposes that the maximum degree is no greater than the square root of the graph size, which does not reflect the situation in many real-world networks.

We address the problem of structural negative correlations in the directed erased configuration model in Chapter 9. Moreover, we use experiments to find the size dependence of the fluctuations of the measured degree-degree correlations around their averages in the erased model for out- and in-degree distributions with both infinite and finite variance. This last research is an important step towards proving central limit theorems for measured degree-degree correlations, which is needed to do a proper analysis of the significance of measured values in complex networks.

1.3.3 Length of directed paths

In a collection of three papers [78, 80, 81] the hopcount of graphs of size n , constructed by the undirected configuration model with an i.i.d. degree sequence, is analyzed. When the degree distribution has finite variance the hopcount grows as $\log n$, [80], while this is $\log \log n$ when the degrees have finite mean and infinite variance [81]. When the degree distribution has infinite mean, the hopcount is either 2 or 3 [78]. These results show that the distances are in general very small compared to the graph size and become shorter as the variability of the degrees increases, due to shortcuts created by nodes with large degrees.

In this thesis, we provide an analysis of the distance between two randomly chosen nodes in the directed configuration model, under the assumption that the covariance between in- and out-degree is finite. This dependence between the out- and in-degree is an important difference between the undirected and directed case and plays a crucial role in the behavior of the distance between nodes. We show that the length of the shortest directed path between two nodes grows logarithmically in the number of nodes, which, unlike in the undirected case, can occur even when the variance of the degrees is infinite.

1.4 Methodology

In this thesis we use a general framework for the analysis of asymptotic properties of random graphs. This is, for instance, reflected in our notations for graphs and the very general regularity assumptions we impose on them, see Chapter 2. Another ingredient of this framework is the use of a uniform format for the statements of our results.

Here we describe the format of our statements and discuss, in detail, some proof techniques used in this thesis.

1.4.1 Convergence statements

Suppose we have random variables $\{X_n\}_{n \geq 1}$ and X such that X_n converges in probability to X , as $n \rightarrow \infty$. Then, for every $\delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \delta) = 0. \quad (1.3)$$

However there are some improvements to be made with respect to this statement. First, (1.3) does not state at which rate the probability converges to zero. Moreover, since (1.3) holds for each fixed $\delta > 0$ we do not know how the random variable $|X_n - X|$ scales with n . Therefore, in this thesis we are interested in statements of the form

$$\mathbb{P}(|X_n - X| > a_n) \leq b_n, \quad (1.4)$$

for sequences a_n and b_n that converge to zero as $n \rightarrow \infty$.

We refer to a_n as the scaling of $|X_n - X|$ and b_n as the convergence rate of $|X_n - X| \xrightarrow{\mathbb{P}} 0$. In particular we consider sequences that scale as inverse powers of n , that is $a_n = O(n^{-\delta})$ and $b_n = O(n^{-\varepsilon})$ as $n \rightarrow \infty$, for some $\delta, \varepsilon > 0$, where

$f(x) = O(g(x))$ as $x \rightarrow a$ means that

$$\limsup_{x \rightarrow a} \frac{f(x)}{g(x)} < \infty.$$

In many cases we use some assumptions on the random variables X_n . These assumptions will often be stated in terms of an event A_n and the condition that $\mathbb{P}(A_n) \rightarrow 1$, as $n \rightarrow \infty$. In this case, the probability of the complement of this event, A_n^c , will be part of the rate of convergence. Moreover, we try to be as general as possible with respect to the range of δ and ε for which the converges holds. Therefore our statements will often be of the following form:

For every $0 < \delta < K_1$, $0 < \varepsilon < K_2$ and $K > 0$

$$\mathbb{P}(|X_n - X| > Kn^{-\delta}) \leq O(n^{-\varepsilon} + \mathbb{P}(A_n^c)), \quad (1.5)$$

as $n \rightarrow \infty$.

We remark that we sometimes replace (1.5) with the equivalent statement

$$\mathbb{P}(|X_n - X| \leq Kn^{-\delta}) \geq 1 - O(n^{-\varepsilon} + \mathbb{P}(A_n^c)).$$

In addition, we emphasize that the scaling terms we derive in our results (δ and ε in the statement above) might not always be the best achievable.

Sometimes we might not be able to prove the rate of convergence and we can only prove that $|X_n - X| \xrightarrow{\mathbb{P}} 0$, conditioned on the event A_n . In this case our results are often stated as follows:

There exists a $\delta > 0$ such that, on the event A_n

$$|X_n - X| = O(n^{-\delta}),$$

as $n \rightarrow \infty$.

1.4.2 Proof strategy and typical arguments

In this thesis, most of the time, we use a uniform set of proof techniques. Here we give an example of the two most often used techniques.

Conditioning on the right event

Suppose that we have two real numbers $x, y > 0$ which satisfy $x < y$. In addition, let $\{X_n\}_{n \geq 1}$ and $\{Y_n\}_{n \geq 1}$ be random variables and let A_n denote an event which satisfies $\mathbb{P}(A_n) \rightarrow 1$ and on which we have that $|X_n - x| \leq n^{-\varepsilon}$ and $|Y_n - y| \leq n^{-\varepsilon}$. Then, if A_n^c denotes the complement, and $\mathbb{1}_{A_n}$ is the indicator of the event A_n , we derive

$$\mathbb{P}(X_n > Y_n) = \mathbb{P}((X_n - x) - (Y_n - y) > y - x) \quad (1.6)$$

$$\leq \mathbb{P}((X_n - x) - (Y_n - y) > y - x, A_n) + \mathbb{P}(A_n^c). \quad (1.7)$$

We now bound the first probability as follows:

$$\mathbb{P}\left(|X_n - x| > \frac{y - x}{2}, A_n\right) + \mathbb{P}\left(|Y_n - y| > \frac{y - x}{2}, A_n\right) \quad (1.8)$$

$$= \frac{2(\mathbb{E}[|X_n - x|\mathbb{1}_{A_n}] + \mathbb{E}[|Y_n - y|\mathbb{1}_{A_n}])}{y - x} \quad (1.9)$$

$$\leq \frac{4n^{-\varepsilon}}{y - x} \quad (1.10)$$

$$= O(n^{-\varepsilon}), \quad (1.11)$$

as $n \rightarrow \infty$. Hence we conclude that

$$\mathbb{P}(X_n \leq Y_n) \geq 1 - O(n^{-\varepsilon} + \mathbb{P}(A_n^c)),$$

as $n \rightarrow \infty$.

Adding and subtracting x and y and reordering gives (1.6). For (1.7) we have used that, for any two events A and B ,

$$\mathbb{P}(A) = \mathbb{P}(A, B) + \mathbb{P}(A, B^c) \leq \mathbb{P}(A, B) + \mathbb{P}(B^c). \quad (1.12)$$

Equation (1.8) is due to the union bound and the fact that

$$A > \frac{x}{2} \quad \text{and} \quad B > \frac{x}{2} \quad \Rightarrow \quad A + B > x. \quad (1.13)$$

We then applied Markov's inequality to get (1.9). The next line (1.10) holds since

$$|X_n - x| \leq n^{-\varepsilon} \quad \text{and} \quad |Y_n - y| \leq n^{-\varepsilon},$$

on the event A_n and $\mathbb{1}_{A_n} \leq 1$. Finally, because $y - x > 0$, we have (1.11).

Finding approximate random variables

Suppose now, we want to prove that for some $\delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > n^{-\delta}) = 0,$$

and we want to get a bound on the rate of convergence. We then often look for random variables $\{Y_n\}_{n \geq 1}$, which are easier to analyze, and that approximate $\{X_n\}_{n \geq 1}$ on an event whose probability converges to one.

As an example, suppose there exist some random variables $\{Y_n\}_{n \geq 1}$ and an event A_n which satisfies $\mathbb{P}(A_n) \rightarrow 1$ and, for all $K > 0$,

$$\mathbb{P}(|Y_n - X| > Kn^{-\delta}, A_n) \leq O(n^{-\kappa}), \quad (1.14)$$

In addition, suppose that on the event A_n we have that $|X_n - Y_n| \leq n^{-\kappa}$, for some $\kappa > 0$. Then we use the following computations:

$$\mathbb{P}(|X_n - X| > n^{-\delta}) \leq \mathbb{P}(|X_n - X| > n^{-\delta}, A_n) + \mathbb{P}(A_n^c) \quad (1.15)$$

$$\leq \mathbb{P}(|Y_n - X_n| + |Y_n - X| > n^{-\delta}, A_n) + \mathbb{P}(A_n^c) \quad (1.16)$$

$$\begin{aligned} &\leq \mathbb{P}\left(|Y_n - X_n| > \frac{n^{-\delta}}{2}, A_n\right) \\ &\quad + \mathbb{P}\left(|Y_n - X| > \frac{n^{-\delta}}{2}, A_n\right) + \mathbb{P}(A_n^c) \end{aligned} \quad (1.17)$$

$$\leq 2n^\delta \mathbb{E}[|Y_n - X_n| \mathbb{1}_{A_n}] + O(n^{-\kappa} + \mathbb{P}(A_n^c)) \quad (1.18)$$

$$= O(n^{-\kappa+\delta} + n^{-\kappa} + \mathbb{P}(A_n^c)) \quad (1.19)$$

$$= O(n^{-\kappa+\delta} + \mathbb{P}(A_n^c)), \quad (1.20)$$

as $n \rightarrow \infty$. From this we conclude that for all $\delta < \kappa$

$$\mathbb{P}(|X_n - X| > n^{-\delta}) \leq O(n^{-\kappa+\delta} + \mathbb{P}(A_n^c)),$$

as $n \rightarrow \infty$.

Here, in (1.15) we have used (1.12), while (1.16) follows by the triangle inequality. The next line (1.17) is, similar to the third line in the previous example, due to the union bound and (1.13). For (1.18) we applied Markov's inequality on the first probability in (1.17) and (1.14) on the second probability. Since, on the event A_n , we have that $|X_n - Y_n| \leq n^{-\kappa}$, we get (1.19). The last line (1.20) follows since $-\kappa + \delta > -\kappa$.

Although both proof strategies described above follow well-known arguments, the main technical difficulty lies in defining the appropriate event A_n and, in the case of the last strategy, the random variables $\{Y_n\}_{n \geq 1}$. This requires us to carefully analyze, in each specific setting, the random variables of interest $\{X_n\}_{n \geq 1}$ as well as their proposed limit X .

1.4.3 Algorithms and experiments

This thesis contains several algorithms, for instance, for generating degree-sequence and graphs. These algorithms are presented in pseudo code in a separate environment so that they can be easily referred to. We emphasize that the algorithms as we have stated them are in no way optimized. We have written them in such a way that the important steps are easy to understand. However many steps need separate implementation, when writing a full implementation of the algorithm in a specific programming language. For instance, let G be a graph and consider the following lines:

```

if  $G$  is simple then
  run Algorithm 1
else
  run Algorithm 2
end if

```

In general there are many algorithms for checking whether G is simple, some more intricate and clever than others. Still it is clear from our pseudo code that the next step in these simple lines depends on whether the graph G is simple or not. This is the message we want to get across when writing such statements. We leave it to the creative and skilled programmers to find a suitable implementation.

In addition to algorithms and proofs, we also do several numerical experiments on large graphs and networks to illustrate our results and investigate structural properties of graphs. For running our experiments we implement the algorithms that are presented in this thesis in JAVA (java.com) and make use of the libraries from the *Laboratory for Web Algorithmics* (law.di.unimi.it). We use state of the art algorithms and techniques for working with large graphs such as the *WebGraph* framework [21] and the HyperBall algorithm [22] for computing

approximate distance distributions on large graphs. In addition we often used parallel programming to speed up the computations and we ran our numerical experiments on a server with the following specifications: 1 TB RAM, 40 Intel(R) Xeon(R) CPU E7- 4870 @ 2.40GHz

Writing solid algorithms and code for experiments on large graphs is a separate topic which often requires some very novel approaches. For instance the HyperBall algorithm [22] uses Hyperloglog counters [43] to compute the approximate size of the neighborhood of a node. Since this thesis focuses on mathematical results, our implementations are by no means optimal and are hence not available online. The interested reader can request our code by sending an email to the author.

1.5 Outline of the thesis

We close this chapter by giving an outline of the rest of the thesis and summarizing the main results per chapter.

We start with some preliminaries in **Part I**. In **Chapter 2** we recall the definition of regularly-varying random variables and consider several relevant results for them. We introduce notations related to (random) graphs that are used throughout the entire thesis and define the main regularity assumptions on the degree sequences and their empirical distribution functions in Assumption 2.2 and Assumption 2.3.

We discuss two algorithms for generating degree-sequences (undirected graphs) and bi-degree sequences (directed graphs), that are close to being i.i.d. samples, in **Chapter 3**. We prove that sequences generated by these algorithms satisfy the regularity assumptions proposed in Chapter 2.

In **Part II** we analyze measures for degree-degree correlations in directed random graphs. First, in **Chapter 4**, we introduce measures for degree-degree correlations in directed graphs, for all four degree-degree correlation types. These include Pearson's correlation coefficient (Definition 4.1), two versions of Spearman's rho (Definition 4.3 and 4.4) and Kendall's tau (Definition 4.5).

In **Chapter 5** we consider general random graphs and analyze the asymptotic behavior of the measures introduced in Chapter 4. We show that Pearson's correlation coefficients for degree-degree correlations in directed networks with regularly varying degree distribution, with infinite second moment, are size dependent and converge to a non-negative number as the size of the graph tends to infinity. We also show that the requirements, in terms of the existence of certain moments for the out- and in-degree distribution, are different for the four degree-degree correlation types.

In addition, we introduce Assumption 5.1, which states that the joint degree distribution on the graph converges to some limit distribution. Then we prove that for graphs which satisfy this regularity assumption the rank-correlation measures converge. Moreover, we give an expression of their limits, which are determined by Spearman's rho and Kendall's tau on integer-valued random variables whose joint distribution is given by the distribution from Assumption 5.1.

We consider the directed configuration model in **Chapter 6**. For the general version of the model, we prove that for all four directed degree-degree correlation

types the rank-correlation measures converge to zero. We use this result, together with known results from the literature, to prove that rank-correlation measures also converge to zero in both the repeated and erased configuration model.

In **Chapter 7** we introduce an algorithm for generating graphs with a specific degree sequence, for which the Spearman's rho rank-correlation measure is minimal, with respect to the degree sequence. We prove that when the degree sequences are sampled in an i.i.d. fashion, then graphs generated by this model satisfy Assumption 5.1 and give a full description of the limiting joint distribution. Then we show that Spearman's rho on any graph that satisfies Assumption 5.1 is asymptotically bounded from below by the value of Spearman's rho obtained by our algorithm. We combine this result with a transformation procedure on the degree distribution to prove that this lower bound is independent of the tail behavior of the degree distribution. Numerical experiments show that the lower bound strongly depends on the shape of the degree distribution and can be far away from the absolute minimum -1 .

Finally, in **Part III**, we further analyze the directed configuration model. We start in **Chapter 8** with deriving upper bounds on the scaling of the number of edges that are removed in the erased configuration model, when the degree distributions are regularly varying. We give an expression for the number of removed edges and use this expression to derive probabilistic upper bounds on the scaling of the average number of removed edges in, respectively, the undirected and directed erased configuration model, for different regimes of the exponents of the degree distributions.

Chapter 9 contains empirical results. Using extensive experiments on large directed erased configuration model graphs we show that this model gives rise to structural negative Out-In degree-degree correlations, when the degree distributions are regularly varying with finite mean and infinite variance. In addition we use the scaling terms from Chapter 8 to investigate the scaling of the fluctuations of rank-correlations around their mean. Based on the large data sets that we generate, we are able to show an interesting phase transition in this scaling.

Path lengths in the directed configuration model are analyzed in **Chapter 10**. We prove that when the covariance between the out- and in-degrees is finite, the shortest path between two nodes in the directed configuration model scales logarithmically with the size of the graph. In addition, we characterize the fluctuations of these path lengths around their typical length, as the graph size tends to infinity, and use experiments to show that this theoretical result is a very accurate approximation for graphs of finite size.

For the proof of distances in the directed configuration model we use a novel coupling between a breadth-first graph exploration process and a Galton-Watson tree. The coupled branching process has a deterministic offspring distribution that does not depend on n or the degree sequences. We show that this coupling holds for a number of steps in the graph exploration process equivalent to discovering $n^{1-\epsilon}$ nodes, for arbitrarily small $\epsilon > 0$.

Part I

Preliminaries

Chapter 2

Notations and definitions

2.1 Probabilistic tools

Given a random variable X , we denote by f_X and F_X , respectively, its probability mass function and cumulative distribution function. In addition we denote by $h_{X,Y}$ and $H_{X,Y}$ the joint probability mass function and cumulative distribution of two random variables X and Y . Often, when there is no confusion on which random variable we are considering, we omit the subscripts.

Remark 2.1. We always use lowercase letters (f) to denote probability mass functions and the uppercase version (F) to denote the corresponding cumulative distribution function.

Remark 2.2. Since most of the characteristics of graphs are integer valued and non-negative, we often work with random variables whose probability mass functions have support on the non-negative integers. In addition, we reserve k, ℓ to denote such integers and write $k \geq 0$, for $k \in \{0, 1, 2, \dots\}$.

We adopt the convention that if A is some event, then A^c will denote the complement of this event. Consequently, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

In addition we use $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{d} X$ to denote that a sequence X_n of random variables converges to X in, respectively, probability and distribution, as $n \rightarrow \infty$.

2.1.1 Convergence of probability mass functions

In this thesis we often need to consider the convergence of sequences of probability mass functions. For this we use two notions of a distance on the space of distributions.

The first is the total variation distance which, for two probability mass functions f and g , on the non-negative integers, can be defined as

$$d_{\text{tv}}(f, g) = \frac{1}{2} \sum_{k=0}^{\infty} |f(k) - g(k)|. \quad (2.1)$$

In addition we also use the Kantorovich-Rubinstein distance (also known as Wasserstein metric of order one). Let us first recall the formal definition.

Definition 2.1. Let $M(f, g)$ denote the set of joint probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals f and g . Then, the Kantorovich-Rubinstein distance between f and g is given by

$$d_1(f, g) = \inf_{\pi \in M(f, g)} \int_{\mathbb{R} \times \mathbb{R}} |x - y| d\pi(x, y).$$

We point out that, strictly speaking, d_1 is only a distance when both f and g have finite first absolute moments. Moreover, it is well known that

$$d_1(f, g) = \int_0^1 |F^{-1}(u) - G^{-1}(u)| du = \int_{-\infty}^{\infty} |F(x) - G(x)| dx,$$

where F and G are the cumulative distribution functions of f and g , respectively, and $\phi^{-1}(t) = \inf\{x \in \mathbb{R} : \phi(x) \geq t\}$ denotes the pseudo-inverse of the function ϕ . It follows that the optimal coupling of two real random variables X and Y , with respect to this distance, is given by $(X, Y) = (F^{-1}(U), G^{-1}(U))$, where U is uniformly distributed in $[0, 1]$. In particular, for probability mass functions f and g on the non-negative integers, we have,

$$d_1(f, g) = \sum_{k=0}^{\infty} \left| \sum_{t=0}^k f(t) - g(t) \right|. \quad (2.2)$$

Note that since f and g are probability mass functions, (2.2) is equivalent to

$$d_1(f, g) = \sum_{k=0}^{\infty} \left| \sum_{t=k+1}^{\infty} f(t) - g(t) \right|. \quad (2.3)$$

Observe that convergence in the Kantorovich-Rubinstein distance implies convergence in total variation distance. Moreover, if X_n and X have, respectively, probability mass functions f_n and f , then convergence of f_n to f in the Kantorovich-Rubinstein distance is equivalent to weak convergence plus convergence of the first absolute moments of X_n . Note that when X_n is integer-valued and $X_n \xrightarrow{d} X$, as $n \rightarrow \infty$, then $d_{\text{tv}}(X_n, X) \rightarrow 0$.

2.1.2 Regularly-varying random variables

The theory of regularly-varying random variables is an application of the theory of regular variation, which for instance has applications in extreme value theory. We refer the interested reader to the book by Bingham, Goldie and Teugels [14]. Other resources for this section are the book by Whitt on limits of stochastic processes [93] and the book on phenomena related to heavy-tails by Resnick [71].

The definition of regularly-varying random variables involves slowly-varying functions. To define the latter let ϕ and ψ be two functions. Then we write $\phi(x) \sim \psi(x)$, as $x \rightarrow \infty$, when

$$\lim_{x \rightarrow \infty} \frac{\phi(x)}{\psi(x)} = 1.$$

With this notation we can introduce the concept of a slowly-varying function and a regularly-varying random variable:

Definition 2.2. A function $\mathcal{L} : \mathbb{R} \rightarrow \mathbb{R}$ is called slowly varying if, for all $t > 0$

$$\mathcal{L}(tx) \sim \mathcal{L}(x), \quad \text{as } x \rightarrow \infty.$$

Definition 2.3. A random variable X is said to have a regularly-varying distribution if there exists a slowly-varying function $\mathcal{L}$, so that

$$\mathbb{P}(X > x) \sim \mathcal{L}(x)x^{-\gamma} \quad \text{as } x \rightarrow \infty. \quad (2.4)$$

The parameter γ is called the exponent of the distribution.

A function $\phi(x)$, defined on $[x_0, \infty)$, is called regularly varying at infinity with index κ , when

$$\phi(tx) \sim \phi(x)t^\kappa,$$

as $x \rightarrow \infty$, for all $t > 0$. Hence, since

$$\mathbb{P}(X > tx) \sim \mathcal{L}(tx)(tx)^{-\gamma} \sim t^{-\gamma}\mathcal{L}(x)x^{-\gamma} \sim t^{-\gamma}\mathbb{P}(X > x),$$

Definition 2.3 states that a regularly-varying random variable with exponent γ is a random variable X for which $1 - F_X(x)$ is regularly varying at infinity with index $-\gamma$.

An important result in the theory of regular variation is Karamata's Theorem, which states that integrals of regularly-varying functions are again regularly varying. We state this result adjusted to the setting in which we use it.

Theorem 2.1 (Karamata's Theorem, [14, 71]). *Let X be regularly varying with exponent γ and slowly-varying function $\mathcal{L}$. Then, as $t \rightarrow \infty$,*

i) *for $\gamma < 1$*

$$\mathbb{E}[X \mathbb{1}_{\{X \leq t\}}] \sim \frac{\mathcal{L}(t)t^{1-\gamma}}{1-\gamma},$$

ii) *for $\gamma > 1$*

$$\mathbb{E}[X \mathbb{1}_{\{X \geq t\}}] \sim \frac{\mathcal{L}(t)t^{1-\gamma}}{\gamma-1} \quad \text{and}$$

iii) *for $\gamma = 1$ we have that $\mathbb{E}[X \mathbb{1}_{\{X \leq t\}}]$ is slowly varying, while this holds for $\mathbb{E}[X \mathbb{1}_{\{X \geq t\}}]$ if $\mathbb{E}[X] < \infty$.*

Another theorem, Karamata's Tauberian Theorem, gives an equivalent definition for a regularly-varying random variable, which involve its Laplace-Stieltjes transform. In this thesis we use a slight adaptation of this theorem, due to [13]:

Theorem 2.2 (Tauberian Theorem, [13]). *Let X be a non-negative random variable with only finite mean and $\mathcal{L}$ a slowly-varying function. Then, for $1 < \gamma < 2$, the following are equivalent,*

i) $\mathbb{P}(X > t) \sim \mathcal{L}(t)t^{-\gamma} \quad \text{as } t \rightarrow \infty,$

ii) $\frac{\mathbb{E}[X]}{t} - 1 + \mathbb{E}\left[\exp\left\{-\frac{X}{t}\right\}\right] \sim \mathcal{L}\left(\frac{1}{t}\right)t^{-\gamma} \quad \text{as } t \rightarrow \infty.$

The last result we mention is a Central Limit Theorem for regularly-varying random variables, which is also called the Stable Law CLT. We state it for the specific case where the random variables are non-negative, since this is the setting that we use when dealing with degrees in graphs:

Theorem 2.3 (Stable Law CLT [93]). *Let $\{X_i\}_{i \geq 1}$ be a sequence of i.i.d. copies of a non-negative regularly-varying random variable X with exponent $0 < \gamma < 2$ and slowly-varying function $\mathcal{L}$. Then there exists a slowly-varying function $\mathcal{L}_0$, in general different from $\mathcal{L}$, such that*

$$\frac{\sum_{i=1}^n X_i - m_n}{\mathcal{L}_0(n)n^{\frac{1}{\gamma}}} \xrightarrow{d} \mathcal{S}_\gamma,$$

where $\mathcal{S}_\gamma$ has a stable distribution with shape parameter γ and

$$m_n = \begin{cases} 0 & \text{if } 0 < \gamma < 1 \\ n^2 \mathbb{E} \left[\sin \left(\frac{X}{\mathcal{L}_0(n)n} \right) \right] & \text{if } \gamma = 1 \\ n \mathbb{E}[X] & \text{if } 1 < \gamma < 2. \end{cases}$$

2.1.3 Scaling of regularly-varying random variables

In this thesis we are often interested in determining the scaling of functionals of $\sum_{i=1}^n X_i^p$, for some integer $p \geq 1$ and sequence $\mathbf{X}_n = \{X_1, X_2, \dots, X_n\}$ of i.i.d. copies of a non-negative regularly-varying random variable X .

From Theorem 2.3 we can infer several scaling results. First, by Potter's Theorem [14, Theorem 1.5.6] we have that for any slowly-varying function $\mathcal{L}$ and $\kappa > 0$,

$$\lim_{n \rightarrow \infty} n^{-\kappa} \mathcal{L}(n) = 0. \quad (2.5)$$

Therefore it follows, by Theorem 2.3 and Slutsky's Theorem, that for any $\kappa > 0$,

$$n^{-\kappa} \frac{\sum_{i=1}^n X_i - m_n}{n^{\frac{1}{\gamma}}} = n^{-\kappa/2} \mathcal{L}_0(n) \frac{\sum_{i=1}^n X_i - m_n}{\mathcal{L}_0(n)n^{\frac{1}{\gamma} + \frac{\kappa}{2}}} \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty, \quad (2.6)$$

which implies that for any $K > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \sum_{i=1}^n X_i - m_n \right| > K n^{\frac{1}{\gamma} + \kappa} \right) = 0.$$

When $\gamma < 1$, $m_n = 0$ so that we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i=1}^n X_i > K n^{\frac{1}{\gamma} + \kappa} \right) = 0,$$

while for $1 < \gamma < 2$,

$$\mathbb{P} \left(\sum_{i=1}^n X_i > K n^{1+\kappa} \right) \leq \mathbb{P} \left(\left| \sum_{i=1}^n X_i - n \mathbb{E}[X] \right| > \frac{K n^{1+\kappa}}{2} \right)$$

$$\begin{aligned}
& + \mathbb{P} \left(n\mathbb{E}[X] > \frac{Kn^{1+\kappa}}{2} \right) \\
& \leq \mathbb{P} \left(\left| \sum_{i=1}^n X_i - n\mathbb{E}[X] \right| > \frac{Kn^{1+\kappa}}{2} \right) + \frac{2n^{-\kappa}\mathbb{E}[X]}{K},
\end{aligned}$$

which converges to zero as $n \rightarrow \infty$.

For the boundary case $\gamma = 1$, we first use Markov's inequality and the union bound to get

$$\begin{aligned}
\mathbb{P} \left(\sum_{i=1}^n X_i > Kn^{1+\kappa} \right) & \leq \mathbb{P} \left(\sum_{i=1}^n X_i > Kn^{1+\kappa}, \max_{1 \leq i \leq n} X_i \leq n^{1+\kappa} \right) \\
& + \mathbb{P} \left(\max_{1 \leq i \leq n} X_i \leq n^{1+\kappa} \right) \\
& \leq \frac{n^{-\kappa}}{K} \mathbb{E} [X_1 \mathbb{1}_{\{X_1 \leq n^{1+\kappa}\}}] + n\mathbb{P}(X_1 > n^{1+\kappa}).
\end{aligned}$$

By Theorem 2.1 iii), $\mathbb{E} [X_1 \mathbb{1}_{\{X_1 \leq n^{1+\kappa}\}}]$ is slowly varying and hence the first term converges to zero by (2.5). For the second term, as $n \rightarrow \infty$,

$$n\mathbb{P}(X_1 > n^{1+\kappa}) \sim \mathcal{L}(n^{1+\kappa})n^{-\kappa} = o(1),$$

where we have again used (2.5) for the last equality.

Therefore, for any sequence $\mathbf{X}_n$ of i.i.d. copies of a non-negative regularly-varying random variable X with exponent $0 < \gamma < 2$ and for any $\kappa, K > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i=1}^n X_i > Kn^{(\frac{1}{\gamma} \vee 1) + \kappa} \right) = 0,$$

where $x \vee y$ denotes the maximum of x and y .

In addition, when $\gamma < 1$, $\mathcal{S}_\gamma$ has a positive density on $(0, \infty)$, so that by the continuous mapping theorem,

$$\frac{\mathcal{L}_0(n)n^{\frac{1}{\gamma}}}{\sum_{i=1}^n X_i} \xrightarrow{d} \frac{1}{\mathcal{S}_\gamma},$$

as $n \rightarrow \infty$. In this case Slutsky's Theorem implies that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i=1}^n X_i < Kn^{\frac{1}{\gamma} - \kappa} \right) = 0,$$

for all $K, \kappa > 0$.

We finally remark that if X is regularly varying with exponent γ , then X^p is regularly varying with exponent γ/p , for $p > 0$. We can now summarize the above as the following useful result:

Proposition 2.4. *Let $\{X_i\}_{i \geq 1}$ be a sequence of i.i.d. copies of a non-negative regularly-varying random variable X with exponent $\gamma > 1$. Then,*

i) for any integer $p \geq 1$ and $K, \kappa > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i=1}^n X_i^p \leq K n^{\left(\frac{p}{\gamma} \vee 1\right) + \kappa} \right) = 1,$$

and

ii) for any integer $p \geq 1$ with $\gamma < p$ and $K, \kappa > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i=1}^n X_i^p \geq K n^{\frac{p}{\gamma} - \kappa} \right) = 1.$$

The following definition is useful for expressing the scaling of functionals related to regularly-varying random variables.

Definition 2.4. Let $\{X_n\}_{n \geq 1}$ be random variables and let $\kappa \in \mathbb{R}$. Then we define

$$X_n = O_{\mathbb{P}}(n^{\kappa}) \iff \text{for all } \delta > 0 \quad n^{-\kappa - \delta} X_n \xrightarrow{\mathbb{P}} 0, \quad \text{as } n \rightarrow \infty.$$

Observe that with this definition, for instance,

$$X_n = O_{\mathbb{P}}(n^{\kappa_x}) \quad \text{and} \quad Y_n = O_{\mathbb{P}}(n^{\kappa_y}) \quad \Rightarrow \quad X_n Y_n = O_{\mathbb{P}}(n^{\kappa_x + \kappa_y}).$$

In general, standard operations, as used for the big- O notation, also apply to $O_{\mathbb{P}}$.

Moreover, by (2.5) it follows that for random variables $\{Y_n\}_{n \geq 1}$ and a slowly-varying function $\mathcal{L}$,

$$O_{\mathbb{P}}(\mathcal{L}(n)Y_n) = O_{\mathbb{P}}(Y_n).$$

Hence, Definition 2.4 is useful when we want to express the scaling of $\sum_{i=1}^n X_i^p$, for sequences $\mathbf{X}_n$ of regularly-varying random variables, only in terms of $n^{-\kappa}$ or we do not know the exact expression for the slowly-varying functions involved. In addition, we note that Proposition 2.4 now states that for any integer $p \geq 1$,

$$\sum_{i=1}^n X_i^p = O_{\mathbb{P}}\left(n^{\left(\frac{p}{\gamma} \vee 1\right)}\right), \tag{2.7}$$

which reflects the statement found in the literature that $\sum_{i=1}^n X_i^p$ scales as $n^{\frac{p}{\gamma}}$, for $p \geq 2$ and as n when $p = 1$.

When X has finite mean we also obtain the following useful results:

Lemma 2.5. Let $\{X_i\}_{i \geq 1}$ be i.i.d. copies of a non-negative regularly-varying random variable X with exponent $\gamma > 1$, denote by $S_n = \sum_{i=1}^n X_i$ and $\nu = \mathbb{E}[X]$. Then, for any two integers $p, q \geq 0$,

$$\text{i) } |S_n - \nu n| = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma}}\right),$$

$$\text{ii) } \left(\frac{1}{S_n}\right)^q = O_{\mathbb{P}}(n^{-q}) \text{ and}$$

$$\text{iii)} \quad \left(\frac{1}{S_n}\right)^q \sum_{i=1}^n X_i^p = O_{\mathbb{P}}\left(n^{\left(\frac{p}{\gamma} \vee 1\right)-q}\right).$$

Proof. The first statement follows from (2.6) with the observation that $m_n = n\mathbb{E}[X]$ for $\gamma > 1$. Moreover, combining ii) with (2.7) yields iii). Therefore, we only need to prove ii).

For this let $\kappa = (\gamma - 1)/(\gamma + 1)$ and denote by A_n the event

$$A_n = \{|S_n - \nu n| \leq n^{1-\kappa}\}.$$

Note that since $\mathbb{E}[X^{1/(1-\kappa)}] = \mathbb{E}[X^{(1+\gamma)/2}] < \infty$, it follows from the Kolmogorov-Marcinkiewicz-Zygmund strong law of large numbers, cf. [52], that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{S_n - \nu n}{n^{1-\kappa}} = 0\right) = 1,$$

and hence

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 1, \tag{2.8}$$

as $n \rightarrow \infty$. Since on the event A_n we have that

$$\left(\frac{1}{S_n}\right)^q \leq \left(\frac{1}{\nu n - n^{1-\kappa}}\right)^q,$$

it follows that, for any $\delta, K > 0$,

$$\begin{aligned} \mathbb{P}\left(\left(\frac{1}{S_n}\right)^q > K n^{-q+\delta}\right) &\leq \mathbb{P}\left(\left(\frac{1}{S_n}\right)^q > K n^{-q+\delta}, A_n\right) + \mathbb{P}(A_n^c) \\ &\leq O(n^{-\delta} + \mathbb{P}(A_n^c)), \end{aligned}$$

as $n \rightarrow \infty$, which proves ii). $\square$

The last result of this section is due to Karamata's theorem. Although we use it in Chapter 8 to analyze convergence of specific terms, it can be of independent interest:

Lemma 2.6. *Let X be regularly varying with exponent $1 < \gamma < 2$ and slowly-varying function $\mathcal{L}$. Then,*

$$\mathbb{E}\left[X\left(1 \wedge \frac{X}{t}\right)\right] \sim \frac{\gamma}{3\gamma - 2 - \gamma^2} \mathcal{L}(t) t^{1-\gamma},$$

as $t \rightarrow \infty$ and as a consequence, for any $\delta > 0$,

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}\left[X\left(1 \wedge \frac{X}{t}\right)\right]}{t^{1-\gamma+\delta}} = 0.$$

Proof. First observe that the second statement follows from the first by (2.5)

$$\lim_{t \rightarrow \infty} t^{-\delta} \mathcal{L}(t) = 0.$$

For the first statement note that X^2 is regularly varying with exponent $\gamma/2 < 1$ and slowly-varying function $\mathcal{L}_2(t) = \mathcal{L}(\sqrt{t})$. Hence it follows from Karamata's Theorem that, as $t \rightarrow \infty$,

$$\begin{aligned}
\mathbb{E} \left[X \left(1 \wedge \frac{X}{t} \right) \right] &= t^{-1} \mathbb{E} [X^2 \mathbb{1}_{\{X \leq t\}}] + \mathbb{E} [X \mathbb{1}_{\{X > t\}}] \\
&= t^{-1} \mathbb{E} [X^2 \mathbb{1}_{\{X^2 \leq t^2\}}] + \mathbb{E} [X \mathbb{1}_{\{X > t\}}] \\
&\sim t^{-1} \frac{\mathcal{L}_2(t^2) t^{2-\gamma}}{1 - \frac{\gamma}{2}} + \frac{\mathcal{L}(t) t^{1-\gamma}}{\gamma - 1} \\
&\sim \frac{\mathcal{L}_2(t^2) t^{1-\gamma}}{1 - \frac{\gamma}{2}} + \frac{\mathcal{L}(t) t^{1-\gamma}}{\gamma - 1} \\
&\sim \left(\frac{1}{1 - \frac{\gamma}{2}} + \frac{1}{\gamma - 1} \right) \mathcal{L}(t) t^{1-\gamma} \\
&\sim \frac{\gamma}{3\gamma - 2 - \gamma^2} \mathcal{L}(t) t^{1-\gamma}.
\end{aligned}$$

□

2.1.4 Continuization

In this thesis we use a continuization principle, as applied for instance in [57], where we transform given integer-valued random variables into continuous ones.

Definition 2.5. Let X be an integer-valued random variable and U a uniformly-distributed random variable on $[0, 1)$ independent of X . Then we define the *continuization* of X as

$$\tilde{X} = X + U.$$

We refer to U as the *continuous part* of $\tilde{X}$. We remark that although we have chosen U to be uniform we could instead take any continuous random variable on $[0, 1)$ with strictly increasing F , cf. [36].

We now establish several relations between the distribution functions of integer-valued random variables and their continuizations. Recall that for two random variables X and Y , we denote their cumulative distribution functions by, respectively, F_X and F_Y and their joint distribution by $H_{X,Y}$.

In addition we make use of the following functions, related to these distributions,

$$\begin{aligned}
\mathcal{F}_X(k) &= F_X(k) + F_X(k-1), \\
\mathcal{F}_Y(\ell) &= F_Y(\ell) + F_Y(\ell-1), \\
\mathcal{H}_{X,Y}(k, \ell) &= H_{X,Y}(k, \ell) + H_{X,Y}(k-1, \ell) \\
&\quad + H_{X,Y}(k, \ell-1) + H_{X,Y}(k-1, \ell-1).
\end{aligned}$$

Let $\tilde{X} = X + U$ be as in Definition 2.5, take $t \in \mathbb{Z}$ and define $I_t = [t, t+1)$. Then for, $x \in I_t$,

$$\tilde{F}_{\tilde{X}}(x) = (x-t)F_X(t) + (t+1-x)F_X(t-1). \quad (2.9)$$

As a consequence, it follows that for $x \in I_t$,

$$dF_{\widetilde{X}}(x) = (F_X(t) - F_X(t-1)) dx = \mathbb{P}(X = t) dx. \quad (2.10)$$

These identities capture the essential relations between X and its continuization $\widetilde{X}$. As a first result we have the following:

Lemma 2.7. *Let X be an integer-valued random variable and $m \in \mathbb{N}$. Then,*

$$\mathbb{E} \left[F_{\widetilde{X}}(\widetilde{X})^m \right] = \frac{1}{m+1} \sum_{i=0}^m \mathbb{E} \left[F_X(X)^i F_X(X-1)^{m-i} \right].$$

Proof. Using (2.9) we obtain,

$$\begin{aligned} \int_{I_t} F_{\widetilde{X}}(x)^m dx &= \int_{I_t} ((x-t)F_X(t) + (t+1-x)F_X(t-1))^m dx \\ &= \sum_{i=0}^m \binom{m}{i} F_X(t)^i F_X(t-1)^{m-i} \int_0^1 (y)^i (1-y)^{m-i} dy \\ &= \sum_{i=0}^m \frac{m!}{i!(m-i)!} F_X(t)^i F_X(t-1)^{m-i} \frac{\Gamma(i+1)\Gamma(m-i+1)}{\Gamma(m+2)} \\ &= \frac{1}{m+1} \sum_{i=0}^m F_X(t)^i F_X(t-1)^{m-i}, \end{aligned}$$

where $\Gamma(t)$ denotes the gamma function. Combining this with (2.10), we get

$$\begin{aligned} \mathbb{E} \left[F_{\widetilde{X}}(\widetilde{X})^m \right] &= \sum_{t \in \mathbb{Z}} \int_{I_t} F_{\widetilde{X}}(x)^m dF_{\widetilde{X}}(x) \\ &= \sum_{t \in \mathbb{Z}} \int_{I_t} F_{\widetilde{X}}(x)^m \mathbb{P}(X = t) dx \\ &= \frac{1}{m+1} \sum_{i=0}^m \mathbb{E} \left[F_X(X)^i F_X(X-1)^{m-i} \right]. \end{aligned}$$

□

Since $F_{\widetilde{X}}(\widetilde{X})$ is uniformly distributed on $(0, 1)$

$$\frac{1}{m+1} = \mathbb{E} \left[F_{\widetilde{X}}(\widetilde{X})^m \right],$$

and hence it follows from Lemma 2.7 that for all $m \geq 0$,

$$\sum_{i=0}^m \mathbb{E} \left[F_X(X)^i F_X(X-1)^{m-i} \right] = 1. \quad (2.11)$$

2.2 Graphs and degree sequences

A common way to characterize a graph is to write $G = (V, E)$, where V denotes the set of vertices and E is the set of edges. When G is directed, E consists of ordered pairs (v, w) , for $v, w \in V$, while in the undirected case one writes $\{v, w\}$ for an edge between v and w . When using this characterization of graphs, one often writes $e \in E$ when working with edges and only specifies $e = (v, w)$ when the vertices that are part of the edge need to be considered.

In this thesis, however, we use a different notation for graphs, vertices and a non-standard notation for edges, which we now explain.

2.2.1 A different characterization of graphs

For most of the results in this thesis we are interested in analyzing properties of graphs as their size goes to infinity. Therefore, instead of specifying the set of vertices V , we denote by G_n a graph of size n , where the nodes are enumerated by the set $\{1, 2, \dots, n\}$.

Remark 2.3. Whenever graphs are involved, n will *always* denote the size of the graph, unless explicitly stated otherwise.

For the edges we use the following notation. When G_n is directed, we let $i \rightarrow j$ denote an edge from i to j , while for undirected graphs we write $i - j$ for an edge between i and j . In addition we write $i \rightarrow j \in G_n$ to indicate that the edge $i \rightarrow j$ exists in the graph G_n .

With these notations, when φ is a function on the edges of the graph G_n , we write $\varphi(i \rightarrow j)$ to denote the value of this function on a specific edge $i \rightarrow j$.

Given a graph G_n we use G_{ij} to denote the number of edges from i to j . In addition we write $\sum_{i \rightarrow j}$ to denote the sum over all edges in G_n . To be more precise, when $\phi(i, j)$ is a function on pairs of nodes of the graph G_n , then

$$\sum_{i \rightarrow j} \phi(i, j) = \sum_{i, j=1}^n G_{ij} \phi(i, j).$$

Observe that for undirected graphs this convention means that we count each edge $i - j$ twice. We sometimes write $\sum_{i \rightarrow j \in G_n}$ when we need to specify over which graph we take the sum. Finally, by Remark 2.3, the subscript n will only be used to indicate the size of the graph, so that there is a clear distinction between G_n and G_{ij} .

2.2.2 Degrees and degree sequences

We first consider the setting of directed graphs and at the end briefly repeat the definitions for undirected graphs.

Directed graphs

Given a directed graph G_n , we denote by D_i^+ and D_i^- the out- and in-degree of node i , respectively. In terms of G_{ij} these satisfy,

$$D_i^+ = \sum_{j=1}^n G_{ij} \quad \text{and} \quad D_i^- = \sum_{j=1}^n G_{ji}.$$

In addition we let

$$L_n = \sum_{i=1}^n D_i^+ = \sum_{i=1}^n D_i^-$$

denote the sum of the degrees, which is equal to the number of edges in G_n .

Since we often have to consider different combinations of out- and in-degrees we index the degree type with $\alpha \in \{+, -\}$ and write D_i^α to denote the α -degree of node i .

Remark 2.4. Throughout this thesis $\alpha, \beta \in \{+, -\}$ will always be used to index the degree type of nodes.

An important characterization of a graph G_n is its bi-degree sequence

$$(\mathbf{D}_n^+, \mathbf{D}_n^-) = \{(D_1^+, D_1^-), (D_2^+, D_2^-), \dots, (D_n^+, D_n^-)\}.$$

The bi-degree sequence can be seen as a first order characterization of the structure of graphs, since it describes the degrees of the nodes but not the structure of the edges.

Given a bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ we define, for $k, \ell \geq 0$, the following empirical probability mass functions,

$$f_n^\alpha(k) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i^\alpha=k\}}, \quad (2.12)$$

$$h_n(k, \ell) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i^+=k\}} \mathbb{1}_{\{D_i^-=\ell\}}, \quad (2.13)$$

$$f_n^{\alpha, \beta}(k) = \frac{1}{L_n} \sum_{i=1}^n D_i^\alpha \mathbb{1}_{\{D_i^\beta=k\}}, \quad (2.14)$$

and let F_n^α , H_n and $F_n^{\alpha, \beta}$ denote the corresponding cumulative distribution functions. Furthermore, we define the empirical expectation and joint expectation, with respect to the by

$$\nu_n = \frac{L_n}{n} \quad \text{and} \quad \mu_n = \frac{1}{L_n} \sum_{i=1}^n D_i^+ D_i^-.$$

With these definitions, $f_n^\alpha(k)$ is the probability that if we pick a node i , uniformly at random amongst all nodes, its α -degree will be k . We therefore refer to this as the empirical degree probability mass function or distribution. Similarly h_n is the empirical joint degree probability mass function of a randomly sampled

node. The probability mass functions $f_n^{\alpha,\beta}(k)$ are the so-called size-biased degree distributions. To explain this, let $\alpha = +$, then

$$f_n^{+,\beta}(k) = \frac{1}{L_n} \sum_{i=1}^n \sum_{j=1}^n G_{ij} \mathbb{1}_{\{D_i^\beta=k\}} = \frac{1}{L_n} \sum_{i \rightarrow j} \mathbb{1}_{\{D_i^\beta=k\}}.$$

Hence, $f_n^{+,\beta}(k)$ is the probability that if we pick an edge $i \rightarrow j$, uniformly at random, the β -degree of the source i equals k . Therefore, $f_n^{+,\beta}(k)$ is the probability mass function of the β -degree of a randomly sampled node, biased on the out-degrees. When $\alpha = -$, $f_n^{-,\beta}(k)$ denotes the probability mass function of the β -degree of a randomly sampled node, biased on the in-degrees.

We also consider the empirical joint α - β degree probability mass function $h_n^{\alpha,\beta}$, which is the joint probability of the α -degree of i and the β -degree of j , for an edge $i \rightarrow j$, sampled uniformly at random,

$$h_n^{\alpha,\beta}(k, \ell) = \frac{1}{L_n} \sum_{i \rightarrow j} \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}}. \quad (2.15)$$

Similar to the other probability mass functions, we let $H_n^{\alpha,\beta}$ denote the cumulative distribution function of $h_n^{\alpha,\beta}$. We remark that, unlike the size-biased probability mass functions $f_n^{\alpha,\beta}$, the joint degree probability mass function depends on the full structure of the graph and not only on the bi-degree sequence.

In some expressions, especially in Chapter 10, we use the index $\pm$ and $\mp$ to indicate both the expression with a $+$ and a $-$ sign, simultaneously, where the index $\pm$ and $\mp$ should be substituted consistently. For instance when we write $f_n^{\pm,\mp}$, we mean either $f_n^{+,-}$ or $f_n^{-,+}$.

Undirected graphs

When G_n is undirected we denote the degree of node i by D_i , and let $L_n = \sum_{i=1}^n D_i$. Observe that L_n is even and the number of edges for undirected graphs equals $L_n/2$. We denote the degree sequence of G_n by $\mathbf{D}_n$, define the corresponding empirical probability mass functions by

$$f_n(k) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i=k\}} \quad (2.16)$$

$$f_n^*(k) = \frac{1}{L_n} \sum_{i=1}^n D_i \mathbb{1}_{\{D_i=k\}} \quad (2.17)$$

$$h_n^*(k, \ell) = \frac{1}{L_n} \sum_{i \rightarrow j} \mathbb{1}_{\{D_i=k\}} \mathbb{1}_{\{D_j=\ell\}}, \quad (2.18)$$

and let F_n , F_n^* and H_n^* denote the corresponding cumulative distribution functions.

Remark 2.5. In the rest of this thesis, we use the term bi-degree sequence to refer to a sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ which satisfies $\sum_{i=1}^n D_i^+ = \sum_{i=1}^n D_i^-$. Likewise, with a degree sequence we mean a sequence $\mathbf{D}_n$ such that $\sum_{i=1}^n D_i$ is even.

2.3 Random graphs and regularity assumptions

In this thesis we often deal with random graphs where we have to analyze the convergence of some expressions that depend on the structure of these graphs. To enable such analysis we need to impose certain regularity assumptions on the graphs. In this section we consider several of such assumptions and explain the intuition behind them. As before, we first focus on directed graphs and then discuss undirected graphs

Remark 2.6. For the asymptotic analysis of graphs we often consider objects X_n and say that they satisfy a certain property P , asymptotically, when

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in P) = 1.$$

By this we implicitly mean that there is a sequence $\{X_n\}_{n \geq 0}$ of such objects for which the above statement holds. We use this convention to avoid having to explicitly write that we consider such sequences, which is especially cumbersome when X_n denotes, for instance, a bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$.

Directed graphs

As we have mentioned in the previous section, a bi-degree sequence does not completely specify a graph. They are however an important characterization of graphs. In light of this, we consider two sources of randomness in graphs; one will be in the bi-degree sequence, while the other is in the structure of the edges. Some random graph models will have only one of these flavors, for instance, when the bi-degree sequence is created using a random process while the connection of nodes follows a deterministic algorithm, given the bi-degree sequence. Others will have both, the assignment of the degrees as well the connection of nodes follows a random process. Here we will discuss a set of regularity assumptions for the bi-degree sequence. The regularity assumptions for the connection of the nodes will be treated separately, in the chapters where we use them.

Suppose that $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ is created using some random process. Then the empirical probability mass functions f_n^α , h_n and $f_n^{\alpha, \beta}$ will also be random. We use $\mathbb{E}_n$ and $\mathbb{P}_n$ to denote the expectation and probability conditioned on the bi-degree sequence.

In order to analyze asymptotic properties of graphs with a random bi-degree sequence, we would like that the random empirical distributions converge, in some sense, to limiting distributions, so that we can use the latter to do the analysis. In addition we would also like to have convergence of the empirical means ν_n , and sometimes μ_n . An example of a general set of such conditions is the following:

Assumption 2.1. There exist integer-valued random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ with distributions F^+ and F^- , respectively, and joint distribution H that satisfy

$$\mathbb{E}[\mathcal{D}^+] = \mathbb{E}[\mathcal{D}^-] = \nu < \infty, \quad \mathbb{E}[\mathcal{D}^+ \mathcal{D}^-] = \mu < \infty, \quad (2.19)$$

such that for any $k, \ell \geq 0$, as $n \rightarrow \infty$,

$$\text{D1) } F_n^+(k) \xrightarrow{\mathbb{P}} F^+(k) \text{ and } F_n^-(k) \xrightarrow{\mathbb{P}} F^-(k),$$

$$\text{D2)} \quad H_n(k, \ell) \xrightarrow{\mathbb{P}} H(k, \ell),$$

$$\text{D3)} \quad \nu_n \xrightarrow{\mathbb{P}} \nu, \quad \text{and}$$

$$\text{D4)} \quad \mu_n \xrightarrow{\mathbb{P}} \mu.$$

When there exist such random variables $\mathcal{D}^+$ and $\mathcal{D}^-$, we say that the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfies Assumption 2.1.

This assumption, and related types, are often used in the literature, c.f. [79, 28]. One reason for this is that when the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ is constructed in an i.i.d. fashion by sampling from two random variables $\mathcal{D}^+$ and $\mathcal{D}^-$, which satisfy (2.19) then it can be shown, see [28], that $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfies Assumption 2.1. See Chapter 3 for an example of such an algorithm.

Unfortunately, since Assumption 2.1 deals with convergence in probability, the results that follow from it will often be no stronger than that. As we mentioned in Section 1.4.1, we are interested in results of the form

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > n^{-\delta}) = O(n^{-\varepsilon}),$$

for some $\delta, \varepsilon > 0$, which is stronger than $X_n \xrightarrow{\mathbb{P}} X$. In addition we also consider the size-biased degree probability mass functions, which are not included in Assumption 2.1. Therefore, we introduce a different, slightly stronger assumption, which deals with the convergence of the Kantorovich-Rubinstein and total-variation distance between the empirical and limiting distributions.

Assumption 2.2. There exist probability mass functions $f^\alpha, f^{\alpha, \beta}$ on the non-negative integers for all $\alpha, \beta \in \{+, -\}$, that satisfy, for some $\eta > 0$

$$\nu := \sum_{k=0}^{\infty} k f^+(k) = \sum_{k=0}^{\infty} k f^-(k) < \infty, \quad (2.20)$$

$$\sum_{k=0}^{\infty} k^{1+\eta} f^+(k) < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} k^{1+\eta} f^-(k) < \infty. \quad (2.21)$$

In addition, there exists an $\varepsilon > 0$ such that, if

$$\Omega_n = \left\{ \left(\max_{\alpha \in \{+, -\}} d_1(f_n^\alpha, f^\alpha) \right) \vee \left(\max_{\alpha, \beta \in \{+, -\}} d_{tv}(f_n^{\alpha, \beta}, f^{\alpha, \beta}) \right) \leq n^{-\varepsilon} \right\}, \quad (2.22)$$

then

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Omega_n) = 1.$$

Remark 2.7. Although Assumption 2.2 seems stronger than 2.1, it is still satisfied by bi-degree sequences created using the i.i.d. sampling techniques, see Theorem 3.2

Observe that on the event Ω_n the empirical degree distributions and size-biased degree distributions converge weakly, while, due to the use of the Kantorovich-Rubinstein distance, weak convergence also holds for the empirical mean ν_n .

Another observation is that Assumption 2.2 uses the Kantorovich-Rubinstein distance only for the degree distributions. This is because, in general, we want to work in the setting where $\mathbb{E}[(\mathcal{D}^+)^2]$ and $\mathbb{E}[(\mathcal{D}^-)^2]$ are infinite, which implies that $F^{+,+}$ and $F^{-,-}$ have infinite mean. However, depending on whether certain joint moments exist, we can use stronger versions of the assumption. We will specify such assumptions when needed.

Definition 2.6. Let $(\mathbf{D}_n^+, \mathbf{D}_n^+)$ be a bi-degree sequence and $\varepsilon, \eta > 0$. Then we write $(\mathbf{D}_n^+, \mathbf{D}_n^+) \in \mathcal{D}_{\eta, \varepsilon}$, if the bi-degree sequence satisfies Assumption 2.2 with these ε and η , for some probability mass functions f^α and $f^{\alpha, \beta}$. In addition we write $G_n \in \mathcal{G}_{\eta, \varepsilon}$ if G_n is a directed graph with bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^+) \in \mathcal{D}_{\eta, \varepsilon}$.

We often write $(\mathbf{D}_n^+, \mathbf{D}_n^+) \in \mathcal{D}_{\eta, \varepsilon}$, without specifying the constants or the probability mass functions. Here it is understood that there exist some $\eta, \varepsilon > 0$, f^α and $f^{\alpha, \beta}$ such that Assumption 2.2 is satisfied.

The term $n^{-\varepsilon}$ will give us the desired control over the concentration. For instance, since $|L_n - \mu n| \leq n^{1-\varepsilon}$ on the event Ω_n it follows that, for any $\delta < \varepsilon$,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{L_n}{n} - \mu\right| > n^{-\delta}\right) &\leq \mathbb{P}\left(\left|\frac{L_n}{n} - \mu\right| > n^{-\delta}, \Omega_n\right) + \mathbb{P}(\Omega_n^c) \\ &\leq O(n^{-\varepsilon+\delta}) + \mathbb{P}(\Omega_n^c), \end{aligned}$$

as $n \rightarrow \infty$.

Condition (2.21), in Assumption 2.2 states that if $\mathcal{D}^+$ and $\mathcal{D}^-$ are integer valued random variables with probability mass function f^+ and f^- , respectively, then $\mathbb{E}[(\mathcal{D}^+)^{1+\eta}]$, $\mathbb{E}[(\mathcal{D}^-)^{1+\eta}] < \infty$. This condition is there to allow for some room to pick appropriate scaling constants, which will be important in this thesis. For instance, in this setting we have that

$$\mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+ > n\}}] = n^{-\eta} \mathbb{E}[n^\eta \mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+ > n\}}] < n^{-\eta} \mathbb{E}[(\mathcal{D}^+)^{1+\eta}] = O(n^{-\eta}),$$

as $n \rightarrow \infty$.

There are of course many variations on Assumption 2.2. For instance, instead of considering $d_1(f_n^+, f^+)$ and $d_1(f_n^-, f^-)$, we could take $d_1(h_n, h)$ for some joint degree probability mass function h with

$$\begin{aligned} \sum_{k, \ell=0}^{\infty} k \ell h(k, \ell) &= \mu < \infty \quad \text{and} \\ \sum_{k, \ell=0}^{\infty} (k \ell)^{1+\eta} h(k, \ell) &< \infty. \end{aligned}$$

We stated Assumption 2.2 here since it will be the most general assumption used throughout this thesis.

Undirected graphs

The translation of Assumption 2.2 to the undirected case is straightforward.

Assumption 2.3. There exist probability mass functions f, f^* on the non-negative integers, that satisfy, for some $\eta > 0$,

$$\sum_{k=0}^{\infty} k f(k) = \nu < \infty \quad \text{and} \\ \sum_{k=0}^{\infty} k^{1+\eta} f(k) < \infty.$$

In addition, there exists an $\varepsilon > 0$ such that, if

$$\Omega_n = \{(d_1(f_n, f) \vee d_{\text{tv}}(f_n^*, f^*) \leq n^{-\varepsilon})\}, \quad (2.23)$$

we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Omega_n) = 1.$$

Similar to the directed case we write $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$, when the degree sequence satisfies Assumption 2.3 for some $\eta, \varepsilon > 0$ and certain probability mass functions f, f^* . When a graph G_n has a degree sequence $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$, we write $G_n \in \mathcal{G}_{\eta, \varepsilon}$.

Chapter 3

Generating degree sequences

In this chapter we will discuss algorithms for generating degree and bi-degree sequences that resemble i.i.d. samples from given random variables. We will show that, under appropriate conditions on the random variables from which we sample, such sequences satisfy the regularity assumptions that we have considered in the previous chapter. We start with the undirected case.

3.1 Undirected graphs

Let f be a probability mass function on the positive integers. A well-known algorithm for generating degree sequences with limit probability mass function f is by sampling the degrees i.i.d. from the distribution and then increase the last degree by 1 if the sum is not even. We will refer to this as the IID algorithm, see Algorithm 1. We will consider degrees that are sampled from random variables that satisfy the following:

Assumption 3.1. There exists an $0 < \eta \leq 1$ such that the random variable $\mathcal{D}$ satisfies,

$$\nu := \mathbb{E}[\mathcal{D}] < \infty, \quad \text{and} \quad \mathbb{E}[\mathcal{D}^{1+\eta}] < \infty.$$

Algorithm 1 Undirected IID Algorithm

Input: an integer-valued random variable $\mathcal{D}$ with distribution F that satisfies Assumption 3.1 and $n \geq 1$.
Sample $\mathcal{D}_n$ as an i.i.d. vector according to F
for $1 \leq i \leq n$ **do**
 Set $D_i = \mathcal{D}_i$.
end for
if $\sum_{i=1}^n D_i$ is not even **then**
 Set $D_n = D_n + 1$.
end if
Output: $\mathbf{D}_n$.

We write $\mathbf{D}_n = \text{IID}(\mathcal{D})$, when $\mathbf{D}_n$ is generated by the IID algorithm with $\mathcal{D}$ and n as input.

The only difference between the sequence of i.i.d. copies $\mathbf{D}_n$ and the eventual degree sequence $\mathbf{D}_n$ is in the degree D_n which satisfies

$$D_n = \mathcal{D}_n + \mathbb{1}_{\{\sum_{i=1}^n D_i \text{ is odd}\}}.$$

Since this additional indicator make almost no impact on the degree sequence we will neglect it in the analysis of degree sequences.

Remark 3.1. We will regard a degree sequence $\mathbf{D}_n = \text{IID}(\mathcal{D})$ as a sequences of i.i.d. copies of $\mathcal{D}$.

The following theorem states that when we generate a degree sequence with the IID algorithm, by sampling from a random variable that satisfies Assumption (3.1), then the degree sequence satisfies Assumption 2.3:

Theorem 3.1. *Let $\mathcal{D}$ be an integer-valued random variable with probability mass function f that satisfies Assumption (3.1) and define the size-biased probability mass function of f by*

$$f^*(k) = \frac{\mathbb{E}[\mathcal{D} \mathbb{1}_{\{\mathcal{D}=k\}}]}{\nu}.$$

Then $\text{IID}(\mathcal{D}) \in \mathcal{D}_{\eta, \varepsilon}(f, f^)$ for any $\varepsilon \leq \eta/(8 + 4\eta)$. Moreover, if we let Ω_n be defined as in (2.23), then*

$$\mathbb{P}(\Omega_n) \geq 1 - O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$.

Proof. Fix $\varepsilon \leq \eta/(8 + 4\eta)$, define the events

$$A_n = \left\{ \sum_{k=0}^{\infty} \left| \sum_{t=0}^k f_n(t) - f(t) \right| \leq n^{-\eta/(2+2\eta)} \right\},$$

$$B_n = \left\{ \sum_{k=0}^{\infty} |f_n^*(k) - f^*(k)| \leq n^{-\varepsilon} \right\},$$

and notice that $\mathbb{P}(\Omega_n^c) \leq \mathbb{P}(A_n^c) + \mathbb{P}(A_n \cap B_n^c)$. For the first term we have, using Markov's inequality,

$$\mathbb{P}(A_n^c) \leq n^{\frac{\eta}{2+2\eta}} \mathbb{E}[d_1(f_n, f)] \leq O\left(n^{-\frac{\eta}{2+2\eta}}\right) = O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$, where the second inequality follows from [29, Proposition 4.2] and the last since $-\eta/(2 + 2\eta) < -\eta/(8 + 4\eta)$. Hence, we need to show that, as $n \rightarrow \infty$

$$\mathbb{P}(A_n \cap B_n^c) \leq O(n^{-\varepsilon}).$$

For this we compute that,

$$\sum_{k=0}^{\infty} |f_n^*(k) - f^*(k)| = \sum_{k=0}^{\infty} \left| \frac{1}{L_n} \sum_{i=1}^n D_i \mathbb{1}_{\{D_i=k\}} - \frac{\mathbb{E}[D \mathbb{1}_{\{D=k\}}]}{\nu} \right|$$

$$\begin{aligned}
&\leq \sum_{k=0}^{\infty} \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| \sum_{i=1}^n D_i \mathbb{1}_{\{D_i=k\}} \\
&\quad + \sum_{k=0}^{\infty} \frac{1}{\nu n} \left| \sum_{i=1}^n D_i \mathbb{1}_{\{D_i=k\}} - \mathbb{E} [D \mathbb{1}_{\{D=k\}}] \right| \\
&= \frac{|L_n - \nu n|}{\nu n} + \sum_{k=0}^{\infty} \frac{1}{\mu n} \left| \sum_{i=1}^n X_{ik} \right|,
\end{aligned}$$

where we defined $X_{ik} = D_i \mathbb{1}_{\{D_i=k\}} - \mathbb{E} [D \mathbb{1}_{\{D=k\}}]$. Now observe that conditioned on A_n we have that

$$\begin{aligned}
\frac{|L_n - \nu n|}{\nu n^{1-\varepsilon}} &\leq \frac{n^{1-\eta/(2+2\eta)}}{\nu n^{1-\varepsilon}} \leq \nu^{-1} n^{\varepsilon-\eta/2(1+\eta)} \\
&\leq \nu^{-1} n^{-\eta/4(1+\eta)} \leq \nu^{-1} n^{-\eta/4(1+2\eta)},
\end{aligned}$$

so that

$$\mathbb{P}(A_n \cap B_n^c) \leq O(n^{-\varepsilon}) + \mathbb{P}\left(\sum_{k=0}^{\infty} \frac{1}{\nu n} \left| \sum_{i=1}^n X_{ik} \right| > n^{-\varepsilon}\right),$$

as $n \rightarrow \infty$.

To analyze the last probability take $a_n = \lfloor n^{2\varepsilon/\eta} \rfloor$. Then,

$$\begin{aligned}
\mathbb{P}\left(\sum_{k=0}^{\infty} \frac{1}{\nu n} \left| \sum_{i=1}^n X_{ik} \right| > n^{-\varepsilon}\right) &\leq \frac{1}{\nu n^{1-\varepsilon}} \sum_{k=0}^{a_n} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik} \right|^2 \right]^{1/2} \\
&\quad + \frac{1}{\nu n^{1-\varepsilon}} \sum_{k=a_n+1}^{\infty} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik} \right| \right] \\
&\leq \frac{1}{\nu n^{1/2-\varepsilon}} \sum_{k=0}^{a_n} \text{Var}(X_{1k})^{1/2} + \frac{n^{\varepsilon}}{\nu} \sum_{k=a_n+1}^{\infty} \mathbb{E} [|X_{1k}|] \\
&\leq \frac{1}{\nu n^{1/2-\varepsilon}} \sum_{k=0}^{a_n} k + \frac{2n^{\varepsilon}}{\nu} \sum_{k=a_n+1}^{\infty} \mathbb{E} [D \mathbb{1}_{\{D=k\}}] \\
&\leq \frac{a_n(a_n+1)}{2\nu n^{1/2-\varepsilon}} + \frac{2n^{\varepsilon}}{\nu} \mathbb{E} [D \mathbb{1}_{\{D>a_n\}}] \\
&\leq \frac{a_n(a_n+1)}{2\nu n^{1/2-\varepsilon}} + 2n^{\varepsilon} a_n^{-\eta} \\
&= O(n^{-\varepsilon}),
\end{aligned}$$

as $n \rightarrow \infty$. Here, for the last line, we have used that

$$\frac{4\varepsilon}{\eta} - \frac{1}{2} + \varepsilon \leq \frac{1}{2+\eta} - \frac{1}{2} + \frac{\eta}{8+4\eta} = -\frac{\eta}{8+4\eta} \leq -\varepsilon, \quad (3.1)$$

so that

$$n^{\varepsilon-\frac{1}{2}} a_n^2 = O\left(n^{\frac{4\varepsilon}{\eta}-\frac{1}{2}+\varepsilon}\right) = O(n^{-\varepsilon}), \quad (3.2)$$

as $n \rightarrow \infty$. □

3.2 Directed graphs

Extending the IID algorithm to the directed setting is not completely straightforward. The difficulty lies in the requirement that $\sum_{i=1}^n D_i^+ = \sum_{i=1}^n D_i^-$, which is more involved than the condition that $\sum_{i=1}^n D_i$ is even, which we needed for degree sequences. We will describe the *directed* IID algorithm, which was introduced in the setting of regularly-varying random variables in [28].

3.2.1 The directed IID algorithm

Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be two integer-valued random variables with, respectively, distribution functions F^+ and F^- and joint distribution H . The idea of the *directed* IID is to first take i.i.d. samples $(\mathcal{D}_i^+, \mathcal{D}_i^-)$ from $(\mathcal{D}^+, \mathcal{D}^-)$. Then, if the difference $|\sum_{i=1}^n \mathcal{D}_i^+ - \mathcal{D}_i^-|$, is small enough, we randomly select this number of nodes and increase their out- or in-degree, depending on which sum is the smallest, by one. For a detailed description see Algorithm 2.

The important part of the algorithm is to define when the difference between the i.i.d. sums is *small enough* and ensure that this happens often enough. For this we use the following condition on the random variables $\mathcal{D}^+$ and $\mathcal{D}^-$.

Assumption 3.2. There exists an $0 < \eta \leq 1$ such that the random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy

$$\nu := \mathbb{E}[\mathcal{D}^+] = \mathbb{E}[\mathcal{D}^-] < \infty, \quad \mu := \mathbb{E}[\mathcal{D}^+ \mathcal{D}^-] < \infty, \quad (3.3)$$

$$\mathbb{E}[(\mathcal{D}^+)^{1+\eta}] < \infty, \quad \mathbb{E}[(\mathcal{D}^-)^{1+\eta}] < \infty, \quad (3.4)$$

Algorithm 2 Directed IID Algorithm

Input: two integer-valued random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ with joint distribution H , that satisfy Assumption 3.2 or 3.4 and $n \geq 1$.

Fix $0 < \delta < \eta/(1 + \eta)$.

repeat

 Sample $(\mathcal{D}_n^+, \mathcal{D}_n^-)$ as i.i.d. vectors distributed according to H .

 Set $\Delta_n = \sum_{i=1}^n (\mathcal{D}_i^+ - \mathcal{D}_i^-)$.

until $|\Delta_n| \leq n^{1-\delta}$

Select $|\Delta_n|$ indices from $\{1, 2, \dots, n\}$ uniformly at random (without replacement) and let $\mathcal{I}_n$ denote the set of these indices.

for $1 \leq i \leq n$ **do**

 Set $D_i^+ = \mathcal{D}_i^+$ and $D_i^- = \mathcal{D}_i^-$.

if $i \in \mathcal{I}_n$ and $\Delta_n > 0$ **then**

 Set $D_i^+ = D_i^+ + 1$.

end if

if $i \in \mathcal{I}_n$ and $\Delta_n < 0$ **then**

 Set $D_i^- = D_i^- + 1$.

end if

end for

Output: $(\mathbf{D}_n^+, \mathbf{D}_n^-)$.

Similar to the undirected case, we write $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$ if the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ was generated by the directed IID algorithm with $\mathcal{D}^+$, $\mathcal{D}^-$ and n as input. Moreover, from now on we use the term IID algorithm for both the undirected and directed case, since it will be clear from the setting we are working in which of the two we mean.

Algorithm 2 was first introduced in [28] for the special case where $\mathcal{D}^+$ and $\mathcal{D}^-$ are independent. Here it was shown that this algorithm generates bi-degree sequences where the degrees of nodes are approximately i.i.d. with the prescribed marginal distributions. It is not hard to see that the results in [28] remain valid even if we allow $\mathcal{D}^+$ and $\mathcal{D}^-$ to be dependent. We also point out that the IID algorithm only requires the moment conditions (3.4), and therefore can be used to generate any light-tailed degree sequence as well as the vast majority of scale-free (heavy-tailed) degree distributions.

There are some important observation to be made. First, it follows from the Kolmogorov-Marcinkiewicz-Zygmund strong law of large number and both (3.3) and (3.4), that

$$\frac{\sum_{i=1}^n \mathcal{D}_i^+ - \mathcal{D}_i^-}{n^{1/p}} = \frac{\sum_{i=1}^n \mathcal{D}_i^+ - n\nu}{n^{1/p}} - \frac{\sum_{i=1}^n \mathcal{D}_i^- - n\nu}{n^{1/p}} \rightarrow 0, \quad \text{a.s.},$$

for all $0 < p \leq 1 + \eta$. In particular, when $0 < \delta \leq \eta/(1 + \eta)$, this implies, with $p = 1/(1 - \delta)$, that

$$\frac{|\Delta_n|}{n^{1-\delta}} = \frac{|\sum_{i=1}^n \mathcal{D}_i^+ - \mathcal{D}_i^-|}{n^{1-\delta}} \xrightarrow{\mathbb{P}} 0, \quad (3.5)$$

as $n \rightarrow \infty$. Hence if $\Psi_n = \{|\Delta_n| \leq n^{1-\delta}\}$ denotes the event that the condition in the repeat-until loop is satisfied, then $\mathbb{P}(\Psi_n) \rightarrow 1$ as $n \rightarrow \infty$.

Secondly, if we define the indicators

$$\xi_i^+ = \mathbb{1}_{\{\Delta_n \leq 0 \text{ and } i \in \mathcal{I}_n\}} \quad \text{and} \quad \xi_i^- = \mathbb{1}_{\{\Delta_n > 0 \text{ and } i \in \mathcal{I}_n\}}, \quad (3.6)$$

then

$$D_i^+ = \mathcal{D}_i^+ + \xi_i^+ \quad \text{and} \quad D_i^- = \mathcal{D}_i^- + \xi_i^-.$$

Note that only one of ξ_i^+ and ξ_i^- can be non-zero for any $1 \leq i \leq n$, while the number of nodes for which this holds equals $|\Delta_n|$. This is summarized in the following two identities:

$$\sum_{i=1}^n \xi_i^+ + \xi_i^- = |\Delta_n| \quad \text{and} \quad \sum_{i,j=1}^n \xi_i^+ \xi_j^- = 0. \quad (3.7)$$

3.2.2 Asymptotic results for bi-degree sequences generated by the IID algorithm

We will first show that bi-degree sequences generated by the IID algorithm satisfy Assumption 2.2.

Theorem 3.2. *Suppose that $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.2 and let $f^\pm$ denote the probability mass function of $\mathcal{D}^\pm$ and define, for all combinations $\alpha, \beta \in \{+, -\}$,*

$$f^{\alpha, \beta}(k) = \frac{\mathbb{E}[\mathcal{D}^\alpha \mathbb{1}_{\{D^\beta = k\}}]}{\nu}.$$

Then, for any $\varepsilon < \eta/(8 + 4\eta)$,

$$\text{IID}(\mathcal{D}^+, \mathcal{D}^-) \in \mathcal{D}_{\eta, \varepsilon},$$

with limiting probability mass functions f^α and $f^{\alpha, \beta}$. Moreover,

$$\mathbb{P}(\Omega_n) \geq 1 - O\left(n^{-\varepsilon} + n^{-\frac{\eta}{1+\eta} + \varepsilon}\right),$$

as $n \rightarrow \infty$.

We will first prove the following useful lemma:

Lemma 3.3. Suppose that $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.2 and let $(\mathcal{D}_n^+, \mathcal{D}_n^-)$ be a vector of i.i.d. copies of $(\mathcal{D}^+, \mathcal{D}^-)$. In addition, take $\alpha, \beta \in \{+, -\}$ and define

$$f^{\alpha, \beta}(k) = \frac{\mathbb{E}[\mathcal{D}^\alpha \mathbb{1}_{\{\mathcal{D}^\beta = k\}}]}{\nu} \quad \text{and} \\ \tilde{f}_n^{\alpha, \beta}(k) = \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^\alpha \mathbb{1}_{\{\mathcal{D}_i^\beta = k\}}.$$

Then, for any $0 < \varepsilon \leq \eta/(8 + 4\eta)$,

$$\mathbb{P}\left(d_{\text{tv}}(\tilde{f}_n^{\alpha, \beta}, f^{\alpha, \beta}) \leq n^{-\varepsilon}\right) \geq 1 - O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$.

Proof. We will only treat the mixed case $\alpha = +, \beta = -$ and the joint case $\alpha = + = \beta$. The other two cases will follow by similar arguments.

$\alpha = +, \beta = -$

Define $X_{ik\ell} = \mathcal{D}_i^+ \mathbb{1}_{\{\mathcal{D}_i^+ = k\}} \mathbb{1}_{\{\mathcal{D}_i^- = \ell\}} - \mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^- = k\}}]$ and let $a_n = \lfloor n^{2\varepsilon/\eta} \rfloor$. Then

$$\begin{aligned} \mathbb{P}\left(d_{\text{tv}}(\tilde{f}_n^{+, -}, f^{+, -}) \leq n^{-\varepsilon}\right) &= \mathbb{P}\left(\sum_{k, \ell=0}^{\infty} \frac{1}{\nu n} \left| \sum_{i=1}^n X_{ik\ell} \right| \leq 2n^{-\varepsilon}\right) \\ &\leq \frac{1}{2\nu n^{1-\varepsilon}} \sum_{k=0}^{a_n} \sum_{\ell=0}^{a_n} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik\ell} \right|^2 \right]^{1/2} \end{aligned} \quad (3.8)$$

$$+ \frac{1}{2\nu n^{1-\varepsilon}} \sum_{k > a_n} \sum_{\ell=0}^{\infty} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik\ell} \right|^2 \right] \quad (3.9)$$

$$+ \frac{1}{2\nu n^{1-\varepsilon}} \sum_{k=0}^{a_n} \sum_{\ell > a_n} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik\ell} \right|^2 \right]. \quad (3.10)$$

We will proceed by showing that each of these three terms is $O(n^{-\varepsilon})$.

For (3.8) we use that

$$\text{Var}(X_{1k\ell}) = \mathbb{E}[(\mathcal{D}^+)^2 \mathbb{1}_{\{\mathcal{D}^+ = k\}} \mathbb{1}_{\{\mathcal{D}^- = \ell\}}] = k^2 h^{+, -}(k, \ell),$$

to obtain

$$\begin{aligned}
\frac{1}{2\nu n^{1-\varepsilon}} \sum_{k=0}^{a_n} \sum_{\ell=0}^{a_n} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik\ell} \right|^2 \right]^{1/2} &= \frac{1}{2\nu n^{\frac{1}{2}-\varepsilon}} \sum_{k=0}^{a_n} \sum_{\ell=0}^{a_n} \text{Var}(X_{1k\ell})^{1/2} \\
&= \frac{1}{2\nu n^{\frac{1}{2}-\varepsilon}} \sum_{k=0}^{a_n} \sum_{\ell=0}^{a_n} k h^{+,-}(k, \ell)^{1/2} \\
&\leq \frac{1}{2\nu n^{\frac{1}{2}-\varepsilon}} \sum_{k=0}^{a_n} k \\
&= \frac{a_n(a_n + 1)}{2\nu n^{\frac{1}{2}-\varepsilon}} \\
&\leq O\left(n^{\varepsilon-\frac{1}{2}} a_n^2\right) \\
&= O\left(n^{\frac{4\varepsilon}{\eta}-\frac{1}{2}+\varepsilon}\right) = O\left(n^{-\varepsilon}\right),
\end{aligned}$$

as $n \rightarrow \infty$, where for the last line we have used (3.1) and (3.2) from the proof of Theorem 3.1.

Next, for (3.9), we use the bound

$$\mathbb{E} \left[\left| \sum_{i=1}^n X_{ik\ell} \right| \right] \leq n \mathbb{E}[|X_{ik\ell}|] \leq 2\mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+=k\}} \mathbb{1}_{\{\mathcal{D}^-=\ell\}}], \quad (3.11)$$

so that, as $n \rightarrow \infty$,

$$\begin{aligned}
\frac{1}{2\nu n^{1-\varepsilon}} \sum_{k>a_n} \sum_{\ell=0}^{\infty} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik\ell}^{\alpha,\beta} \right| \right] &\leq \frac{1}{\nu n^{-\varepsilon}} \sum_{k>a_n} \sum_{\ell=0}^{\infty} \mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+=k\}} \mathbb{1}_{\{\mathcal{D}^-=\ell\}}] \\
&\leq \frac{1}{\nu n^{-\varepsilon}} \sum_{k>a_n} \mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+=k\}}] \\
&\leq \frac{1}{\nu n^{-\varepsilon}} \mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+>a_n\}}] \\
&\leq n^\varepsilon a_n^{-\eta} \frac{\mathbb{E}[(\mathcal{D}^+)^{1+\eta}]}{\nu} = O(n^{-\varepsilon}).
\end{aligned}$$

Finally, for (3.10) we get, using (3.11),

$$\begin{aligned}
\frac{1}{2\nu n^{1-\varepsilon}} \sum_{k=0}^{a_n} \sum_{\ell>a_n} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik\ell} \right| \right] &\leq \frac{1}{\nu n^{-\varepsilon}} \sum_{k=0}^{a_n} \sum_{\ell>a_n} \mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+=k\}} \mathbb{1}_{\{\mathcal{D}^-=\ell\}}] \\
&\leq \frac{n^\varepsilon a_n}{\nu} \mathbb{E}[\mathbb{1}_{\{\mathcal{D}^->a_n\}}] \\
&\leq n^\varepsilon a_n^{-\eta} \frac{\mathbb{E}[(\mathcal{D}^-)^{1+\eta}]}{\nu} = O(n^{-\varepsilon}),
\end{aligned}$$

as $n \rightarrow \infty$. This completes the proof for $\alpha = +$, $\beta = -$.

$\alpha = + = \beta$.

Let $a_n = \lfloor n^{2\varepsilon/\eta} \rfloor$ and define $X_{ik} = \mathcal{D}_i^+ \mathbb{1}_{\{\mathcal{D}_i^+=k\}} - \mathbb{E}[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+=k\}}]$. Then, since

$$\text{Var}(X_{1k}) = \mathbb{E}[(\mathcal{D}^+)^2 \mathbb{1}_{\{\mathcal{D}^+=k\}}] = k^2 f^+(k),$$

we can use the same arguments as for the previous case to derive that

$$\begin{aligned}
& \mathbb{P}\left(d_{\text{tv}}(\tilde{f}_n^{+,-}, f^{+,-}) \leq n^{-\varepsilon}\right) \\
&= \mathbb{P}\left(\sum_{k=0}^{\infty} \frac{1}{\nu n} \left|\sum_{i=1}^n X_{ik}\right| \leq 2n^{-\varepsilon}\right) \\
&\leq \frac{1}{2\nu n^{1-\varepsilon}} \sum_{k=0}^{a_n} \mathbb{E}\left[\left|\sum_{i=1}^n X_{ik}\right|^2\right]^{1/2} + \frac{1}{2\nu n^{1-\varepsilon}} \sum_{k>a_n} \mathbb{E}\left[\left|\sum_{i=1}^n X_{ik}\right|\right] \\
&\leq \frac{1}{2\nu n^{1/2-\varepsilon}} \sum_{k=0}^{a_n} \text{Var}(X_{1k})^{1/2} + \frac{1}{\nu n^{-\varepsilon}} \mathbb{E}\left[\mathcal{D}^+ \mathbb{1}_{\{\mathcal{D}^+ > a_n\}}\right] \\
&\leq \frac{a_n(a_n+1)}{4\nu n^{1/2-\varepsilon}} + n^\varepsilon a_n^{-\eta} \frac{\mathbb{E}\left[(\mathcal{D}^-)^{1+\eta}\right]}{\nu} = O(n^{-\varepsilon}),
\end{aligned}$$

as $n \rightarrow \infty$. □

With this result we are ready to prove Theorem 3.2:

Proof of Theorem 3.2. Fix $0 < \varepsilon < \eta/(8+4\eta)$ and define the events

$$\begin{aligned}
\Psi_n &= \left\{|\Delta_n| \leq n^{\frac{1}{1+\eta}}\right\}, \\
A_n &= \left\{\max\{d_1(f_n^+, f^+), d_1(f_n^-, f^-)\} \leq n^{-\varepsilon}\right\} \quad \text{and} \\
B_n &= \left\{\max_{\alpha, \beta \in \{+, -\}} \{d_{\text{tv}}(f_n^{\alpha, \beta}, f^{\alpha, \beta})\} \leq n^{-\varepsilon}\right\}.
\end{aligned}$$

Then by (3.5), with $\delta = \eta/(1+\eta)$, we have that $\mathbb{P}(\Psi_n) \rightarrow 1$ as $n \rightarrow \infty$. Furthermore, $\mathbb{P}(\Omega_n^c) \leq \mathbb{P}(A_n^c) + \mathbb{P}(B_n^c \cap A_n)$ and we will show that, $\mathbb{P}(A_n^c) = O(n^{-\eta(1+\eta)+\varepsilon})$ and $\mathbb{P}(B_n^c \cap A_n) = O(n^{-\varepsilon} + n^{-\eta/(1+\eta)+\varepsilon})$, as $n \rightarrow \infty$.

To this end we define, for $\alpha, \beta \in \{+, -\}$, the empirical distributions

$$\begin{aligned}
\tilde{f}_n^\alpha(k) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\mathcal{D}_i^\alpha = k\}}, \\
\tilde{f}_n^{\alpha, \beta}(k) &= \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^\alpha \mathbb{1}_{\{\mathcal{D}_i^\beta = k\}}.
\end{aligned}$$

We will now prove the claimed two bounds:

$$\mathbb{P}(A_n^c) \leq O(n^{-\eta/(1+\eta)+\varepsilon})$$

It follows from the triangle's inequality and [34, Theorem 2.2] (see [29, Proposition 4.2]), that

$$\begin{aligned}
\mathbb{E}[d_1(f_n^\alpha, f^\alpha)] &\leq \mathbb{E}\left[d_1(f_n^\alpha, \tilde{f}_n^\alpha) \mid \Psi_n\right] + \mathbb{E}\left[d_1(\tilde{f}_n^\alpha, f^\alpha) \mid \Psi_n\right] \\
&\leq \frac{1}{\mathbb{P}(\Psi_n)} \mathbb{E}\left[d_1(f_n^\alpha, \tilde{f}_n^\alpha) \mathbb{1}_{\Psi_n}\right] + O(n^{-\delta}),
\end{aligned}$$

as $n \rightarrow \infty$. Moreover, on the event Ψ_n ,

$$\begin{aligned} d_1(f_n^\alpha, \tilde{f}_n^\alpha) &= \frac{1}{n} \sum_{k=0}^{\infty} \left| \sum_{i=1}^n \mathbb{1}_{\{\mathcal{D}_i^\alpha + \xi_i^\alpha \leq k\}} - \mathbb{1}_{\{\mathcal{D}_i^\alpha \leq k\}} \right| \\ &= \frac{1}{n} \sum_{k=0}^{\infty} \sum_{i=1}^n \mathbb{1}_{\{\mathcal{D}_i^\alpha < k \leq \mathcal{D}_i^\alpha + \xi_i^\alpha\}} \\ &= \frac{1}{n} \sum_{i=1}^n \xi_i^\alpha \leq \frac{|\Delta_n|}{n} \leq n^{-\delta}. \end{aligned}$$

Therefore we obtain that

$$\mathbb{E}[d_1(f_n^\alpha, f^\alpha)] = O\left(n^{-\frac{\eta}{1+\eta}}\right), \quad (3.12)$$

and hence, by Markov's inequality,

$$\mathbb{P}(A_n^c) \leq n^\varepsilon (\mathbb{E}[d_1(f_n^+, f^+)] + \mathbb{E}[d_1(f_n^-, f^-)]) = O\left(n^{-\frac{\eta}{1+\eta} + \varepsilon}\right),$$

as $n \rightarrow \infty$.

$$\mathbb{P}(B_n^c \cap A_n) \leq O(n^{-\varepsilon} + n^{-\eta/(1+\eta) + \varepsilon})$$

Fix $\alpha, \beta \in \{+, -\}$. We will prove that, as $n \rightarrow \infty$,

$$\mathbb{P}(d_{\text{tv}}(f_n^{\alpha, \beta}, f^{\alpha, \beta}) \leq n^{-\varepsilon}, A_n) \leq O(n^{-\varepsilon} + n^{-\delta + \varepsilon}).$$

Since

$$\mathbb{P}(B_n^c \cap A_n) \leq \mathbb{P}(d_{\text{tv}}(f_n^{+, -}, f^{+, -}) \leq n^{-\varepsilon}, A_n) + \mathbb{P}(d_{\text{tv}}(f_n^{-, +}, f^{-, +}) \leq n^{-\varepsilon}, A_n),$$

the result then follows.

Similar to the case for $d_1(f_n^\alpha, f^\alpha)$ we split the probability

$$\mathbb{P}(d_{\text{tv}}(f_n^{\alpha, \beta}, f^{\alpha, \beta}) \leq n^{-\varepsilon}, A_n) \leq \mathbb{P}\left(d_{\text{tv}}(f_n^{\alpha, \beta}, \tilde{f}_n^{\alpha, \beta}) \leq \frac{n^{-\varepsilon}}{2}\right) \quad (3.13)$$

$$+ \mathbb{P}\left(d_{\text{tv}}(\tilde{f}_n^{\alpha, \beta}, f^{\alpha, \beta}) \leq \frac{n^{-\varepsilon}}{2}, A_n\right). \quad (3.14)$$

From Lemma 3.3 it follows that (3.14) is $O(n^{-\varepsilon})$. For the other probability we first recall that $\nu_n = L_n/n$, so that on the event A_n ,

$$\begin{aligned} 2d_{\text{tv}}(f_n^{\alpha, \beta}, \tilde{f}_n^{\alpha, \beta}) &= \sum_{k=0}^{\infty} \left| \frac{1}{L_n} \sum_{i=1}^n D_i^\alpha \mathbb{1}_{\{D_i^\beta = k\}} - \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^\alpha \mathbb{1}_{\{\mathcal{D}_i^\beta = k\}} \right| \\ &\leq \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| L_n + \frac{1}{\nu n} \sum_{k=0}^{\infty} \left| \sum_{i=1}^n D_i^\alpha \mathbb{1}_{\{D_i^\beta = k\}} - \mathcal{D}_i^\alpha \mathbb{1}_{\{\mathcal{D}_i^\beta = k\}} \right| \\ &\leq \frac{|L_n - \nu n|}{\nu n} + \frac{1}{\nu n} \sum_{k=0}^{\infty} \sum_{i=1}^n \xi_i^\alpha \mathbb{1}_{\{D_i^\beta = k\}} \\ &\quad + \frac{1}{\nu n} \sum_{k=0}^{\infty} \sum_{i=1}^n \mathcal{D}_i^\alpha \left| \mathbb{1}_{\{\mathcal{D}_i^\beta + \xi_i^\beta = k\}} - \mathbb{1}_{\{\mathcal{D}_i^\beta = k\}} \right| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{|\nu_n - \nu|}{\nu} + \frac{|\Delta_n|}{\nu n} + \frac{2}{\nu n} \sum_{i=1}^n \mathcal{D}_i^\alpha \mathbb{1}_{\{\xi_i^\beta=1\}} \\
&\leq \frac{d_1(f_n^\alpha, f^\alpha)}{\nu} + \frac{|\Delta_n|}{\nu n} + \frac{2}{\nu n} \sum_{i=1}^n \mathcal{D}_i^\alpha \xi_i^\beta.
\end{aligned}$$

Therefore we have that

$$\begin{aligned}
&\mathbb{P} \left(d_{\text{tv}}(f_n^{\alpha,\beta}, \tilde{f}_n^{\alpha,\beta}) \leq \frac{n^{-\varepsilon}}{2}, A_n \right) \\
&\leq n^\varepsilon \left(\frac{\mathbb{E}[d_1(f_n^\alpha, f^\alpha)]}{\nu} + \frac{\mathbb{E}[|\Delta_n|]}{\nu n} + \frac{2\mathbb{E}[\mathcal{D}_1^\alpha \xi_1^\beta \mathbb{1}_{A_n}]}{\nu} \right) \\
&\leq \frac{n^\varepsilon}{\mathbb{P}(\Psi_n)} \left(\frac{\mathbb{E}[d_1(f_n^\alpha, f^\alpha)]}{\nu} + \frac{\mathbb{E}[|\Delta_n| \mathbb{1}_{\Psi_n}]}{\nu n} + \frac{2\mathbb{E}[\mathcal{D}_1^\alpha \xi_1^\beta \mathbb{1}_{A_n \cap \Psi_n}]}{\nu} \right).
\end{aligned}$$

For the last term we note that

$$\begin{aligned}
\mathbb{E}[\mathcal{D}_1^\alpha \xi_1^\beta \mathbb{1}_{A_n \cap \Psi_n}] &= \mathbb{E}[\mathcal{D}_1^\beta \mathbb{E}[\xi_1^\alpha | (\mathcal{D}_n^+, \mathcal{D}_n^-)] \mathbb{1}_{A_n \cap \Psi_n}] \\
&\leq \mathbb{E}\left[\frac{|\Delta_n|}{L_n} \mathcal{D}_1^\alpha \mathbb{1}_{A_n \cap \Psi_n}\right] \\
&\leq \frac{\nu n^{-\frac{\eta}{1+\eta}}}{\nu - n^{-\varepsilon}}.
\end{aligned} \tag{3.15}$$

Hence, using (3.12), we obtain

$$\begin{aligned}
&\frac{n^\varepsilon}{\mathbb{P}(\Psi_n)} \left(\frac{\mathbb{E}[d_1(f_n^\alpha, f^\alpha)]}{\nu} + \frac{\mathbb{E}[|\Delta_n| \mathbb{1}_{\Psi_n}]}{\nu n} + \frac{2\mathbb{E}[\mathcal{D}_1^\alpha \xi_1^\beta \mathbb{1}_{A_n \cap \Psi_n}]}{\nu} \right) \\
&\leq \frac{n^\varepsilon}{\mathbb{P}(\Psi_n)} \left(\frac{\mathbb{E}[d_1(f_n^\alpha, f^\alpha)]}{\nu} + \frac{n^{-\frac{\eta}{1+\eta}}}{\nu} + \frac{2n^{-\frac{\eta}{1+\eta}}}{\nu - n^{-\varepsilon}} \right) \\
&= O\left(n^{-\frac{\eta}{1+\eta} + \varepsilon}\right),
\end{aligned}$$

as $n \rightarrow \infty$, which implies that

$$\mathbb{P}(d_{\text{tv}}(f_n^{\alpha,\beta}, f^{\alpha,\beta}) \leq n^{-\varepsilon}, A_n) \leq O\left(n^{-\varepsilon} + n^{-\frac{\eta}{1+\eta} + \varepsilon}\right),$$

as $n \rightarrow \infty$. □

3.2.3 IID algorithm with finite covariance of out- and in-degrees

In this section we illustrate the comment made at the end of Chapter 2, that when certain joint moments exist we can prove stronger versions of Assumption 2.2. We consider such an assumption, which we later use in Chapter 10.

Assumption 3.3. There exist, for all $\alpha, \beta \in \{+, -\}$, probability mass functions f^α , $f^{\alpha, \beta}$ and h_n on the non-negative integers, that satisfy, for some $\eta > 0$,

$$\nu := \sum_{k=0}^{\infty} k f^+(k) = \sum_{k=0}^{\infty} k f^-(k) < \infty, \quad (3.16)$$

$$\sum_{k=0}^{\infty} k^{1+\eta} f^+(k) < \infty, \quad \sum_{k=0}^{\infty} k^{1+\eta} f^-(k) < \infty \quad \text{and} \quad (3.17)$$

$$\mu_n^* := \sum_{k, \ell=0}^{\infty} (k\ell)^{1+\eta/2} h_n(k, \ell) < \infty. \quad (3.18)$$

In addition there exists an $\varepsilon > 0$ such that, if

$$\begin{aligned} \Xi_n^{(1)} &= \left\{ \left(\max_{\alpha \in \{+, -\}} d_1(f_n^\alpha, f^\alpha) \right) \vee \left(\max_{\alpha \in \{+, -\}} d_{\text{tv}}(f_n^{\alpha, \alpha}, f^{\alpha, \alpha}) \right) \leq n^{-\varepsilon} \right\}, \\ \Xi_n^{(2)} &= \left\{ \max\{d_1(f_n^{-, +}, f^{-, +}), d_1(f_n^{+, -}, f^{+, -})\} \leq n^{-\varepsilon} \right\}, \\ \Xi_n^{(3)} &= \left\{ \left| \sum_{k, \ell=0}^{\infty} (k\ell)^{1+\eta/2} h_n(k, \ell) - \mu_n^* \right| \leq n^{-\varepsilon} \right\} \quad \text{and} \\ \Xi_n &= \Xi_n^{(1)} \cap \Xi_n^{(2)} \cap \Xi_n^{(3)}. \end{aligned}$$

then

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Xi_n) = 1.$$

Comparing Assumption 3.3 to Assumption 2.2 we observe that we require convergence of $f_n^{-, +}$ and $f_n^{+, -}$ with respect to the Kantorovich-Rubenstein distance, instead of the total variation distance. Moreover, we require the convergence of the empirical joint $1 + \eta$ moment of the in out- and in-degrees of a node, which is not necessarily true under the weaker Assumption 3.2.

It will come as no surprise that, in order for the bi-degree sequences $\text{IID}(\mathcal{D}^+, \mathcal{D}^-)$ to satisfy Assumption 3.3, we at least need that $\mathbb{E}[(\mathcal{D}^+ \mathcal{D}^-)^{1+\eta}] < \infty$. We therefore define the following set of conditions.

Assumption 3.4. There exists an $0 < \eta \leq 1$ such that random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy (3.3) and

$$\mathbb{E}[(\mathcal{D}^+)^{1+\eta}(\mathcal{D}^-)^{1+\eta}] < \infty. \quad (3.19)$$

Since we have

$$\mathbb{E}[|\mathcal{D}^+ - \mathcal{D}^-|^{1+\eta}] \leq \mathbb{E}[(\mathcal{D}^+ \mathcal{D}^-)^{1+\eta}] \leq 2^{1+\eta} \mathbb{E}[(\mathcal{D}^+)^{1+\eta}(\mathcal{D}^-)^{1+\eta}] < \infty$$

it again follows from the Kolmogorov-Marcinkiewicz-Zygmund strong law of large number that when $0 < \delta \leq \eta/(1 + \eta)$, similar to Assumption 3.2,

$$\frac{|\Delta_n|}{n^{1-\delta}} = \frac{|\sum_{i=1}^n \mathcal{D}_i^+ - \mathcal{D}_i^-|}{n^{1-\delta}} \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$. Hence if we define $\Psi_n = \{|\Delta_n| \leq n^{\frac{1}{1+\eta}}\}$, then $\mathbb{P}(\Psi_n) \rightarrow 1$, as $n \rightarrow \infty$. We now have the following result:

Theorem 3.4. *Let $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.4. Denote by $f^\pm$ the probability mass function of $\mathcal{D}^\pm$ and define, for all combinations $\alpha, \beta \in \{+, -\}$, the probability mass function functions*

$$f^{\alpha, \beta}(k) = \frac{\mathbb{E} [\mathcal{D}^\alpha \mathbb{1}_{\{\mathcal{D}^\beta = k\}}]}{\mu}.$$

Then, for any $0 < \delta < \min\{\eta/(1 + \eta), 1/2\}$ and $\varepsilon < \min\{\delta, \eta/(2 + 2\eta)\}$,

$$\mathbb{P}(\Xi_n) = O\left(n^{-\delta+\varepsilon} + n^{-\frac{\eta}{2+2\eta}+\varepsilon} + n^{-\varepsilon}\right),$$

as $n \rightarrow \infty$, and hence IID($\mathcal{D}^+, \mathcal{D}^-$) satisfies Assumption 3.3.

The proof of Theorem 3.4 will follow similar steps as in the proof of Theorem 3.2. Hence, we will first establish some intermediate results. We start with the following technical result:

Lemma 3.5. *Let $\{X_i\}_{i \geq 1}$ be a sequence of i.i.d. mean zero random variables such that $\mathbb{E}[|X_1|^{1+\eta}] < \infty$ for some $0 < \eta \leq 1$. Then, there exists a constant $K < \infty$, which depends only on η , such that*

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > x\right) &\leq \frac{1}{x^{1+\eta}} \mathbb{E}\left[\left|\sum_{i=1}^n X_i\right|^{1+\eta}\right] \\ &\leq K \mathbb{E}[|X_1|^{1+\eta}] \frac{n}{x^{1+\eta}}. \end{aligned}$$

Proof. The result follows by first applying Markov's inequality and then Burkholder's inequality, see [25, Theorem 3.2], on the zero mean random walk $S_n = X_1 + \dots + X_n$. where we use that $(\sum_i y_i)^p \leq \sum_i y_i^p$ for any $y_i \geq 0$ and $0 < p \leq 1$. $\square$

The next lemma is the equivalence of Lemma 3.3 under Assumption 3.4:

Lemma 3.6. *Suppose that $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.4 and let $(\mathcal{D}_n^+, \mathcal{D}_n^-)$ be a vector of i.i.d. copies of $(\mathcal{D}^+, \mathcal{D}^-)$. In addition define for all combinations of $\alpha, \beta \in \{+, -\}$,*

$$\begin{aligned} f^{\alpha, \beta}(k) &= \frac{\mathbb{E} [\mathcal{D}^\alpha \mathbb{1}_{\{\mathcal{D}^\beta = k\}}]}{\nu} \quad \text{and} \\ \tilde{f}_n^{\alpha, \beta}(k) &= \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^\alpha \mathbb{1}_{\{\mathcal{D}_i^\beta = k\}}. \end{aligned}$$

Then, for any $0 < \varepsilon < \eta/(2 + 2\eta)$,

$$\mathbb{P}\left(\max\{d_1(\tilde{f}_n^{-, +}, f^{-, +}), d_1(\tilde{f}_n^{+, -}, f^{+, -})\} \leq n^{-\varepsilon}\right) \geq 1 - O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$.

Proof. Since the terms are symmetric, we will prove that

$$\mathbb{P}\left(d_1(\tilde{f}_n^{-,+}, f^{-,+}) > n^{-\varepsilon}\right) \leq O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$, from which the result follows.

Define $X_{ik} = \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ > k\}} - \mathbb{E}[\mathcal{D}^- \mathbb{1}_{\{\mathcal{D}^+ > k\}}]$. Then, using (2.3), we get

$$\begin{aligned} d_1(\tilde{f}_n^{-,+}, f^{-,+}) &= \sum_{k=0}^{\infty} \left| \sum_{t=k+1}^{\infty} \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ = t\}} - \frac{\mathbb{E}[\mathcal{D}^- \mathbb{1}_{\{\mathcal{D}^+ = t\}}]}{\nu} \right| \\ &= \sum_{k=0}^{\infty} \left| \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ > k\}} - \frac{\mathbb{E}[\mathcal{D}^- \mathbb{1}_{\{\mathcal{D}^+ > k\}}]}{\nu} \right| \\ &= \sum_{k=0}^{\infty} \frac{1}{\nu n} \left| \sum_{i=1}^n X_{ik} \right|. \end{aligned}$$

Now fix $1 < p := 1/(1 - 2\varepsilon) < 1 + \eta$ and first apply Markov's inequality, then Lemma 3.5 on the zero mean random variables X_{ik} and finally Minkowski's inequality to obtain

$$\begin{aligned} \mathbb{P}\left(d_1(\tilde{f}_n^{-,+}, f^{-,+}) > n^{-\varepsilon}\right) &\leq \frac{1}{\nu n^{1-\varepsilon}} \sum_{k=0}^{\infty} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik} \right| \right] \\ &\leq \frac{1}{\nu n^{1-\varepsilon}} \sum_{k=0}^{\infty} \mathbb{E} \left[\left| \sum_{i=1}^n X_{ik} \right|^p \right]^{1/p} \\ &\leq \frac{1}{\nu n^{1-\varepsilon}} \sum_{k=0}^{\infty} (Kn \mathbb{E}[|X_{1k}|^p])^{1/p} \\ &\leq \frac{K^{1/p} n^{1/p}}{\nu n^{1-\varepsilon}} \sum_{k=0}^{\infty} (Kn \mathbb{E}[|X_{1k}|^p])^{1/p} \\ &\leq \frac{2K^{1/p} n^{1/p}}{\nu n^{1-\varepsilon}} \sum_{k=0}^{\infty} \mathbb{E} [|\mathcal{D}^- \mathbb{1}_{\{\mathcal{D}^+ > k\}}|^p]^{1/p} \\ &\leq \frac{2K^{1/p} n^{1/p}}{\nu n^{1-\varepsilon}} \sum_{k=0}^{\infty} \mathbb{E} [(\mathcal{D}^-)^p \mathbb{1}_{\{\mathcal{D}^+ > k\}}]^{1/p}. \quad (3.20) \end{aligned}$$

For the last step observe that

$$\begin{aligned} &\mathbb{E} [(\mathcal{D}^-)^p \mathbb{1}_{\{\mathcal{D}^+ > k\}}] \\ &\leq \mathbb{E} [(\mathcal{D}^-)^p \mathcal{D}^+] \mathbb{1}_{\{k \leq 1\}} + \mathbb{E} [(\mathcal{D}^-)^p \mathbb{1}_{\{\mathcal{D}^+ > k\}}] \mathbb{1}_{\{k > 1\}} \\ &= \mathbb{E} [(\mathcal{D}^-)^p \mathcal{D}^+] \mathbb{1}_{\{k < 1\}} + k^{-(1+\eta)} k^{1+\eta} \mathbb{E} [(\mathcal{D}^-)^p \mathbb{1}_{\{\mathcal{D}^+ > k\}}] \mathbb{1}_{\{k \geq 1\}} \\ &= \mathbb{E} [(\mathcal{D}^-)^p \mathcal{D}^+] \mathbb{1}_{\{k < 1\}} + k^{-(1+\eta)} \mathbb{E} [(\mathcal{D}^-)^p (\mathcal{D}^+)^{1+\eta}] \mathbb{1}_{\{k \geq 1\}}. \end{aligned}$$

Inserting this into (3.20) gives

$$\begin{aligned}
& \mathbb{P} \left(d_1(\tilde{f}_n^{-,+}, f_n^{-,+}) > n^{-\varepsilon} \right) \\
& \leq \frac{2K^{1/p} n^{1/p}}{\nu n^{1-\varepsilon}} \left(\mathbb{E} [|(\mathcal{D}^-)^p \mathcal{D}^+|] + \sum_{k=1}^{\infty} k^{-(1+\eta)} \mathbb{E} [|(\mathcal{D}^-)^p (\mathcal{D}^+)^{1+\eta}|] \right) \\
& = O \left(n^{\varepsilon + \frac{1}{p} - 1} \right) = O \left(n^{-\varepsilon} \right),
\end{aligned}$$

as $n \rightarrow \infty$.

□

Proof of Theorem 3.4. Fix $0 < \varepsilon < \eta/(2+2\eta)$ and define the event

$$\Psi_n = \left\{ |\Delta_n| \leq n^{\frac{1}{1+\eta}} \right\}.$$

Next, note that since Assumption 3.4 implies Assumption 3.2 it follows from Theorem 3.2 that

$$\mathbb{P} \left((\Xi_n^{(1)})^c \right) \leq O \left(n^{-\frac{\eta}{1+\eta} + \varepsilon} + n^{-\varepsilon} \right),$$

as $n \rightarrow \infty$. Therefore it is enough to show that, as $n \rightarrow \infty$

$$\begin{aligned}
\mathbb{P} \left((\Xi_n^{(2)})^c \cap \Xi_n^{(1)} \right) & \leq O \left(n^{-\varepsilon} + n^{-\frac{\eta}{1+\eta} + \varepsilon} \right) \quad \text{and} \\
\mathbb{P} \left((\Xi_n^{(3)})^c \cap \Xi_n^{(1)} \right) & \leq O \left(n^{-\eta/(2+2\eta) + \varepsilon} \right).
\end{aligned}$$

To this end we define, for $\alpha, \beta \in \{+, -\}$, the empirical distributions

$$\begin{aligned}
\tilde{f}_n^\alpha(k) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\mathcal{D}_i^\alpha = k\}}, \\
\tilde{f}_n^{\alpha, \beta}(k) &= \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^\alpha \mathbb{1}_{\{\mathcal{D}_i^\beta = k\}}.
\end{aligned}$$

$$\mathbb{P} \left((\Xi_n^{(2)})^c \cap \Xi_n^{(1)} \right) \leq O \left(n^{-\varepsilon} + n^{-\eta/(1+\eta) + \varepsilon} \right).$$

We will only show that

$$\mathbb{P} \left(d_1(f_n^{-,+}, f_n^{-,+}) > n^{-\varepsilon}, \Xi_n^{(1)} \right) \leq O \left(n^{-\frac{\eta}{1+\eta} + \varepsilon} + n^{-\varepsilon} \right).$$

The proof of the other case follows in a similar fashion.

Note that by Lemma 3.6 and the triangle inequality

$$\begin{aligned}
& \mathbb{P} \left(d_1(f_n^{-,+}, f_n^{-,+}) > n^{-\varepsilon}, \Xi_n^{(1)} \right) \\
& \leq \mathbb{P} \left(d_1(\tilde{f}_n^{-,+}, f_n^{-,+}) > \frac{n^{-\varepsilon}}{2} \right) + \mathbb{P} \left(d_1(f_n^{-,+}, \tilde{f}_n^{-,+}) > \frac{n^{-\varepsilon}}{2}, \Xi_n^{(1)} \right) \\
& \leq \mathbb{P} \left(d_1(f_n^{-,+}, \tilde{f}_n^{-,+}) > \frac{n^{-\varepsilon}}{2}, \Xi_n^{(1)} \right) + O \left(n^{-\varepsilon} \right),
\end{aligned}$$

as $n \rightarrow \infty$. For the other probability we first recall that $\nu_n = L_n/n$, while, on the event $\Xi_n^{(1)}$,

$$\begin{aligned} d_1(f_n^{-,+}, \tilde{f}_n^{-,+}) &= \sum_{k=0}^{\infty} \left| \frac{1}{L_n} \sum_{i=1}^n D_i^- \mathbb{1}_{\{D_i^+ > k\}} - \frac{1}{\nu n} \sum_{i=1}^n \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ > k\}} \right| \\ &\leq \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| L_n + \frac{1}{\nu n} \sum_{k=0}^{\infty} \left| \sum_{i=1}^n D_i^- \mathbb{1}_{\{D_i^+ > k\}} - \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ > k\}} \right|. \end{aligned}$$

For the first term we have

$$\left| \frac{1}{L_n} - \frac{1}{\nu n} \right| L_n \leq \frac{|L_n - \nu n|}{\nu n} = \frac{|\nu_n - \nu|}{\nu} \leq \frac{d_1(f_n^-, f^-)}{\nu},$$

while we can bound the second term from above as follows

$$\begin{aligned} &\frac{1}{\nu n} \sum_{k=0}^{\infty} \left| \sum_{i=1}^n D_i^- \mathbb{1}_{\{D_i^+ > k\}} - \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ > k\}} \right| \\ &\leq \frac{1}{\nu n} \sum_{k=0}^{\infty} \left| \sum_{i=1}^n D_i^- \mathbb{1}_{\{D_i^+ > k\}} - \mathcal{D}_i^- \mathbb{1}_{\{D_i^+ > k\}} \right| \\ &\quad + \frac{1}{\nu n} \sum_{k=0}^{\infty} \left| \sum_{i=1}^n \mathcal{D}_i^- \mathbb{1}_{\{D_i^+ > k\}} - \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ > k\}} \right| \\ &\leq \frac{1}{\nu n} \sum_{k=0}^{\infty} \sum_{i=1}^n \xi_i^- \mathbb{1}_{\{D_i^+ > k\}} + \frac{1}{\nu n} \sum_{k=0}^{\infty} \sum_{i=1}^n \mathcal{D}_i^- \left| \mathbb{1}_{\{\mathcal{D}_i^+ + \xi_i^+ > k\}} - \mathbb{1}_{\{\mathcal{D}_i^+ > k\}} \right| \\ &\leq \frac{1}{\nu n} \sum_{i=1}^n \xi_i^- \mathcal{D}_i^+ + \frac{1}{\nu n} \sum_{k=0}^{\infty} \sum_{i=1}^n \mathcal{D}_i^- \mathbb{1}_{\{\mathcal{D}_i^+ \leq k < \mathcal{D}_i^+ + \xi_i^+\}} \\ &\leq \frac{1}{\nu n} \sum_{i=1}^n (\mathcal{D}_i^+ \xi_i^- + \mathcal{D}_i^- \xi_i^+). \end{aligned}$$

Therefore we have that

$$d_1(f_n^{-,+}, \tilde{f}_n^{-,+}) \leq \frac{d_1(f_n^-, f^-)}{\nu} + \frac{1}{\nu n} \sum_{i=1}^n (\mathcal{D}_i^+ \xi_i^- + \mathcal{D}_i^- \xi_i^+),$$

and hence

$$\begin{aligned} &\mathbb{P} \left(d_1(f_n^{-,+}, \tilde{f}_n^{-,+}) > \frac{n^{-\varepsilon}}{2}, \Xi_n^{(1)} \right) \\ &\leq 2n^\varepsilon \left(\frac{\mathbb{E}[d_1(f_n^-, f^-)]}{\nu} + \frac{2\mathbb{E}[(\mathcal{D}_1^+ \xi_1^- + \mathcal{D}_1^- \xi_1^+) \mathbb{1}_{\Xi_n^{(1)}}]}{\nu} \right) \\ &\leq \frac{2n^\varepsilon}{\mathbb{P}(\Psi_n)} \left(\frac{\mathbb{E}[d_1(f_n^-, f^-)]}{\nu} + \frac{2\mathbb{E}[(\mathcal{D}_1^+ \xi_1^- + \mathcal{D}_1^- \xi_1^+) \mathbb{1}_{\Xi_n^{(1)} \cap \Psi_n}]}{\nu} \right). \end{aligned}$$

Following the same arguments as in the proof of Theorem 3.2, see (3.15) and (3.12) we get, as $n \rightarrow \infty$,

$$\begin{aligned} & \frac{2n^\varepsilon}{\mathbb{P}(\Psi_n)} \left(\frac{\mathbb{E}[d_1(f_n^-, f^-)]}{\nu} + \frac{2\mathbb{E}[(\mathcal{D}_1^+ \xi_1^- + \mathcal{D}_1^- \xi_1^+) \mathbb{1}_{\Xi_n^{(1)} \cap \Psi_n}]}{\nu} \right) \\ & \leq \frac{2n^\varepsilon}{\mathbb{P}(\Psi_n)} \left(O\left(n^{-\frac{\eta}{1+\eta}}\right) + \frac{2\nu n^{-\frac{\eta}{1+\eta}}}{\nu - n^{-\varepsilon}} \right) = O\left(n^{-\frac{\eta}{1+\eta} + \varepsilon}\right), \end{aligned}$$

which completes the argument.

$$\mathbb{P}\left((\Xi_n^{(3)})^c \cap \Xi_n^{(1)}\right) \leq O\left(n^{-\eta/(2+2\eta)+\varepsilon}\right)$$

First we write

$$\begin{aligned} & \left| \sum_{k, \ell=0}^{\infty} (k\ell)^{1+\eta/2} h_n(k, \ell) - \mathbb{E}\left[(\mathcal{D}^+)^{1+\eta/2} (\mathcal{D}^-)^{1+\eta/2}\right] \right| \\ & \leq \left| \frac{1}{n} \sum_{i=1}^n (D_i^+ D_i^-)^{1+\eta/2} - (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} \right| \\ & \quad + \frac{1}{n} \left| \sum_{i=1}^n (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} - \mathbb{E}\left[(\mathcal{D}^+)^{1+\eta/2} (\mathcal{D}^-)^{1+\eta/2}\right] \right|. \end{aligned}$$

It then follows that

$$\begin{aligned} & \mathbb{P}\left((\Xi_n^{(3)})^c \cap \Xi_n^{(1)}\right) \\ & \leq \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n (D_i^+ D_i^-)^{1+\eta/2} - (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} \right| > \frac{Kn^{-\varepsilon}}{2} \middle| \Psi_n\right) \end{aligned} \quad (3.21)$$

$$+ \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} - \mathbb{E}\left[(\mathcal{D}^+)^{1+\eta/2} (\mathcal{D}^-)^{1+\eta/2}\right] \right| > \frac{Kn^{-\varepsilon}}{2}\right). \quad (3.22)$$

We will show that (3.21) is $O\left(n^{-\eta/(1+\eta)+\varepsilon}\right)$ and (3.22) is $O\left(n^{-\frac{\eta}{2+2\eta}+\varepsilon}\right)$. This implies the result, since $\eta/(2+2\eta) < \eta/(1+\eta)$.

For (3.22) define

$$X_i = (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} - \mathbb{E}\left[(\mathcal{D}^+)^{1+\eta/2} (\mathcal{D}^-)^{1+\eta/2}\right],$$

and let $p = (1+\eta)/(1+\eta/2)$ so that $\mathbb{E}[|X_1|^p] < \infty$. We then use Lemma 3.5 to obtain

$$\begin{aligned} & \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} - \mathbb{E}\left[(\mathcal{D}^+)^{1+\eta/2} (\mathcal{D}^-)^{1+\eta/2}\right] \right| > \frac{Kn^{-\varepsilon}}{2}\right) \\ & \leq \frac{2}{Kn^{1-\varepsilon}} \mathbb{E}\left[\left| \sum_{i=1}^n X_i \right|\right] \leq \frac{2}{Kn^{1-\varepsilon}} \mathbb{E}\left[\left| \sum_{i=1}^n X_i \right|^p\right]^{1/p} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{2}{Kn^{1-\varepsilon}} (K_1 n \mathbb{E}[|X_1|^p])^{1/p} \leq \frac{2K_1^{1/p}}{K} n^{\frac{1}{p}+\varepsilon-1} \mathbb{E}[|X_1|^p]^{1/p} \\
&= O\left(n^{-\frac{\eta}{2+2\eta}+\varepsilon}\right),
\end{aligned}$$

as $n \rightarrow \infty$, where we have used that $1/p = 1 - \eta/(2 + 2\eta)$.

We are now left with (3.21). For this we first write

$$\begin{aligned}
&\left| \frac{1}{n} \sum_{i=1}^n (D_i^+ D_i^-)^{1+\eta/2} - (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} \right| \\
&\leq \left| \frac{1}{n} \sum_{i=1}^n (D_i^+ D_i^-)^{1+\eta/2} - (D_i^+)^{1+\eta/2} (\mathcal{D}_i^-)^{1+\eta/2} \right| \\
&\quad + \left| \frac{1}{n} \sum_{i=1}^n (D_i^+)^{1+\eta/2} (\mathcal{D}_i^-)^{1+\eta/2} - (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} \right| \\
&\leq \frac{1}{n} \sum_{i=1}^n (D_i^+)^{1+\eta/2} \left((D_i^-)^{1+\eta/2} - (\mathcal{D}_i^-)^{1+\eta/2} \right) \tag{3.23}
\end{aligned}$$

$$+ \frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^-)^{1+\eta/2} \left((D_i^+)^{1+\eta/2} - (\mathcal{D}_i^+)^{1+\eta/2} \right), \tag{3.24}$$

where we used that $D_i^\pm \geq \mathcal{D}_i^\pm$. We will first analyze (3.23). Since $0 < \eta/2 \leq 1/2$ and $\xi_i^\pm \leq 1$ we have

$$\begin{aligned}
(D_i^\pm)^{1+\eta/2} &= (D_i^\pm + \xi_i^\pm) (D_i^\pm + \xi_i^\pm)^{\eta/2} \\
&\leq (D_i^\pm + \xi_i^\pm) \left((D_i^\pm)^{\eta/2} + \xi_i^\pm \right) \\
&= (D_i^\pm)^{1+\eta/2} + \xi_i^\pm \left((D_i^\pm)^{\eta/2} + 1 \right).
\end{aligned}$$

and hence

$$(D_i^-)^{1+\eta/2} - (\mathcal{D}_i^-)^{1+\eta/2} = \xi_i^\pm \left((D_i^\pm)^{\eta/2} + 1 \right).$$

Now recall that $\xi_i^+ \xi_i^- = 0$, so that

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n (D_i^+)^{1+\eta/2} \left((D_i^-)^{1+\eta/2} - (\mathcal{D}_i^-)^{1+\eta/2} \right) \\
&\leq \frac{1}{n} \sum_{i=1}^n (D_i^+)^{1+\eta/2} \xi_i^- \left((D_i^-)^{\eta/2} + 1 \right) \\
&\leq \frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} \xi_i^- + \frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+)^{1+\eta/2} \xi_i^-.
\end{aligned}$$

Similarly, we have that (3.24) is bounded from above by

$$\frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} \xi_i^+ + \frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^-)^{1+\eta/2} \xi_i^+.$$

Therefore, using Markov's inequality and the union bound, we have that

$$\begin{aligned}
& \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} - (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} \right| > \frac{Kn^{-\varepsilon}}{2}, \Xi_n^{(1)} \middle| \Psi_n \right) \\
& \leq \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+)^{1+\eta/2} \xi_i^- + (\mathcal{D}_i^-)^{1+\eta/2} \xi_i^+ > \frac{Kn^{-\varepsilon}}{4}, \Xi_n^{(1)} \middle| \Psi_n \right) \\
& \quad + \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n (\mathcal{D}_i^+ \mathcal{D}_i^-)^{1+\eta/2} (\xi_i^+ + \xi_i^-) > \frac{Kn^{-\varepsilon}}{4}, \Xi_n^{(1)} \middle| \Psi_n \right) \\
& \leq \frac{4n^\varepsilon}{K\mathbb{P}(\Psi_n)} \mathbb{E} \left[\left((\mathcal{D}_1^+)^{1+\eta/2} \xi_1^- + (\mathcal{D}_1^-)^{1+\eta/2} \xi_1^+ \right) \mathbb{1}_{\Xi_n^{(1)} \cap \Psi_n} \right] \\
& \quad + \frac{4n^\varepsilon}{K\mathbb{P}(\Psi_n)} \mathbb{E} \left[(\mathcal{D}_1^+ \mathcal{D}_1^-)^{1+\eta/2} (\xi_1^+ + \xi_1^-) \mathbb{1}_{\Xi_n^{(1)} \cap \Psi_n} \right]. \\
& = O \left(n^{-\frac{\eta}{1+\eta} + \varepsilon} \right).
\end{aligned}$$

It follows from (3.15) that the first term is $O \left(n^{-\frac{\eta}{1+\eta} + \varepsilon} \right)$. In a similar fashion we compute

$$\begin{aligned}
& \frac{4n^\varepsilon}{K\mathbb{P}(\Psi_n)} \mathbb{E} \left[(\mathcal{D}_1^+ \mathcal{D}_1^-)^{1+\eta/2} (\xi_1^+ + \xi_1^-) \mathbb{1}_{\Xi_n^{(1)} \cap \Psi_n} \right] \\
& \leq \frac{4n^\varepsilon}{K\mathbb{P}(\Psi_n)} \mathbb{E} \left[(\mathcal{D}_1^+ \mathcal{D}_1^-)^{1+\eta/2} \mathbb{E} [\xi_1^+ + \xi_1^- | (\mathcal{D}^+, \mathcal{D}^-)] \mathbb{1}_{\Xi_n^{(1)} \cap \Psi_n} \right] \\
& = \frac{4n^\varepsilon}{K\mathbb{P}(\Psi_n)} \mathbb{E} \left[(\mathcal{D}_1^+ \mathcal{D}_1^-)^{1+\eta/2} \frac{2|\Delta_n|}{L_n} \mathbb{1}_{\Xi_n^{(1)} \cap \Psi_n} \right] \\
& \leq \frac{4n^{-\frac{\eta}{1+\eta} + \varepsilon} \mathbb{E} [(\mathcal{D}_1^+ \mathcal{D}_1^-)^{1+\eta/2}]}{K\mathbb{P}(\Psi_n) (\nu - n^{-\varepsilon})} \\
& = O \left(n^{-\frac{\eta}{1+\eta} + \varepsilon} \right),
\end{aligned}$$

which completes the argument. $\square$

3.2.4 Scaling results for out- and in-degrees IID algorithm

We end this chapter with a result on the scaling of the product of degrees for bi-degree sequences generated by the IID algorithm. These results will be helpful in Chapter 8, when we analyze the scaling of the proportion of removed edges in the erased configuration model.

Lemma 3.7. *Suppose that $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.4 and let $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$. Then for any $\alpha, \beta \in \{+, -\}$, $0 < \delta < \min(1/(1+\eta), 1/2)$ and any increasing sequence a_n ,*

$$\sum_{i,j=1}^n \mathbb{E} \left[D_i^+ D_j^- \mathbb{1}_{\{D_i^\alpha > a_n\}} \right] \leq \begin{cases} O(n^{2-\delta} + n^2 a_n^{-\eta}) & \text{if } \alpha = + \\ O(n^{2-\delta} + n^2 a_n^{-1-\eta}) & \text{if } \alpha = -, \end{cases} \quad (3.25)$$

and

$$\sum_{i,j=1}^n \mathbb{E} \left[D_i^+ D_j^- \mathbb{1}_{\{D_j^\beta > a_n\}} \right] \leq \begin{cases} O(n^{2-\delta} + n^2 a_n^{-1-\eta}) & \text{if } \beta = + \\ O(n^{2-\delta} + n^2 a_n^{-\eta}) & \text{if } \beta = -, \end{cases} \quad (3.26)$$

as $n \rightarrow \infty$.

Note that the difference between (3.25) and (3.26) is that in the latter we condition on the event that $D_j^\beta > a_n$, while in the former this is $D_i^\alpha > a_n$.

Proof of Lemma 3.7. The proofs for both statements are similar. Hence we will only prove the first. For this we define the event

$$\Psi_n = \{|\Delta_n| \leq n^{1-\delta}\},$$

and note that by (3.5) $\mathbb{P}(\Psi_n) \rightarrow 1$ as $n \rightarrow \infty$. Next, note that by definition of the IID algorithm we have that

$$\sum_{i,j=1}^n \mathbb{E} \left[D_i^+ D_j^- \mathbb{1}_{\{D_i^\alpha > a_n\}} \right] \leq \sum_{i,j=1}^n \mathbb{E} \left[D_i^+ D_j^- \mathbb{1}_{\{D_i^\alpha > a_n\}} \middle| \Psi_n \right],$$

since the algorithm only terminates when Ψ_n is satisfied. Therefore it is enough to prove the scaling for the latter term. To this end we write

$$\begin{aligned} & \sum_{i,j=1}^n \mathbb{E} \left[D_i^+ D_j^- \mathbb{1}_{\{D_i^\alpha > a_n\}} \middle| \Psi_n \right] \\ & \leq \sum_{i,j=1}^n \mathbb{E} \left[(\mathcal{D}_i^+ + \xi_i^+) (\mathcal{D}_j^- + \xi_j^-) \mathbb{1}_{\{\mathcal{D}_i^\alpha > a_{n-1}\}} \middle| \Psi_n \right] \\ & \leq \sum_{i,j=1}^n \mathbb{E} \left[\mathcal{D}_i^+ \mathcal{D}_j^- \mathbb{1}_{\{\mathcal{D}_i^\alpha > a_{n-1}\}} \middle| \Psi_n \right] \end{aligned} \quad (3.27)$$

$$+ \sum_{i=1}^n \mathbb{E} \left[\mathcal{D}_i^+ \mathbb{1}_{\{\mathcal{D}_i^\alpha > a_{n-1}\}} |\Delta_n| \middle| \Psi_n \right] \quad (3.28)$$

$$+ \sum_{j=1}^n \mathbb{E} \left[\mathcal{D}_j^- |\Delta_n| \middle| \Psi_n \right]. \quad (3.29)$$

For (3.29), we bound

$$\sum_{j=1}^n \mathbb{E} \left[\mathcal{D}_j^- |\Delta_n| \middle| \Psi_n \right] \leq \frac{n^{1-\delta}}{\mathbb{P}(\Psi_n)} \sum_{j=1}^n \mathbb{E} \left[\mathcal{D}_j^- \right] \leq O(n^{2-\delta}), \quad (3.30)$$

as $n \rightarrow \infty$. In order to analyze the other terms we will need to consider the two cases for α .

Next, for (3.28), we observe that when $\alpha = +$,

$$\mathbb{E} \left[\mathcal{D}_1^+ \mathbb{1}_{\{\mathcal{D}_1^+ > a_{n-1}\}} \right] \leq (a_n - 1)^{-\eta} \mathbb{E} \left[(\mathcal{D}_1^+)^{1+\eta} \right]. \quad (3.31)$$

On the other hand, when $\alpha = -$,

$$\mathbb{E} \left[\mathcal{D}_1^+ \mathbb{1}_{\{\mathcal{D}_1^- > a_{n-1}\}} \right] \leq (a_n - 1)^{-1-\eta} \mathbb{E} \left[(\mathcal{D}_1^+ \mathcal{D}_1^-)^{1+\eta} \right] \quad (3.32)$$

Therefore, since

$$\sum_{i=1}^n \mathbb{E} \left[\mathcal{D}_i^+ \mathbb{1}_{\{\mathcal{D}_i^+ > a_{n-1}\}} |\Delta_n| \middle| \Psi_n \right] \leq \frac{n^{2-\delta}}{\mathbb{P}(\Psi_n)} \mathbb{E} \left[\mathcal{D}_1^+ \mathbb{1}_{\{\mathcal{D}_1^+ > a_{n-1}\}} \right],$$

it follows that, as $n \rightarrow \infty$,

$$\sum_{i=1}^n \mathbb{E} \left[\mathcal{D}_i^+ \mathbb{1}_{\{\mathcal{D}_i^\alpha > a_{n-1}\}} |\Delta_n| \middle| \Psi_n \right] \leq \begin{cases} O(n^{2-\delta} a_n^{-\eta}) & \text{if } \alpha = + \\ O(n^{2-\delta} a_n^{-1-\eta}) & \text{if } \alpha = -. \end{cases} \quad (3.33)$$

Finally we consider (3.27). Here

$$\begin{aligned} \sum_{i,j=1}^n \mathbb{E} \left[\mathcal{D}_i^+ \mathcal{D}_j^- \mathbb{1}_{\{\mathcal{D}_i^\alpha > a_{n-1}\}} \middle| \Psi_n \right] &= \frac{n(n-1)}{\mathbb{P}(\Psi_n)} \mathbb{E} \left[\mathcal{D}_1^+ \mathcal{D}_2^- \mathbb{1}_{\{\mathcal{D}_1^\alpha > a_{n-1}\}} \right] \\ &\quad + \frac{n}{\mathbb{P}(\Psi_n)} \mathbb{E} \left[\mathcal{D}_1^+ \mathcal{D}_1^- \mathbb{1}_{\{\mathcal{D}_1^\alpha > a_{n-1}\}} \right]. \end{aligned}$$

For the second term we use that

$$\mathbb{E} \left[\mathcal{D}_1^+ \mathcal{D}_1^- \mathbb{1}_{\{\mathcal{D}_1^\alpha > a_{n-1}\}} \right] \leq (a_n - 1)^{-\eta} \mathbb{E} \left[(\mathcal{D}_1^+ \mathcal{D}_1^-)^{1+\eta} \right],$$

which implies that

$$\frac{n}{\mathbb{P}(\Psi_n)} \mathbb{E} \left[\mathcal{D}_1^+ \mathcal{D}_1^- \mathbb{1}_{\{\mathcal{D}_1^\alpha > a_{n-1}\}} \right] \leq O(n a_n^{-\eta}),$$

as $n \rightarrow \infty$. Now, since

$$\mathbb{E} \left[\mathcal{D}_1^+ \mathcal{D}_2^- \mathbb{1}_{\{\mathcal{D}_1^\alpha > a_{n-1}\}} \right] = \nu \mathbb{E} \left[\mathcal{D}_1^+ \mathbb{1}_{\{\mathcal{D}_1^\alpha > a_{n-1}\}} \right],$$

we again use (3.31) and (3.32) to obtain,

$$\sum_{i,j=1}^n \mathbb{E} \left[\mathcal{D}_i^+ \mathcal{D}_j^- \mathbb{1}_{\{\mathcal{D}_i^\alpha > a_{n-1}\}} \middle| \Psi_n \right] \leq \begin{cases} O(n^2 a_n^{-\eta}) & \text{if } \alpha = + \\ O(n^2 a_n^{-1-\eta}) & \text{if } \alpha = -, \end{cases} \quad (3.34)$$

as $n \rightarrow \infty$. The result now follows by combining (3.34), (3.33) and (3.30). $\square$

Part II

Degree-degree correlations in directed random graphs

Chapter 4

Measures for degree-degree correlations in directed graphs

Based on [84]:

van der Hoorn, P. and Litvak, N.

**Degree-Degree Dependencies in Directed Networks with
Heavy-Tailed Degrees**

Internet Mathematics 11.2 (2015), pp. 155-179.

1

4.1 Introduction

In this chapter we will introduce several measures for degree-degree correlations in directed graphs. We first consider Pearson's correlation coefficient [59, 69] for all four correlations types, as described in Section 1.2. After that, we will define rank-correlation measures for directed networks, which extend the measures corresponding to Spearman's rho and Kendall's tau, as defined in [70, 82]. To illustrate the differences between the degree assortativity measures considered in this chapter, we will also show results of experiments on several large directed networks.

This chapter is structured as follows. In Section 4.2 we will define Pearson's correlation coefficient. Then in Section 4.3 we will introduce two specific versions of Spearman's rho for degree-degree correlations. Kendall's tau is defined in Section 4.4. Finally, Section 4.5 contains the numerical experiments.

4.2 Pearson's correlation coefficient

Among all the measures for degree-degree correlations, the one proposed by Newman [59, 60] is used the most in the literature. The idea is to compute the statistical estimator for the weighted covariance, Pearson correlation coefficient, on the degrees on both sides of a random edge in the network. We will consider Pearson's correlation coefficients as in [59, 60], adjusted to the setting of directed graphs as in [44, 69].

Consider two random variables X and Y , let $(\mathbf{X}_m, \mathbf{Y}_m)$ be the vector of m independent joint samples from (X, Y) and suppose we want to determine the correlation between X and Y , based on these m joint measurements. Then, if we define, for $p \in \{1, 2\}$,

$$\mathbb{E}_m[XY] = \frac{1}{m} \sum_{i=1}^m X_i Y_i, \quad \mathbb{E}_m[X^p] = \frac{1}{m} \sum_{i=1}^m X_i^p, \quad \mathbb{E}_m[Y^p] = \frac{1}{m} \sum_{i=1}^m Y_i^p, \quad (4.1)$$

we can compute Pearson's correlation coefficient, which is defined as

$$r(\mathbf{X}_m, \mathbf{Y}_m) = \frac{\mathbb{E}_m[XY] - \mathbb{E}_m[X] \mathbb{E}_m[Y]}{\sqrt{\mathbb{E}_m[X^2] - \mathbb{E}_m[X]^2} \sqrt{\mathbb{E}_m[Y^2] - \mathbb{E}_m[Y]^2}}. \quad (4.2)$$

This coefficient measures the weighted covariance of X and Y , based on the m joint measurements.

Now let G_n be a directed graph, take $\alpha, \beta \in \{+, -\}$ and consider the joint degree vector $(D_i^\alpha, D_j^\beta)_{i \rightarrow j}$, which consists of pairs of degrees (D_i^α, D_j^β) for each edge $i \rightarrow j \in G_n$. We can see this vector as a collection of L_n joint measurements of the α -degree of the source and the β -degree of the target of a random edge in G_n . Hence we can use $r((D_i^\alpha)_{i \rightarrow j}, (D_j^\beta)_{i \rightarrow j})$ as a measure for the α - β degree-degree correlation. This is the measure proposed by Newman, which we denote by $r_\alpha^\beta(G_n)$.

Remark 4.1. When considering measures for degree-degree correlations, we will consistently use the superscript index to denote the specific degree type of the target of an edge, while the subscript index denotes the degree type of the source. For

example, $r_+^-(G_n)$ denotes Pearson's correlation coefficient for the Out-In degree-degree correlation type and $r_-^+(G_n)$ that of the In-Out degree-degree correlation type, see Figure 1.4.

Written in terms of the degrees, Newman's measure is

$$r_\alpha^\beta(G_n) = \frac{\sum_{i \rightarrow j} D_i^\alpha D_j^\beta - \frac{1}{L_n} \sum_{i \rightarrow j} D_i^\alpha \sum_{i \rightarrow j} D_j^\beta}{\sigma_\alpha(G_n) \sigma^\beta(G_n)}, \quad (4.3)$$

where

$$\sigma_\alpha(G_n) = \sqrt{\sum_{i \rightarrow j} (D_i^\alpha)^2 - \frac{1}{L_n} \left(\sum_{i \rightarrow j} D_i^\alpha \right)^2} \quad \text{and} \quad (4.4)$$

$$\sigma^\beta(G_n) = \sqrt{\sum_{i \rightarrow j} (D_j^\beta)^2 - \frac{1}{L_n} \left(\sum_{i \rightarrow j} D_j^\beta \right)^2}. \quad (4.5)$$

Although $r_\alpha^\beta(G_n)$ is determined by the specific edge structure of G_n , only one term in (4.3) truly depends on it. To see this, we note that the summations over edges $i \rightarrow j$ of powers of either D_i^α or D_j^β can be rewritten in term of summations over vertices by observing that

$$\sum_{i \rightarrow j} (D_i^\alpha)^p = \sum_{i=1}^n D_i^+ (D_i^\alpha)^p$$

and similarly

$$\sum_{i \rightarrow j} (D_j^\beta)^p = \sum_{j=1}^n D_j^- (D_j^\beta)^p$$

for all $p \geq 0$. Plugging this into (4.3)-(4.5) we arrive at the following definition.

Definition 4.1. Let G_n be a directed graph and let $\alpha, \beta \in \{+, -\}$. Then the Pearson's α - β correlation coefficient is defined by

$$r_\alpha^\beta(G_n) = \frac{\sum_{i \rightarrow j} D_i^\alpha D_j^\beta - \frac{1}{L_n} \sum_{i=1}^n D_i^+ D_i^\alpha \sum_{j=1}^n D_j^- D_j^\beta}{\sigma_\alpha(G_n) \sigma^\beta(G_n)}, \quad (4.6)$$

where

$$\sigma_\alpha(G_n) = \sqrt{\sum_{i=1}^n D_i^+ (D_i^\alpha)^2 - \frac{1}{L_n} \left(\sum_{i=1}^n D_i^+ D_i^\alpha \right)^2} \quad \text{and} \quad (4.7)$$

$$\sigma^\beta(G_n) = \sqrt{\sum_{j=1}^n D_j^- (D_j^\beta)^2 - \frac{1}{L_n} \left(\sum_{j=1}^n D_j^- D_j^\beta \right)^2}. \quad (4.8)$$

When either $\sigma_\alpha(G_n) = 0$ or $\sigma^\beta(G_n) = 0$ we define $r_\alpha^\beta(G_n) = 0$.

An important observation is that just as in the undirected case, c.f. [53, 82], the structure of the graph only contributes to the positive part of (4.6), while all other terms are completely determined by the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$. In undirected graphs $\sum_{i \rightarrow j} D_i^\alpha D_j^\beta$ is often referred to as the s metric of the graph G_n , see for instances [2]. The fact that the negative part of Pearson's α - β correlation coefficient depends only on the bi-degree sequence will enable us to analyze the behavior of $r_\alpha^\beta(G)$, for networks with scale-free degree distributions, in Chapter 5. In contrast to undirected graphs we see that in the directed case the correlation between the out- and in-degrees of a node can play a role, take for instance $\alpha = -$ and $\beta = +$.

Note that there are graphs for which $\sigma_\alpha(G_n)$ or $\sigma^\beta(G_n)$ equals zero. For example, when G_n is a directed regular graph of arbitrary size. From (4.7) it follows that $\sigma_\alpha(G_n) \neq 0$ is only possible if $D_i^\alpha \neq D_j^\alpha$ for some $1 \leq i < j \leq n$. Moreover, node i and j must have non-zero out-degree for at least one such pair i, j , so that D_i^α and D_j^α are counted when we sum over the edges. This argument is formalized in the following lemma, which provides necessary and sufficient conditions so that $\sigma_\alpha(G_n), \sigma^\beta(G_n) \neq 0$.

Lemma 4.1. *Let G_n be a directed graph of size n and take $\alpha, \beta \in \{+, -\}$. Then*

$$\frac{1}{L_n} \left(\sum_{i=1}^n D_i^\alpha D_i^\beta \right)^2 \leq \sum_{j=1}^n D_j^\alpha (D_j^\beta)^2, \quad (4.9)$$

and strict inequality holds if and only if there exists distinct $1 \leq i, j \leq n$ such that $D_i^\alpha, D_j^\alpha > 0$ and $D_i^\beta \neq D_j^\beta$.

Proof. Recall that $L_n = \sum_{i=1}^n D_i^\alpha$ for any $\alpha \in \{+, -\}$. Therefore we have,

$$\begin{aligned} & L_n \sum_{i=1}^n D_i^\alpha (D_i^\beta)^2 - \left(\sum_{j=1}^n D_j^\alpha D_j^\beta \right)^2 \\ &= \sum_{i=1}^n \sum_{j \neq i} D_i^\alpha D_j^\alpha (D_j^\beta)^2 - D_i^\alpha D_i^\beta D_j^\alpha D_j^\beta \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i} D_i^\alpha D_j^\alpha \left((D_i^\beta)^2 - 2D_i^\beta D_j^\beta + (D_j^\beta)^2 \right) \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i} D_i^\alpha D_j^\alpha \left(D_i^\beta - D_j^\beta \right)^2 \geq 0, \end{aligned}$$

which proves (4.9). From the last line one easily sees that strict inequality holds if and only if there exists distinct $1 \leq i, j \leq n$, such that $D_i^\alpha, D_j^\alpha > 0$ and $D_i^\beta \neq D_j^\beta$. $\square$

4.3 Spearman's rho and related rank-correlation measures

In this section we consider another class of measures for degree-degree correlations, which are related to Spearman's rho [75]. These measures belong to the class of so-called rank-correlation measures, since they are based on the ranks of the degrees rather than their actual value.

4.3.1 Definition of Spearman's rho

Let us go back to the setting in which we have m joint measurements $(\mathbf{X}_m, \mathbf{Y}_m)$ of (X, Y) and we want to measure the correlation between X and Y . Spearman introduced the following approach for this [75]. First order $\mathbf{X}_m$ in descending order

$$X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(m)},$$

and let $1 \leq \varrho_X(i) \leq m$ denote the rank of X_i in this order, i.e. $X_i = X_{(\varrho_X(i))}$ and similarly for $\mathbf{Y}_m$, where the rank is denoted by $\varrho_Y(i)$. This constitutes a transformation

$$(\mathbf{X}_m, \mathbf{Y}_m) \mapsto (\varrho\mathbf{X}_m, \varrho\mathbf{Y}_m),$$

and Spearman's rho is then defined as Pearson correlation coefficient of the vector of ranks related to the joint measurements $(\mathbf{X}_m, \mathbf{Y}_m)$,

$$\rho(\mathbf{X}_m, \mathbf{Y}_m) = r(\varrho\mathbf{X}_m, \varrho\mathbf{Y}_m),$$

where $r(\cdot, \cdot)$ is defined by (4.2). Note that in this setting many terms in (4.2) become constants, for instance $\sum_{i=1}^m \varrho_X(i) = m(m+1)/2$. We will see specific examples of this in Section 4.3.2 and 4.3.3.

Although the original idea of Spearman is to assign to each value a unique rank, the concept can be generalized. For this, let ϱ_X denote the ranking function of the vector $\mathbf{X}_m$. Then we can see the ranking function ϱ_X as a bijection between the set $\{X_1, \dots, X_m\}$ and $\{1, \dots, m\}$, where the eventual ranks are determined by the natural order on the latter. Observe however, that there is nothing special about $\{1, \dots, m\}$ and actually, all that we need is that ϱ_X maps the entries in $\{X_1, \dots, X_m\}$ to some ordered set such that, for any distinct $1 \leq i, j \leq m$,

$$\varrho_X(X_i) < \varrho_X(X_j) \Rightarrow X_i > X_j.$$

We refer to such a function as a *ranking*.

With this setup, and using similar notations as in (4.1), we obtain the following general version of Spearman's rho for any two rankings ϱ_X and ϱ_Y :

$$\rho(\mathbf{X}_m, \mathbf{Y}_m | \varrho) = \frac{\mathbb{E}_m[\varrho_X \varrho_Y] - \mathbb{E}_m[\varrho_X] \mathbb{E}_m[\varrho_Y]}{\sqrt{\mathbb{E}_m[\varrho_X^2] - \mathbb{E}_m[\varrho_X]^2} \sqrt{\mathbb{E}_m[\varrho_Y^2] - \mathbb{E}_m[\varrho_Y]^2}}.$$

Similar to the previous section, we can adapt this measure to the setting of degree-degree correlations in directed graphs by applying it to the joint degree vector $(D_i^\alpha, D_j^\beta)_{i \rightarrow j}$, for $\alpha, \beta \in \{+, -\}$. We emphasize that if ϱ_α is a ranking of

$(D_i^\alpha)_{i \rightarrow j}$, $\varrho_\alpha(i \rightarrow j)$ denotes the rank of the edge $i \rightarrow j$ and if we have more edges $i \rightarrow j$, then the rank will, in general, be different for each of them. Since for any function $f : \mathbb{R} \rightarrow \mathbb{R}$ and $\mathbf{X}_m$

$$\mathbb{E}_m[f(X)] = \frac{1}{m} \sum_{i=1}^m f(X_i),$$

we have the following general definition of Spearman's rho for degree-degree correlations in directed graphs:

Definition 4.2. Let G_n be a directed graph, $\alpha, \beta \in \{+, -\}$ and let ϱ_α and ϱ^β be rankings of, respectively, $(D_i^\alpha)_{i \rightarrow j}$ and $(D_j^\beta)_{i \rightarrow j}$. Then we define the α - β Spearman's rho, with respect to ϱ_α and ϱ^β , by

$$\rho_\alpha^\beta(G_n|\varrho) = \frac{\sum_{i \rightarrow j} \varrho_\alpha(i \rightarrow j) \varrho^\beta(i \rightarrow j) - \frac{1}{L_n} \sum_{i \rightarrow j} \varrho_\alpha(i \rightarrow j) \sum_{i \rightarrow j} \varrho^\beta(i \rightarrow j)}{\sigma_\alpha(G_n|\varrho) \sigma^\beta(G_n|\varrho)},$$

where

$$\sigma_\alpha(G_n|\varrho) = \sqrt{\sum_{i \rightarrow j} \varrho_\alpha(i \rightarrow j)^2 - \frac{1}{L_n} \left(\sum_{i \rightarrow j} \varrho_\alpha(i \rightarrow j) \right)^2} \quad \text{and}$$

$$\sigma^\beta(G_n|\varrho) = \sqrt{\sum_{i \rightarrow j} \varrho^\beta(i \rightarrow j)^2 - \frac{1}{L_n} \left(\sum_{i \rightarrow j} \varrho^\beta(i \rightarrow j) \right)^2}.$$

With this definition we have a family of measures for degree-degree correlations in directed graphs, determined by the rankings ϱ_α and ϱ^β . In the next two sections we will consider two specific types of ranking functions, which will give rise to two different rank-correlation measures.

4.3.2 Spearman's rho with uniform tie resolution

Here we will consider the original setting of Spearman's rho. That is, given a directed graph G_n and joint degree vector $(D_i^\alpha, D_j^\beta)_{i \rightarrow j}$, we want to order the vectors $(D_i^\alpha)_{i \rightarrow j}$ and $(D_j^\beta)_{i \rightarrow j}$ in descending order and assign a unique rank to each element. The problem that we run into here is that these vectors contain many equal values. For example, consider an edge $i \rightarrow j$. Then there exist D_i^+ edges $i \rightarrow t$ and hence we will encounter the value D_i^α at least D_i^+ times. If we want to assign unique ranks to each edge $i \rightarrow j$, then we need to resolve these ties. For this we will use the continuization principle, see Section 2.1.4 for more details. Let (U, W) be two independent uniform random variables on the interval $[0, 1)$, and let $(\mathbf{U}_{L_n}, \mathbf{W}_{L_n})$ be a vector of L_n joint i.i.d. copies of (U, W) , which we will index using the edges as $U_{i \rightarrow j}$ and $W_{i \rightarrow j}$. Then we define the uniform rankings

$$R_\alpha(i \rightarrow j) = \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha + U_{s \rightarrow t} \geq D_i^\alpha + U_{i \rightarrow j}\}}, \quad (4.10)$$

$$R^\beta(i \rightarrow j) = \sum_{s \rightarrow t} \mathbb{1}_{\{D_t^\beta + W_{s \rightarrow t} \geq D_j^\beta + W_{i \rightarrow j}\}}. \quad (4.11)$$

To clarify these rankings, consider, instead of the original vector of degrees, the joint vector

$$(D_i^\alpha + U_{i \rightarrow j}, D_j^\beta + W_{i \rightarrow j})_{i \rightarrow j}.$$

Then, if we order both vectors in descending order, $R_\alpha(i \rightarrow j)$ and $R^\beta(i \rightarrow j)$ denote the rank of the edge $i \rightarrow j$ among the vector $(D_i^\alpha + U_{L_n})_{i \rightarrow j}$ and $(D_j^\beta + W_{L_n})_{i \rightarrow j}$, respectively.

Observe that since,

$$\mathbb{P}(U_{i \rightarrow j} = U_{s \rightarrow t}) = 0 = \mathbb{P}(W_{i \rightarrow j} = W_{s \rightarrow t})$$

for any two distinct edges $i \rightarrow j$ and $s \rightarrow t$, it follows that the vectors $(D_i^\alpha + U_{L_n})_{i \rightarrow j}$ and $(D_j^\beta + W_{L_n})_{i \rightarrow j}$ will, with probability one, not contain any ties. Therefore the rankings $R_\alpha(i \rightarrow j)$ and $R^\beta(i \rightarrow j)$ are bijections from the set of edges to $\{1, \dots, L_n\}$. As a consequence we have that

$$\begin{aligned} \sum_{i \rightarrow j} R_\alpha(i \rightarrow j) &= \sum_{t=1}^{L_n} t = \frac{L_n(L_n + 1)}{2}, \\ \sum_{i \rightarrow j} R_\alpha(i \rightarrow j)^2 &= \sum_{t=1}^{L_n} t^2 = \frac{L_n^3}{3} + \frac{L_n^2}{2} + \frac{L_n}{6}, \end{aligned}$$

and similar for $R^\beta(i \rightarrow j)$. Plugging this into the expression for $\sigma_\alpha(G_n|\rho)$, from Definition 4.2, we get,

$$\begin{aligned} \sigma_\alpha(G_n|R)^2 &= \sum_{i \rightarrow j} R_\alpha(i \rightarrow j)^2 - \frac{1}{L_n} \left(\sum_{i \rightarrow j} R_\alpha(i \rightarrow j) \right)^2 \\ &= \frac{L_n^3}{3} + \frac{L_n^2}{2} + \frac{L_n}{6} - \frac{L_n(L_n + 1)^2}{4} \\ &= \frac{L_n^3 - L_n}{12} = \sigma^\beta(G_n|R). \end{aligned}$$

Since, in addition,

$$\frac{1}{L_n} \sum_{i \rightarrow j} R_\alpha(i \rightarrow j) \sum_{i \rightarrow j} R^\beta(i \rightarrow j) = \frac{L_n(L_n + 1)^2}{4},$$

we obtain the following measure:

Definition 4.3. Let G_n be a directed graph, $\alpha, \beta \in \{+, -\}$ and let R_α and R^β be defined as in, respectively, (4.10) and (4.11). Then we define the α - β Spearman's rho, with uniform tie resolution, by

$$\rho_\alpha^\beta(G_n) = \frac{12 \sum_{i \rightarrow j} R_\alpha(i \rightarrow j) R^\beta(i \rightarrow j) - 3L_n(L_n + 1)^2}{L_n^3 - L_n}. \quad (4.12)$$

We emphasize that the ties are resolved uniformly at random each time ρ_α^β is computed. Hence, if we compute this measure several times an edge can receive a different rank for each separate computation. This could potentially give rise to very different results among several instances, see [84, Section 5.1] for a worse-case example. Therefore, in experiments, one could take an average of $\rho_\alpha^\beta(G_n)$ over several instances of the uniform rankings.

4.3.3 Spearman's rho with average ranking

A different approach for resolving ties is to assign the same, average, rank to all tied values. As an example, suppose we need to assign ranks 1, 2, 3, 4, 5 and 6 to the elements in the vector (1, 2, 1, 3, 3, 1). Since 3 is the largest value we should assign ranks 1 and 2 to these two elements. Instead, we assign the average of these rank to both of them. This scheme implies the following transformation

$$(3, 3, 2, 1, 1, 1) \mapsto \left(\frac{3}{2}, \frac{3}{2}, 3, 5, 5, 5\right).$$

Now let G_n be a directed graph and $\alpha, \beta \in \{+, -\}$. Then, using the procedure as described above, we define the average rankings

$$\bar{R}_\alpha(i \rightarrow j) = \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha > D_t^\alpha\}} + \frac{1}{2} \left(1 + \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha = D_t^\alpha\}} \right) \quad (4.13)$$

$$\bar{R}^\beta(i \rightarrow j) = \sum_{s \rightarrow t} \mathbb{1}_{\{D_t^\beta > D_j^\beta\}} + \frac{1}{2} \left(1 + \sum_{s \rightarrow t} \mathbb{1}_{\{D_t^\beta = D_j^\beta\}} \right). \quad (4.14)$$

Observe that with the above definitions we have, similar to the rankings R_α and R^β ,

$$\sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j) = \frac{L_n(L_n + 1)}{2} = \sum_{i \rightarrow j} \bar{R}^\beta(i \rightarrow j).$$

The difference with resolving ties uniformly at random is that we in general do not know $\sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j)^2$, as this depends on how many ties we have. Inserting these ranks into the general definition of Spearman's rho we obtain a second rank-correlation measure for degree-degree correlations.

Definition 4.4. Let G_n be a directed graph, $\alpha, \beta \in \{+, -\}$ and let $\bar{R}_\alpha$ and $\bar{R}^\beta$ be defined as in, respectively, (4.13) and (4.14). Then we define the α - β Spearman's rho, with average ranking, by

$$\bar{\rho}_\alpha^\beta(G_n) = \frac{4 \sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j) \bar{R}^\beta(i \rightarrow j) - L_n(L_n + 1)^2}{\bar{\sigma}_\alpha(G_n) \bar{\sigma}^\beta(G_n)}, \quad (4.15)$$

where

$$\begin{aligned} \bar{\sigma}_\alpha(G) &= \sqrt{4 \sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j)^2 - L_n(L_n + 1)^2} \quad \text{and} \\ \bar{\sigma}^\beta(G) &= \sqrt{4 \sum_{i \rightarrow j} \bar{R}^\beta(i \rightarrow j)^2 - L_n(L_n + 1)^2}. \end{aligned}$$

Note that $\bar{\rho}_\alpha^\beta(G_n)$ does not suffer from any randomness in the ranking of the degrees. Hence, in contrast to (4.12), here we do not need to take an average over multiple instances.

Now consider an edge $i \rightarrow j$. Then since $\mathbf{U}_{s \rightarrow t}$ are i.i.d. copies of a uniform random variable U on $[0, 1)$ we have

$$\begin{aligned} \mathbb{E}[R_\alpha(i \rightarrow j)] &= \mathbb{E}\left[\sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha + U_{s \rightarrow t} \geq D_i^\alpha + U_{i \rightarrow j}\}}\right] \\ &= \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha > D_i^\alpha\}} + \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha = D_i^\alpha\}} \mathbb{E}[\mathbb{1}_{\{U_{s \rightarrow t} \geq U_{i \rightarrow j}\}}] \\ &= \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha > D_i^\alpha\}} + \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha = D_i^\alpha\}} \left(\mathbb{1}_{\{s=i, t=j\}} + \frac{1 - \mathbb{1}_{\{s=i, t=j\}}}{2}\right) \\ &= \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha > D_i^\alpha\}} + \frac{1}{2} \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha = D_i^\alpha\}} + \frac{1}{2} \\ &= \bar{R}_\alpha(i \rightarrow j), \end{aligned}$$

and a similar result holds for $\mathbb{E}[R^\beta(i \rightarrow j)]$. Hence the expectation over the uniform rankings are equal to the average rankings. Moreover, since U and W are independent it is easy to see that the above results imply that

$$\mathbb{E}[R_\alpha(i \rightarrow j)R^\beta(i \rightarrow j)] = \bar{R}_\alpha(i \rightarrow j)\bar{R}^\beta(i \rightarrow j),$$

and hence

$$\mathbb{E}[\rho_\alpha^\beta(G_n)] = \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3 - L_n} \bar{\rho}_\alpha^\beta(G_n). \quad (4.16)$$

Therefore, instead of computing $\rho_\alpha^\beta(G_n)$ several times and taking the average we can just apply the average rankings, compute $\bar{\sigma}_\alpha(G_n)$ and $\bar{\sigma}^\beta(G)$, and use (4.16). We end this section by observing that from (4.16) it follows that the difference between $\rho_\alpha^\beta(G_n)$ and $\bar{\rho}_\alpha^\beta(G_n)$ is completely determined by the number of ties in the vectors $(\mathbf{D}_i^\alpha)_{i \rightarrow j}$ and $(\mathbf{D}_j^\beta)_{i \rightarrow j}$.

4.4 Kendall's tau

Another common rank-correlation measure is Kendall's tau [49]. This measure computes the weighted difference between the number of concordant and discordant pairs of the joint observations. More precisely, given a vector $(\mathbf{X}_m, \mathbf{Y}_m)$ of joint measurements, we say that a pair (X_i, Y_i) and (X_j, Y_j) of joint observations is concordant if

$$(X_i - X_j)(Y_i - Y_j) > 0,$$

and discordant if

$$(X_i - X_j)(Y_i - Y_j) < 0.$$

If we let $\mathfrak{C}(\mathbf{X}_m, \mathbf{Y}_m)$ and $\mathfrak{D}(\mathbf{X}_m, \mathbf{Y}_m)$ denote, respectively, the number of concordant and discordant pairs among the joint vector $(\mathbf{X}_m, \mathbf{Y}_m)$, then Kendall's tau is defined as

$$\tau(\mathbf{X}_m, \mathbf{Y}_m) = \frac{\mathfrak{C}(\mathbf{X}_m, \mathbf{Y}_m) - \mathfrak{D}(\mathbf{X}_m, \mathbf{Y}_m)}{m(m-1)/2}. \quad (4.17)$$

Now let ϱ_X and ϱ_Y be, respectively, the rankings of $\mathbf{X}_m$ and $\mathbf{Y}_m$, as in Section 4.3.1, and consider the joint vector of ranks $(\varrho_{\mathbf{X}_m}, \varrho_{\mathbf{Y}_m})$. When the ranks are in perfect agreement, $\varrho_{\mathbf{X}_m} = \varrho_{\mathbf{Y}_m}$, the number of discordant pairs equals zero. If, on the other hand, we have that the ranks are inverse to each other, the number of concordant pairs is zero. Therefore, when all values in the two vectors $\mathbf{X}_m$ and $\mathbf{Y}_m$ are distinct, the latter case implies $\tau(\mathbf{X}_m, \mathbf{Y}_m) = -1$, while for the former we have $\tau(\mathbf{X}_m, \mathbf{Y}_m) = 1$. It can be shown that Kendall's tau is proportional to the number of adjacent swaps needed to match one of the vectors $\varrho_{\mathbf{X}_m}$, $\varrho_{\mathbf{Y}_m}$, to the other. This insight is used in [51] to design an efficient algorithm, using merge sort, for computing Kendall's tau which is known as Knight's algorithm.

Observe that, by definition, if either $X_i = X_j$ or $Y_i = Y_j$ the pair of joint observations (X_i, Y_i) and (X_j, Y_j) is considered neither concordant nor discordant. Hence, $\mathfrak{C}(\mathbf{X}_m, \mathbf{Y}_m)$ and $\mathfrak{D}(\mathbf{X}_m, \mathbf{Y}_m)$ will only count those observations that do not contain any ties, while the normalization $m(m-1)/2$ is the total number joint observations. When we include the number of ties in the normalization we obtain a different measure which is referred to as Kendall's tau b. In this thesis, however, we will consider only (4.17), translated to the setting of directed graphs as follows:

Definition 4.5. Let G_n be a directed graph, $\alpha, \beta \in \{-, +\}$ and denote by $\mathfrak{C}_\alpha^\beta(G_n)$ and $\mathfrak{D}_\alpha^\beta(G_n)$, respectively, the number of concordant and discordant pairs among $(\mathbf{D}_i^\alpha, \mathbf{D}_j^\beta)_{i \rightarrow j}$. Then we define the α - β Kendall's tau of G_n by

$$\tau_\alpha^\beta(G_n) = \frac{2(\mathfrak{C}_\alpha^\beta(G_n) - \mathfrak{D}_\alpha^\beta(G_n))}{L_n(L_n - 1)}. \quad (4.18)$$

4.5 Measured degree-degree correlation in real-world networks

We now illustrate how the degree-degree correlation measures from Definitions 4.1, 4.3, 4.4 and 4.5 perform on real networks. For this we consider the full Wikipedia graphs, wikipedia.org, of nine different languages. Here nodes correspond to different pages and the directed edges are represented by the hyperlinks between the pages. The graphs are obtained from the The Laboratory for Web Algorithmics (LAW) dataset database, law.di.unimi.it/datasets.php.

The networks are classified by their language, using the ISO 639-1 language codes¹. For instance EN Wiki refers to the Wikipedia graph for the English language. For each Wikipedia graph we have calculated all four degree-degree correlations using the four measures introduced in this paper. Since these graphs are large we use the implementation of Knights algorithm, as provided by the software libraries from the Laboratory for Web Algorithmics (law.di.unimi.it), for efficiently computing Kendall's tau. The full set of results can be found in Table 4.1. For further illustration we have displayed the results for two networks in Figure 4.1.

The first observation is that for each Wikipedia graph and degree-degree correlation type, the measures ρ , $\bar{\rho}$ and τ always have the same sign, while r has

¹http://www.iso.org/iso/home/standards/language_codes.htm

Figure 4.1: Plot of the values for the four measures for degree-degree correlation measures in Wikipedia graphs.

a different sign in six cases. See, for instance, the values for the Out-Out degree-degree correlation type for the NL Wiki network, Figure 4.1b, or the In-In degree-degree correlation type for the HU Wiki network in Table 4.1. Furthermore, there are many cases where the absolute value of the three rank correlations is at least an order of magnitude larger than that of Pearson's correlation coefficients. See for instance the Out-In degree-degree correlation type for the DE, EN, FR and NL Wikipedia networks or the values for the In-Out degree-degree correlation type for KO and RU. Moreover our experiments illustrate that the values for ρ and $\bar{\rho}$ are almost in full agreement with each other for all four degree-degree correlation types. This would then suggest, looking back at equation (4.16), that $3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n) \approx L_n^3 - L_n$ for the Wikipedia networks.

Based on these experiments we conclude that different measures for degree-degree correlations in directed networks give different results and it is not clear, initially, where these differences come from. Moreover, in several cases there even seems to be a disagreement between Pearson's correlation coefficient r_α^β and the rank-correlation measures ρ_α^β , $\bar{r}_\alpha^\beta$ and τ_α^β , in the sign of the correlation which is important for characterizing a network as either assortative or disassortative. Therefore we need to conduct a more thorough analysis of these four measures to see whether they are actually good measures for degree-degree correlations.

Graph	α/β	ρ_α^β	$\bar{\rho}_\alpha^\beta$	τ_α^β	r_α^β
DE wiki	+/-	-0.1362	-0.1362	-0.0934	-0.0518
	-/+	0.0446	0.0447	0.0294	0.0052
	+/+	-0.0570	-0.0570	-0.0395	-0.0399
	-/-	-0.0071	-0.0071	-0.0059	-0.0117
EN wiki	+/-	-0.1999	-0.1999	-0.1364	-0.0557
	-/+	0.0239	0.0240	0.0163	-0.0007
	+/+	-0.0855	-0.0855	-0.0581	-0.0713
	-/-	-0.0664	-0.0666	-0.0457	-0.0074
ES wiki	+/-	-0.1429	-0.1429	-0.0972	-0.1031
	-/+	-0.0407	-0.0417	-0.0294	-0.0033
	+/+	0.0178	0.0178	0.0119	-0.0272
	-/-	-0.1627	-0.1669	-0.1174	-0.0262
FR wiki	+/-	-0.1065	-0.1065	-0.0720	-0.0536
	-/+	0.0119	0.0121	0.0085	0.0048
	+/+	-0.0126	-0.0126	-0.0087	-0.0512
	-/-	-0.0262	-0.0267	-0.0186	-0.0094
HU wiki	+/-	-0.1280	-0.1280	-0.0877	-0.1048
	-/+	0.0525	0.0595	0.0442	0.0120
	+/+	-0.0207	-0.0207	-0.0140	-0.0579
	-/-	0.0051	0.0060	0.0050	-0.0279
IT wiki	+/-	-0.0964	-0.0964	-0.0653	-0.0711
	-/+	0.0468	0.0469	0.0319	0.0048
	+/+	-0.0277	-0.0277	-0.0189	-0.0704
	-/-	-0.0428	-0.0429	-0.0296	-0.0115
KO wiki	+/-	-0.2696	-0.2722	-0.1985	-0.0805
	-/+	0.1760	0.2323	0.1902	0.0157
	+/+	0.0016	0.0191	0.0170	-0.1697
	-/-	-0.0493	-0.0618	-0.0463	-0.0138
NL wiki	+/-	-0.2943	-0.2944	-0.2035	-0.0645
	-/+	0.0646	0.0647	0.0442	0.0074
	+/+	0.0161	0.0161	0.0111	-0.0538
	-/-	-0.1843	-0.1846	-0.1275	-0.0245
RU wiki	+/-	-0.1080	-0.1084	-0.0755	-0.0911
	-/+	0.1977	0.2200	0.1655	0.0398
	+/+	0.2472	0.2480	0.1736	0.0082
	-/-	0.0236	0.0255	0.0187	-0.0242

Table 4.1: Degree-degree correlations for Wikipedia graphs.

Chapter 5

Convergence of correlation measures in directed random graphs

Based on [84]:

van der Hoorn, P. and Litvak, N.

Degree-Degree Dependencies in Directed Networks with Heavy-Tailed Degrees

Internet Mathematics 11.2 (2015), pp. 155-179.

and [83]:

van der Hoorn, P. and Litvak, N.

Convergence of rank based degree-degree correlations in random directed networks

Moscow Journal of Combinatorics and Number Theory 4.4 (2014), pp 45-83.

5.1 Introduction

Currently, the most commonly used measure for degree-degree correlations is Pearson's correlation coefficient r_α^β (Definition 4.1), introduced in [59, 60]. The structure of many networks has been analyzed using this measure, see for instance [60, Table I] and [61, Table II]. However, since r_α^β essentially deals with empirical estimators of the second, and sometimes third moment of the degrees, one can expect this measure to suffer when the network has nodes with very large degree. Under the assumption that degrees in an undirected network follow a regularly-varying distribution with exponent $1 < \gamma < 3$ it is proved that Pearson's correlation coefficient converges to a non-negative number as the size of the graph goes to infinity, see [53, 82].

This result implies that for large graphs with scale-free degree distribution and exponent $1 < \gamma < 3$, if Pearson's correlation coefficient is negative, it will be very small in absolute value. Because most real-world networks have been reported to have scale-free degree distributions with exponent in $1 < \gamma < 3$, c.f. [1, 61, Table II], this could then explain why large networks are rarely classified as disassortative. Moreover, the above results indicate that Pearson's correlation coefficients depends on the graph size and hence might not be the right measure for degree-degree correlations in networks.

In [82] it was proved that under general regularity assumptions, Spearman's rho does converges to a correct limit, which captures the asymptotic correlation between the degrees of connected nodes. Moreover, both the undirected configuration model and preferential attachment model were shown to satisfy these assumptions.

In this chapter we will extend the analysis in [53] and [82] to the setting of directed networks. We will show, similar to what was proven in [82], that Pearson's correlation coefficient r_α^β depends on the size of the graph and converges to a non-negative number under certain conditions on the existence of moments of the degree distributions. These conditions will differ per degree-degree correlation type, although they always include the infinite variance case. In contrast, we will show that under general assumptions on the bi-degree sequences and the joint degree distribution, the rank-correlations measures ρ_α^β , $\bar{\rho}_\alpha^\beta$ and τ_α^β converges to a limit value, which properly reflects the correlations between the α and β degree of connected nodes. These limits are related to Spearman's rho and Kendall's tau for integer-valued random variables.

The structure of this chapter is as follows. We start in Section 5.2 with the analysis of Pearson's correlation coefficients. Then we move to rank-correlations. We will state our main convergence result in Section 5.3 and give an overview of the proof strategy in Section 5.4. The proof can be found in Section 5.5. Finally, we briefly discuss convergence results of rank-correlations under weaker assumptions in Section 5.6.

5.2 Non-negativity of Pearson's correlation coefficient

We will first consider Pearson's α - β coefficient (Definition 4.1), on directed random graphs with regularly-varying degree distributions. Our main result says that, depending on the exponents γ_+ and γ_- of, respectively, the out- and in-degree distribution, $r_\alpha^\beta(G_n)$ converges to a non-negative value as the graph size goes to infinity. In order to state the result we first define the following areas $\mathfrak{G}_\alpha^\beta \subseteq \mathbb{R}^+ \times \mathbb{R}^+$:

$$\begin{aligned}\mathfrak{G}_+^- &= \{1 < \gamma_+ < 3 \text{ and } \gamma_- > 1\} \cup \{1 < \gamma_- < 3 \text{ and } \gamma_+ > 1\}, \\ \mathfrak{G}_-^+ &= \{1 < \gamma_+ < 2 \text{ and } \gamma_- > 1\} \cup \{1 < \gamma_- < 2 \text{ and } \gamma_+ > 1\}, \\ \mathfrak{G}_+^+ &= \{1 < \gamma_+ < 3 \text{ and } \gamma_- > 1\}, \\ \mathfrak{G}_-^- &= \{1 < \gamma_- < 3 \text{ and } \gamma_+ > 1\},\end{aligned}$$

see Figure 5.1 for an illustration.

Theorem 5.1. *Let $(\gamma_+, \gamma_-) \in \mathfrak{G}_\alpha^\beta$ and $\mathcal{D}^+$ and $\mathcal{D}^-$ be independent regularly varying random variables with, respectively, exponent γ_+ and γ_- . Then, for any graph G_n with bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$ and any*

$$0 < \delta < \min \left\{ \frac{|3 - \gamma_+|}{3\gamma_+}, \frac{|3 - \gamma_-|}{3\gamma_-}, \frac{\gamma_+ - 1}{3\gamma_+}, \frac{\gamma_- - 1}{3\gamma_-}, \frac{1}{2} \right\}, \quad (5.1)$$

we have that

$$\lim_{n \rightarrow \infty} \mathbb{P}(r_\alpha^\beta(G_n) > -n^{-\delta}) = 1.$$

An important observation is that the conditions of Theorem 5.1 depend solely on the bi-degree sequence. This means that no matter what the structure of the graph is, as long as the bi-degree sequence satisfies the assumptions of the theorem, Pearson's correlation coefficients will converge to a non-negative value as the graph size increases. In particular, since the range where both $1 < \gamma_+, \gamma_- < 2$ is contained in all four areas, this holds for any graph G_n with degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$, where $\mathcal{D}^+$ and $\mathcal{D}^-$ are independent regularly varying random variables with these exponents γ_+ and γ_- . That is, even when the graph is wired in a disassortative way, this will not, asymptotically, be reflected by the value of r_α^β . This could explain why in most large networks strong disassortativity has not been observed. Another consequence of Theorem 5.1 is that we cannot compare networks of different sizes, for which the degree sequences satisfy the assumptions of the theorem, because in that case, $r_\alpha^\beta(G_n)$ will change with the network size.

For undirected graphs [82, Theorem 3.1] shows that when the the maximum of the degree scales at a certain rate, which holds for degree sequences sampled from a regularly-varying random variable with exponent $1 < \gamma < 3$, then any limit point of Pearson's correlation coefficient is non-negative. Our result, Theorem 5.1, establishes a similar result for directed graphs while also providing a scaling on the lower bound for ρ_α^β .

Figure 5.1: Representation of the four areas $\mathfrak{G}_\alpha^\beta$, where r_α^β converges to a non-negative number. The boundaries are not included.

For the proof of Theorem 5.1 recall, from (4.6), that the negative contribution to r_α^β comes from

$$\tilde{r}_\alpha^\beta(G_n) = \frac{\frac{1}{L_n} \sum_{i=1}^n D_i^+ D_i^\alpha \sum_{j=1}^n D_j^- D_j^\beta}{\sigma_\alpha(G_n) \sigma^\beta(G_n)}, \quad (5.2)$$

where the terms in the denominator are defined in (4.7) and (4.8). Therefore it is enough to prove that, under the conditions in Theorem 5.1,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\tilde{r}_\alpha^\beta(G_n) > n^{-\delta}) = 0. \quad (5.3)$$

To explain the idea behind the proof, let us fix $\alpha, \beta \in \{+, -\}$. Then the main argument for proving (5.3) it to show that when the exponents of the out- and in-degree distribution, γ_+ and γ_- , are contained in the area $\mathfrak{G}_\alpha^\beta$, the terms in the denominator of (5.2) scale at a larger rate than those in the numerator.

For the proof of Theorem 5.1 we need the following technical lemma, which we will prove at the end of this section:

Lemma 5.2. *Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be independent regularly-varying random variables with, respectively, exponent γ_+ and γ_- . Let $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$ and let*

$$0 < \delta < \min \left\{ \frac{|3 - \gamma_+|}{3\gamma_+}, \frac{|3 - \gamma_-|}{3\gamma_-}, \frac{\gamma_+ - 1}{3\gamma_+}, \frac{\gamma_- - 1}{3\gamma_-}, \frac{1}{2} \right\}.$$

Then, if we define the events

$$A_n^\pm = \left\{ \frac{(\sum_{i=1}^n (D_i^\pm)^2)^2}{L_n \sum_{i=1}^n (D_i^\pm)^3} \leq n^{-\delta} \right\} \quad \text{and} \quad B_n^\pm = \left\{ \frac{(\sum_{i=1}^n D_i^+ D_i^-)^2}{L_n \sum_{i=1}^n D_i^\mp (D_i^\pm)^2} \leq n^{-\delta} \right\},$$

we have that

- i) $\lim_{n \rightarrow \infty} \mathbb{P}(A_n^\pm) = 1$ for all $1 < \gamma_\pm < 3$ and $\gamma_\mp > 1$, and
- ii) $\lim_{n \rightarrow \infty} \mathbb{P}(B_n^\pm) = 1$ for all $1 < \gamma_\pm < 2$ and $\gamma_\mp > 1$.

On the event $A_n^\pm$ we have that $L_n \sum_{i=1}^n (D_i^\pm)^3$ scales faster with n than $(\sum_{i=1}^n (D_i^\pm)^2)^2$. As we mentioned, the idea behind the proof of (5.3) is to show that the terms in the denominator scales at a larger rate than those in the numerator. We can now condition on specific combination of the events $A_n^\pm$ and $B_n^\pm$ to ensure this happens, while the probability of these events converges to one by Lemma 5.2.

Proof of Theorem 5.1. Let $\alpha, \beta \in \{+, -\}$ and take $(\gamma_+, \gamma_-) \in \mathfrak{G}_\alpha^\beta$. Define

$$m(\gamma_+, \gamma_-) = \min \left\{ \frac{|3 - \gamma_+|}{3\gamma_+}, \frac{|3 - \gamma_-|}{3\gamma_-}, \frac{\gamma_+ - 1}{3\gamma_+}, \frac{\gamma_- - 1}{3\gamma_-}, \frac{1}{2} \right\},$$

let $0 < \delta < m(\gamma_+, \gamma_-)$ and take $\kappa = (m(\gamma_+, \gamma_-) + \delta)/2$. In addition we define, for any $\alpha', \beta' \in \{+, -\}$

$$K_n^{\alpha', \beta'} = \frac{\left(\sum_{i=1}^n D_i^{\alpha'} D_i^{\beta'} \right)^2}{L_n \sum_{i=1}^n D_i^{\alpha'} (D_i^{\beta'})^2},$$

and observe that

$$\tilde{r}_\alpha^\beta(G_n) = \left(\frac{K_n^{+, \alpha}}{1 - K_n^{+, \alpha}} \right)^{1/2} \left(\frac{K_n^{-, \beta}}{1 - K_n^{-, \beta}} \right)^{1/2}.$$

Therefore we have that, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{P}(\tilde{r}_\alpha^\beta(G_n) > n^{-\delta}) &\leq \mathbb{P}(\tilde{r}_\alpha^\beta(G_n) > n^{-\delta}, K_n^{+, \alpha} \leq n^{-\kappa}, K_n^{-, \beta} \leq n^{-\kappa}) \\ &\quad + \mathbb{P}(K_n^{+, \alpha} > n^{-\kappa}) + \mathbb{P}(K_n^{-, \beta} > n^{-\kappa}) \\ &\leq \mathbb{P}\left(\frac{n^{-\kappa}}{1 - n^{-\kappa}} > n^{-\delta}\right) + \mathbb{P}(K_n^{+, \alpha} > n^{-\kappa}) + \mathbb{P}(K_n^{-, \beta} > n^{-\kappa}) \\ &\leq \mathbb{P}(K_n^{+, \alpha} > n^{-\kappa}) + \mathbb{P}(K_n^{-, \beta} > n^{-\kappa}). \end{aligned} \tag{5.4}$$

Since $\delta < \kappa$ it holds that $n^{-\kappa+\delta} \rightarrow 0$ and all that is left is to show that the other two probabilities converge to zero, as $n \rightarrow \infty$. For this we observe that the events from Lemma 5.2 equal

$$A_n^\pm = \{K^{\pm, \pm} \leq n^{-\delta}\} \quad \text{and} \quad B_n^\pm = \{K^{\mp, \pm} \leq n^{-\delta}\}.$$

Hence, since $\kappa < m(\gamma_+, \gamma_-)$, Lemma 5.2 implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}(K_n^{+, \alpha} \leq n^{-\kappa}) = 1 = \lim_{n \rightarrow \infty} \mathbb{P}(K_n^{-, \beta} \leq n^{-\kappa}),$$

whenever $(\gamma_+, \gamma_-) \in \mathfrak{G}_\alpha^\beta$, which proves the result. □

We now give the proof of Lemma 5.2:

Proof of Lemma 5.2. We will prove the statements for A_n^+ and B_n^+ . The proof of the other two statements follow using similar arguments. Recall that when $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$ then

$$D_i^+ = \mathcal{D}_i^+ + \xi_i^+ \quad \text{and} \quad D_i^- = \mathcal{D}_i^- + \xi_i^-,$$

where $(\mathcal{D}_n^+, \mathcal{D}_n^-)$ are i.i.d. samples from $(\mathcal{D}^+, \mathcal{D}^-)$ and $\xi_i^\pm \in \{0, 1\}$ are indicators, used to make the sums of out- and in-degrees equal. We start by observing that since $\mathcal{D}^+$ and $\mathcal{D}^-$ are independent, $\mathcal{D}^+ \mathcal{D}^-$ is regularly varying with exponent $\gamma = \gamma_+ \wedge \gamma_-$, see for instance [41]. In addition, since $0 \leq \xi_i^\pm \leq 1$, it follows from Proposition 2.4 that for any integers $p, q \geq 0$ and any $\kappa, K > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i=1}^n (D_i^+)^p (D_i^-)^q \leq K n^{\left(\frac{p}{\gamma_+} \vee \frac{q}{\gamma_-} \vee 1\right) + \kappa}\right) = 1, \quad (5.5)$$

and if $(p/\gamma_+) \vee (q/\gamma_-) > 1$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i=1}^n (D_i^+)^p (D_i^-)^q \geq K n^{\left(\frac{p}{\gamma_+} \vee \frac{q}{\gamma_-}\right) - \kappa}\right) = 1. \quad (5.6)$$

Hence, if we define the event

$$C_n = \left\{ \sum_{i=1}^n D_i^+ \geq n^{1 - \frac{\delta}{2}} \right\}, \quad (5.7)$$

then it follows, by (5.6), that $\mathbb{P}(C_n) \rightarrow 1$, since $\gamma_+ > 1$.

For A_n^+ , we define the event

$$M_n = \left\{ \sum_{i=1}^n (D_i^+)^3 \geq n^{\frac{3}{\gamma_+} - \frac{\delta}{2}} \right\},$$

and observe that since $\gamma_+ < 3$ we have that (5.6) implies $\mathbb{P}(M_n) \rightarrow 1$. Because

$$\mathbb{P}(A_n^+) \leq \mathbb{P}(A_n^+, C_n, M_n) + \mathbb{P}(C_n^c) + \mathbb{P}(M_n^c),$$

we only need to analyze $\mathbb{P}(A_n^+, C_n, M_n)$. For this we have

$$\begin{aligned} & \mathbb{P}(A_n^+, C_n, M_n) \\ &= \mathbb{P}\left(\left(\sum_{i=1}^n (D_i^+)^2\right)^2 > n^{-\delta} L_n \sum_{i=1}^n (D_i^+)^3, C_n, M_n\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^n (D_i^+)^2 > n^{-\frac{\delta}{2}} (\mu n - n^{1-\varepsilon})^{\frac{1}{2}} n^{\frac{3}{2\gamma_+} - \frac{\delta}{2}}\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^n (D_i^+)^2 > n^{\frac{1}{2} + \frac{3}{2\gamma_+} - \delta} (\mu - n^{-\varepsilon})^{\frac{1}{2}}\right), \end{aligned}$$

and hence, it follows from (5.5) that it is enough to show that

$$\frac{1}{2} + \frac{3}{2\gamma_+} - \delta > \left(\frac{2}{\gamma_+} \vee 1\right).$$

This follows since, by the assumption on δ ,

$$\frac{1}{2} + \frac{3}{2\gamma_+} - \delta > \frac{\gamma_+ + 3}{2\gamma_+} - \frac{\gamma_+ - 1}{2\gamma_+} > \frac{2}{\gamma_+},$$

and additionally,

$$\frac{1}{2} + \frac{3}{2\gamma_+} - \delta > \frac{\gamma_+ + 3}{2\gamma_+} - \frac{3 - \gamma_+}{2\gamma_+} > 1.$$

Next we will consider the event B_n^+ . We will follow a similar strategy as for the proof of A_n^+ . Let C_n be defined as in (5.7) and define the event M_n now as

$$M_n = \left\{ \sum_{i=1}^n D_i^- (D_i^+)^2 \geq n^{\frac{2}{\gamma_+} - \frac{\delta}{2}} \right\},$$

so that, again by (5.6) and the fact that $1 < \gamma_+ < 2$ and $\gamma_1 > 1$, we have that $\mathbb{P}(M_n) \rightarrow 1$. Since

$$\begin{aligned} & \mathbb{P}(B_n^+, C_n, M_n) \\ &= \mathbb{P}\left(\left(\sum_{i=1}^n D_i^+ D_i^-\right)^2 > n^{-\delta} L_n \sum_{i=1}^n D_i^- (D_i^+)^2, C_n, M_n\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^n D_i^+ D_i^- > n^{\frac{1}{2} + \frac{1}{\gamma_+} - \delta} (\mu - n^{-\varepsilon})^{\frac{1}{2}}\right), \end{aligned}$$

we are left to show that

$$\frac{1}{2} + \frac{1}{\gamma_+} - \delta > \left(\frac{1}{\gamma_+} \vee \frac{1}{\gamma_-}\right).$$

Here, on one hand we have

$$\frac{1}{2} + \frac{1}{\gamma_+} - \delta > \frac{1}{2} + \frac{1}{\gamma_+} - \frac{1}{2} > \frac{1}{\gamma_+},$$

while on the other

$$\frac{1}{2} + \frac{1}{\gamma_+} - \delta > \frac{1}{2} + \frac{1}{2\gamma_-} - \frac{\gamma_- - 1}{2\gamma_-} = \frac{1}{\gamma_-}.$$

□

5.3 Convergence of rank-correlation measures in random graphs

We will now direct our attention to the rank-correlation measures Spearman's rho and Kendall's tau, see Definitions 4.3, 4.4 and 4.5. In this section we will state the main result, Theorem 5.3, which shows that, in contrast to Pearson's correlation coefficients, under very general conditions on the bi-degree sequence and the joint degree distribution, rank-correlation measures converge to a limit value that does properly describe the correlations between the degrees of connected nodes. Before we can give this result we will state the general definitions of Spearman's rho and Kendall's tau for integer valued random variables.

5.3.1 Rank-correlation measures for integer valued random variables

Given two integer-valued random variables X and Y , define the following functions, related to their distributions,

$$\begin{aligned}\mathcal{F}_X(k) &= F_X(k) + F_X(k-1), \\ \mathcal{F}_Y(\ell) &= F_Y(\ell) + F_Y(\ell-1), \\ \mathcal{H}_{X,Y}(k, \ell) &= H_{X,Y}(k, \ell) + H_{X,Y}(k-1, \ell) \\ &\quad + H_{X,Y}(k, \ell-1) + H_{X,Y}(k-1, \ell-1).\end{aligned}$$

These functions arise from the continuization principle, see Section 2.1.4, and are involved in the definition of Spearman's rho and Kendall's tau for integer-valued random variables.

Given two integer-valued random variables X and Y , Spearman's rho $\rho(X, Y)$ is defined as, c.f. [57]

$$\begin{aligned}\rho(X, Y) &= 3(\mathbb{P}(X < X', Y < Y'') + \mathbb{P}(X \leq X', Y < Y'') \\ &\quad + \mathbb{P}(X < X', Y \leq Y'') + \mathbb{P}(X \leq X', Y \leq Y'') - 1),\end{aligned}$$

where (X', Y') and (X'', Y'') are independent copies of (X, Y) . We will rewrite this expression, starting with a single term:

$$\mathbb{P}(X < X', Y < Y'') = \mathbb{E}[\mathbb{1}_{\{X < X'\}} \mathbb{1}_{\{Y < Y''\}}]$$

$$\begin{aligned}
 &= 1 - \mathbb{E} [\mathbb{1}_{\{X' \leq X\}}] - \mathbb{E} [\mathbb{1}_{\{Y'' \leq Y\}}] + \mathbb{E} [\mathbb{1}_{\{X' \leq X\}} \mathbb{1}_{\{Y'' \leq Y\}}] \\
 &= 1 - \mathbb{E} [F_X(X)] - \mathbb{E} [F_Y(Y)] + \mathbb{E} [F_X(X) F_Y(Y)].
 \end{aligned}$$

If we do the same for the other three terms and use that

$$\mathbb{E} [\mathcal{F}_X(X)] = 1 = \mathbb{E} [\mathcal{F}_Y(Y)],$$

see (2.11), we obtain,

$$\rho(X, Y) = 3\mathbb{E} [\mathcal{F}_X(X) \mathcal{F}_Y(Y)] - 3. \quad (5.8)$$

It follows from [57, Proposition 2.2] that Kendall's tau is related to $\mathcal{H}_{X,Y}(X, Y)$ by,

$$\tau(X, Y) = \mathbb{E} [\mathcal{H}_{X,Y}(X, Y)] - 1. \quad (5.9)$$

5.3.2 Main result

We are now ready to address the convergence of rank-correlation measures in random directed graphs. The main idea is to show that, for instance, $\rho_\alpha^\beta(G_n)$ converges to $\rho(D_\alpha, D^\beta)$, for certain integer valued random variables D_α and D^β . For this we first define, similar to $\mathcal{F}$ and $\mathcal{H}$ above,

$$\begin{aligned}
 \mathcal{F}_n^{\alpha,\beta}(k) &= F_n^{\alpha,\beta}(k) + F_n^{\alpha,\beta}(k-1) \\
 \mathcal{H}_n(k, \ell) &= H_n(k, \ell) + H_n(k-1, \ell) + H_n(k, \ell-1) + H_n(k-1, \ell-1).
 \end{aligned}$$

Next, in addition to Assumption 2.2, we also need a regularity assumption on the wiring of the graph. In particular we need the existence of limiting random variables D_α and D^β such that the empirical joint degree probability mass function $h_n^{\alpha,\beta}$ converges to the joint probability mass function of D_α and D^β .

Assumption 5.1. There exists integer-valued random variables D_α and D^β , which are not concentrated in a single point, with joint probability mass function $h^{\alpha,\beta}$ and a $\varepsilon > 0$, such that for any bounded function $K > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{k, \ell=0}^n \left| h_n^{\alpha,\beta}(k, \ell) - h^{\alpha,\beta}(k, \ell) \right| \leq K n^{-\varepsilon} \right) = 1. \quad (5.10)$$

We will use the notation $G_n \in \mathcal{G}_{\eta, \varepsilon}(D_\alpha, D^\beta)$ to indicate that the bi-degree sequence satisfies Assumption 2.2 while G_n satisfies Assumption 5.1, where the random variables D_α and D^β have, respectively, probability mass functions $f^{+, \alpha}$ and $f^{-, \beta}$. In this setting we will slightly abuse notation and let ε denote the minimum of the two versions used in the definitions of Assumption 2.2 and 5.1.

For the convergence of $\bar{\rho}_\alpha^\beta(G_n)$, we will need one final definition to deal with the limit of the terms $\bar{\sigma}_\alpha(G_n)$ and $\bar{\sigma}^\beta(G_n)$ in the denominator. For this we introduce the functions

$$S_\alpha(k) = F^{+, \alpha}(k) F^{+, \alpha}(k-1) \quad \text{and} \quad S^\beta(\ell) = F^{-, \beta}(\ell) F^{-, \beta}(\ell-1), \quad (5.11)$$

and define

$$\bar{\rho}(D^\alpha, D^\beta) = \frac{\rho(D_\alpha, D^\beta)}{3 \sqrt{\mathbb{E} [S_\alpha(D_\alpha)] \mathbb{E} [S^\beta(D^\beta)]}}. \quad (5.12)$$

Observe that when D_α and D^β are not concentrated in a single point, we have that $\mathbb{E}[S_\alpha(D_\alpha)], \mathbb{E}[S^\beta(D^\beta)] > 0$, so that (5.12) is well-defined.

With this, we have the following convergence results:

Theorem 5.3. *Let $\alpha, \beta \in \{+, -\}$ and $G_n \in \mathcal{G}_{n,\varepsilon}(D_\alpha, D^\beta)$. Then, for any $0 < \delta < \min(\varepsilon, 1/2)$ and $K > 0$,*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(|\rho_\alpha^\beta(G_n) - \rho(D^\alpha, D^\beta)| \leq Kn^{-\delta}) &= 1, \\ \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{\rho}_\alpha^\beta(G_n) - \bar{\rho}(D^\alpha, D^\beta)| \leq Kn^{-\delta}) &= 1, \\ \lim_{n \rightarrow \infty} \mathbb{P}(|\tau_\alpha^\beta(G_n) - \tau(D^\alpha, D^\beta)| \leq Kn^{-\delta}) &= 1. \end{aligned}$$

Moreover, if we define the event

$$\mathcal{J}_n^{\alpha,\beta} = \left\{ \sum_{k,\ell=0}^{\infty} |h^{\alpha,\beta}(k, \ell) - h_n^{\alpha,\beta}(k, \ell)| > \frac{Kn^{-\delta}}{16} \right\}, \quad (5.13)$$

and take $\kappa = \min(1 - 2\delta, \varepsilon - \delta)$, then, as $n \rightarrow \infty$,

- i) $\mathbb{P}(|\rho_\alpha^\beta(G_n) - \rho(D^\alpha, D^\beta)| > Kn^{-\delta}) \leq O(n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta}))$,
- ii) $\mathbb{P}(|\bar{\rho}_\alpha^\beta(G_n) - \bar{\rho}(D^\alpha, D^\beta)| > Kn^{-\delta}) \leq O(n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta}))$,
- iii) $\mathbb{P}(|\tau_\alpha^\beta(G_n) - \tau(D^\alpha, D^\beta)| > Kn^{-\delta}) \leq O(n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta}))$.

Before we proof Theorem 5.3, let us explain why rank-correlation measures are a better tool for the analysis of degree-degree correlation in directed networks, by comparing Theorem 5.3 to Theorem 5.1.

Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be two independent regularly-varying random variables with exponents $1 < \gamma_\pm < 2$, and consider a bi-degree sequence generated by the IID algorithm, $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$. If we now generate graphs with this bi-degree sequence that satisfy Assumption 5.1, then by Theorem 5.3 the rank-correlation measures $\rho_\alpha^\beta(G_n)$, $\bar{\rho}_\alpha^\beta(G_n)$ and $\tau_\alpha^\beta(G_n)$ will converge to a limit that is determined by the joint distribution of the limiting random variables D_α and D^β . On the other hand, by Theorem 5.1, Pearson's correlation coefficients $r_\alpha^\beta(G_n)$ will, in this case, always converge to a non-negative value, independent of the relation between D_α and D^β . As a consequence, the value of Pearson's correlation coefficient does not reflect the degree-degree correlations in these graphs.

Now let $\mathcal{J}_n^{\alpha,\beta}$ be as defined in Theorem 5.3. Then, since rank-correlations are uniformly bounded in absolute value and

$$|\rho_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| \leq 2,$$

we have, using Theorem 5.3 i), that

$$\begin{aligned} \mathbb{E}[|\rho_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)|] &\leq n^{-\delta} + 2\mathbb{P}(|\rho_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| > n^{-\delta}) \\ &\leq O(n^{-\delta} + n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}((\mathcal{J}_n^{\alpha,\beta})^c)), \end{aligned}$$

as $n \rightarrow \infty$. Using similar arguments for $\bar{\rho}_\alpha^\beta$ and τ_α^β , we obtain convergence in expectation, as a corollary to Theorem 5.3.

Corollary 5.4. *Let $\alpha, \beta \in \{+, -\}$ and $G_n \in \mathcal{G}_{n,\varepsilon}(D_\alpha, D^\beta)$. Then, for any $0 < \delta < \min(\varepsilon, 1/2)$ and $\kappa = \min(1 - \delta, \varepsilon - \delta)$, as $n \rightarrow \infty$,*

- i) $\mathbb{E} [|\rho_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)|] \leq O(n^{-\delta} + n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta}))$
- ii) $\mathbb{E} [|\bar{\rho}_\alpha^\beta(G_n) - \bar{\rho}(D_\alpha, D^\beta)|] \leq O(n^{-\delta} + n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta}))$
- iii) $\mathbb{E} [|\tau_\alpha^\beta(G_n) - \tau(D_\alpha, D^\beta)|] \leq O(n^{-\delta} + n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta})),$

where $\mathcal{J}_n^{\alpha,\beta}$ is defined as in Theorem 5.3.

5.4 Road map for the proof of Theorem 5.3

In this section we will explain the ideas behind the proof of Theorem 5.3 and introduce the steps we take to reach the final result.

Recall that for a directed graph G_n , we interpret the joint α - β degree vector $(D_i^\alpha, D_j^\beta)_{i \rightarrow j}$ as L_n measurements of the α - β degree-degree correlation. Therefore, similar to the Pearson's α - β correlation coefficient, we should have that rank-correlation measures $\rho_\alpha^\beta(G_n)$ and $\bar{\rho}_\alpha^\beta(G_n)$ are related to empirical versions of Spearman's rho, as defined in Section 5.3.1. We will show that when the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfies Assumption 2.2 this is indeed the case.

Let G_n be a directed graph, then

$$\frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) = \sum_{k, \ell=0}^{\infty} \mathcal{F}_n^{+, \alpha}(k) \mathcal{F}_n^{-, \beta}(\ell) h_n^{\alpha, \beta}(k, \ell), \quad (5.14)$$

where $h_n^{\alpha, \beta}$ denotes the empirical probability mass function of the α - and β -degree of, respectively, the source and target of an edge in G_n , see (2.15). In light of Assumption 5.1, this expression can be interpreted as an empirical version of $\mathbb{E} [\mathcal{F}^{+, \alpha}(D_\alpha) \mathcal{F}^{-, \beta}(D^\beta)]$, where $\mathcal{F}^{+, \alpha}$ and $\mathcal{F}^{-, \beta}$ are defined in a similar way as, for instance, $\mathcal{F}_n^{+, \alpha}$. We will indeed show (Proposition 5.5) that under Assumption 5.1

$$\frac{3}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) - 3 \xrightarrow{\mathbb{P}} \rho(D_\alpha, D^\beta), \quad (5.15)$$

as $n \rightarrow \infty$.

Theorem 5.3 i)

We will show (Lemma 5.7) that

$$\left| \frac{4}{L_n^3} \sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j) \bar{R}^\beta(i \rightarrow j) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) \right| \xrightarrow{\mathbb{P}} 0.$$

Since

$$\left| \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) - \frac{3}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) + 3 \right|$$

$$\leq \left| \frac{4}{L_n^3} \sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j) \bar{R}^\beta(i \rightarrow j) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) \right| + \left| \frac{3L_n(L_n + 1)^2}{L_n^3} - 3 \right|,$$

it now follows, using (5.15), that under Assumption 5.1

$$\frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \rho(D_\alpha, D^\beta). \quad (5.16)$$

For the last step we show (Proposition 5.6) that

$$\left| \rho_\alpha^\beta(G_n) - \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) \right| \xrightarrow{\mathbb{P}} 0,$$

which then implies that

$$\rho_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \rho(D_\alpha, D^\beta).$$

Theorem 5.3 ii)

In order to prove Theorem 5.3 ii) we first establish (Lemma 5.8) that under Assumption 5.1

$$\frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} \xrightarrow{\mathbb{P}} \mathbb{E}[S_\alpha(D_\alpha)] \quad \text{and} \quad \frac{\bar{\sigma}^\beta(G_n)^2}{L_n^3} \xrightarrow{\mathbb{P}} \mathbb{E}[S^\beta(D^\beta)].$$

Since, by Assumption 5.1, the random variables D_α and D^β are not concentrated in one point we have that $\mathbb{E}[S_\alpha(D_\alpha)], \mathbb{E}[S^\beta(D^\beta)] \neq 0$ and hence

$$\frac{L_n^3 \sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]}}{\bar{\sigma}_\alpha(G_n) \bar{\sigma}^\beta(G_n)} \xrightarrow{\mathbb{P}} 1.$$

Using (5.16) we then obtain

$$\begin{aligned} \bar{\rho}_\alpha^\beta(G_n) &= \left(\frac{L_n^3 \sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]}}{\bar{\sigma}_\alpha(G_n) \bar{\sigma}^\beta(G_n)} \right) \frac{\bar{\sigma}_\alpha(G_n) \bar{\sigma}^\beta(G_n) \bar{\rho}_\alpha^\beta(G_n)}{L_n^3 \sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]}} \\ &\xrightarrow{\mathbb{P}} \frac{\rho(D_\alpha, D^\beta)}{\sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]}} \\ &= \bar{\rho}(D_\alpha, D^\beta), \end{aligned}$$

which establishes the result.

Theorem 5.3 iii)

Here we will first show (Proposition 5.9) that

$$\left| \tau_\alpha^\beta(G_n) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) + 1 \right| \xrightarrow{\mathbb{P}} 0.$$

Noting that,

$$\frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) - 1 = \sum_{k, \ell}^{\infty} \mathcal{H}_n^{\alpha, \beta}(k, \ell) h_n^{\alpha, \beta}(k, \ell) - 1$$

we then prove that under Assumption 5.1,

$$\frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(k, \ell) - 1 \xrightarrow{\mathbb{P}} \sum_{k, \ell=0}^{\infty} \mathcal{H}_n^{\alpha, \beta}(k, \ell) h_n^{\alpha, \beta}(k, \ell) - 1 = \tau(D_\alpha, D^\beta).$$

5.5 Proof of the main result

We will first establish (5.15) with the appropriate scaling terms. Let G_n be a directed graph and define, for any $\alpha, \beta \in \{+, -\}$,

$$\tilde{\rho}_\alpha^\beta(G_n) = \frac{3}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) - 3. \quad (5.17)$$

Then we have the following result

Proposition 5.5. *Let $G_n \in \mathcal{G}_{\eta, \varepsilon}$, let $\alpha, \beta \in \{+, -\}$, $0 < \delta < 1$, $K > 0$ and define the event*

$$\mathcal{J}_n^{\alpha, \beta} = \left\{ \sum_{k, \ell=0}^{\infty} |h^{\alpha, \beta}(k, \ell) - h_n^{\alpha, \beta}(k, \ell)| > K \frac{n^{-\delta}}{12} \right\}.$$

Then, if we take $\kappa = \min(1 - 2\delta, \varepsilon - \delta)$,

$$\mathbb{P}(|\tilde{\rho}_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| \leq K n^{-\delta}) \geq 1 - O(n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha, \beta})),$$

as $n \rightarrow \infty$.

Proof. By definition of $\rho(D_\alpha, D^\beta)$ we have

$$\begin{aligned} & |\tilde{\rho}_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| \\ &= 3 \left| \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) - \mathbb{E}[\mathcal{F}^{+, \alpha}(D_\alpha) \mathcal{F}^{-, \beta}(D^\beta)] \right| \\ &= 3 \left| \sum_{k, \ell=0}^{\infty} \mathcal{F}_n^{+, \alpha}(k) \mathcal{F}_n^{-, \beta}(\ell) h_n^{\alpha, \beta}(k, \ell) - \mathcal{F}^{+, \alpha}(k) \mathcal{F}^{-, \beta}(\ell) h^{\alpha, \beta}(k, \ell) \right| \\ &\leq 3 \left| \sum_{k, \ell=0}^{\infty} \mathcal{F}^{+, \alpha}(k) \mathcal{F}^{-, \beta}(\ell) (h_n^{\alpha, \beta}(k, \ell) - h^{\alpha, \beta}(k, \ell)) \right| \end{aligned} \quad (5.18)$$

$$+ 3 \left| \sum_{k, \ell=0}^{\infty} (\mathcal{F}_n^{+, \alpha}(k) \mathcal{F}_n^{-, \beta}(\ell) - \mathcal{F}^{+, \alpha}(k) \mathcal{F}^{-, \beta}(\ell)) h_n^{\alpha, \beta}(k, \ell) \right|. \quad (5.19)$$

The first term (5.18) is bounded from above by

$$12 \sum_{k, \ell=0}^{\infty} |h_n^{\alpha, \beta}(k, \ell) - h^{\alpha, \beta}(k, \ell)|,$$

while for (5.19) we have that

$$\begin{aligned} & 3 \left| \sum_{k, \ell=0}^{\infty} \left(\mathcal{F}_n^{+, \alpha}(k) \mathcal{F}_n^{-, \beta}(\ell) - \mathcal{F}^{+, \alpha}(k) \mathcal{F}^{-, \beta}(\ell) \right) h_n^{\alpha, \beta}(k, \ell) \right| \\ & \leq 3 \sup_{k, \ell \geq 0} |\mathcal{F}_n^{+, \alpha}(k) \mathcal{F}_n^{-, \beta}(\ell) - \mathcal{F}^{+, \alpha}(k) \mathcal{F}^{-, \beta}(\ell)| \\ & \leq 6 \sup_{k \geq 0} |\mathcal{F}_n^{+, \alpha}(k) - \mathcal{F}^{+, \alpha}(k)| + 6 \sup_{\ell \geq 0} |\mathcal{F}_n^{-, \beta}(\ell) - \mathcal{F}^{-, \beta}(\ell)| \\ & \leq 12 \sup_{k \geq 0} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| + 12 \sup_{\ell \geq 0} |f_n^{-, \beta}(\ell) - f^{-, \beta}(\ell)| \\ & \leq 12 \sum_{k=0}^{\infty} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| + 12 \sum_{\ell=0}^{\infty} |f_n^{-, \beta}(\ell) - f^{-, \beta}(\ell)|. \end{aligned}$$

We therefore get that, as $n \rightarrow \infty$,

$$\begin{aligned} & \mathbb{P}(|\tilde{\rho}_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| > Kn^{-\delta}) \\ & \leq \mathbb{P}\left(\sum_{k, \ell=0}^{\infty} |h_n^{\alpha, \beta}(k, \ell) - h^{\alpha, \beta}(k, \ell)| > \frac{Kn^{-\delta}}{12}\right) \\ & \quad + \mathbb{P}\left(\sum_{k=0}^{\infty} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| > \frac{Kn^{-\delta}}{12}, \Omega_n\right) + \mathbb{P}(\Omega_n^c) \\ & \quad + \mathbb{P}\left(\sum_{\ell=0}^{\infty} |f_n^{-, \beta}(\ell) - f^{-, \beta}(\ell)| > \frac{Kn^{-\delta}}{12}, \Omega_n\right) + \mathbb{P}(\Omega_n^c) \\ & \leq \frac{12n^\delta}{K} \mathbb{E}\left[\sum_{k=0}^{\infty} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| \mathbb{1}_{\Omega_n}\right] \\ & \quad + \frac{12n^\delta}{K} \mathbb{E}\left[\sum_{\ell=0}^{\infty} |f_n^{-, \beta}(\ell) - f^{-, \beta}(\ell)| \mathbb{1}_{\Omega_n}\right] \\ & \quad + 2\mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha, \beta}) \\ & \leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha, \beta})), \end{aligned}$$

from which the result follows. $\square$

5.5.1 Proof of Theorem 5.3 i)

We already saw in Chapter 4 that for non-random graphs,

$$\mathbb{E}[\rho_\alpha^\beta(G_n)] = \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3 - L_n} \bar{\rho}_\alpha^\beta(G_n).$$

In the setting of random graphs we can obtain the following, similar, result.

Proposition 5.6. *Let $G_n \in \mathcal{G}_{\eta, \varepsilon}$ and let $\alpha, \beta \in \{+, -\}$. Then for any $0 < \delta < 1/2$ and $K > 0$,*

$$\mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) - \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) \right| \leq Kn^{-\delta} \right) \geq 1 - O(n^{-1+2\delta} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$, where $\bar{\sigma}_\alpha(G_n)$ and $\bar{\sigma}^\beta(G_n)$ are defined as in Definition 4.4.

The proof of Proposition 5.6 is postponed till the end of this section. We will first use the result to prove Theorem 5.3 i).

Proof of Theorem 5.3 i). We will show that for any $0 < \delta < \min(\varepsilon, 1/2)$ and $\kappa = \min(1 - 2\delta, \varepsilon - \delta)$, as $n \rightarrow \infty$

$$\mathbb{P}(|\rho_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| > Kn^{-\delta}) \leq O(n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha, \beta})). \quad (5.20)$$

First note that

$$\frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) = \frac{3}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) - \frac{3(L_n - 1)^2}{L_n^2 - 1}. \quad (5.21)$$

Hence, the strategy for proving (5.20) is as follows. We will use Proposition 5.6 and (5.30) to transition from $\rho_\alpha^\beta(G_n)$ to the term

$$\frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) = \frac{3}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \mathcal{F}_n^{-, \beta}(D_j^\beta) - \frac{3(L_n - 1)^2}{L_n^2 - 1}.$$

Then we will compare this to $\tilde{\rho}_\alpha^\beta(G_n)$ and use Proposition 5.5 for the final convergence.

With these ideas in mind, we bound the left hand side of (5.20) as follows:

$$\begin{aligned} & \mathbb{P}(|\rho_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| > Kn^{-\delta}) \\ & \leq \mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) - \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) \right| > Kn^{-\delta} \right) \\ & \quad + \mathbb{P} \left(\left| \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta) \right| > Kn^{-\delta} \right) \\ & \leq \mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) - \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) \right| > Kn^{-\delta} \right) \end{aligned} \quad (5.22)$$

$$+ \mathbb{P}(|\tilde{\rho}_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| > Kn^{-\delta}) \quad (5.23)$$

$$+ \mathbb{P} \left(3 \left| \frac{(L_n + 1)^2}{L_n^2 - 1} - 1 \right| > Kn^{-\delta} \right). \quad (5.24)$$

Proposition 5.6 gives that (5.22) is $O(n^{-1+2\delta} + \mathbb{P}(\Omega_n^c))$ while, due to Proposition 5.5, we have that (5.23) is $O(n^{-1+\delta} + \mathbb{P}(\mathcal{J}_n^{\alpha, \beta}) + \mathbb{P}(\Omega_n^c))$, with $\mathcal{J}_n^{\alpha, \beta}$ as defined in (5.13). Finally, for (5.24) it holds

$$\mathbb{P} \left(3 \left| \frac{(L_n + 1)^2}{L_n^2 - 1} - 1 \right| > n^{-\delta} \right) \leq \mathbb{P} \left(\frac{6}{L_n - 1} > n^{-\delta}, \Omega_n \right) + \mathbb{P}(\Omega_n^c)$$

$$\begin{aligned} &\leq 6n^\delta \mathbb{E} \left[\frac{1}{L_n - 1} \mathbb{1}_{\Omega_n} \right] + \mathbb{P}(\Omega_n^c) \\ &\leq O(n^{-1+\delta} + \mathbb{P}(\Omega_n^c)), \end{aligned}$$

as $n \rightarrow \infty$. Therefore we conclude that, as $n \rightarrow \infty$,

$$\mathbb{P}(|\rho_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta)| > Kn^{-\delta}) = O(n^{-1+\delta} + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta}) + \mathbb{P}(\Omega_n^c)),$$

which proves (5.20). $\square$

Before we prove Proposition 5.6 we give the following lemma, which we will need for the proof:

Lemma 5.7. *Let $G_n \in \mathcal{G}_{n,\varepsilon}$ and let $\alpha, \beta \in \{+, -\}$. For any $0 < \delta < 1$ and $K > 0$ define the event*

$$\overline{\mathcal{R}}_n^{\alpha,\beta} = \left\{ \left| \frac{4}{L_n^3} \sum_{i \rightarrow j} \overline{R}_\alpha(i \rightarrow j) \overline{R}^\beta(i \rightarrow j) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+,\alpha}(D_i^\alpha) \mathcal{F}_n^{-,\beta}(D_j^\beta) \right| > Kn^{-\delta} \right\}.$$

Then

$$\mathbb{P}(\overline{\mathcal{R}}_n^{\alpha,\beta}) \leq O(n^{-1+\delta} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$

Proof. From (4.13) it follows that

$$\begin{aligned} \frac{\overline{R}_\alpha(i, j)}{L_n} &= \frac{1}{2L_n} + \frac{1}{L_n} \sum_{s \rightarrow t} \left(\mathbb{1}_{\{D_s^\alpha > D_i^\alpha\}} + \frac{1}{2} \mathbb{1}_{\{D_s^\alpha = D_i^\alpha\}} \right) \\ &= 1 + \frac{1}{2L_n} - \frac{1}{2L_n} \sum_{s \rightarrow t} \left(\mathbb{1}_{\{D_s^\alpha \leq D_i^\alpha\}} + \mathbb{1}_{\{D_s^\alpha \leq D_i^\alpha - 1\}} \right) \\ &= 1 + \frac{1}{2L_n} - \frac{1}{2L_n} \sum_{s=1}^n D_s^+ \mathbb{1}_{\{D_s^\alpha \leq D_i^\alpha\}} - \frac{1}{2L_n} \sum_{s=1}^n D_s^+ \mathbb{1}_{\{D_s^\alpha \leq D_i^\alpha - 1\}} \\ &= 1 + \frac{1}{2L_n} - \frac{1}{2} \mathcal{F}_n^{+,\alpha}(D_i^\alpha). \end{aligned}$$

Using similar calculations for $\overline{R}_\alpha(i, j)/L_n$ we have,

$$\frac{\overline{R}_\alpha(i, j)}{L_n} = 1 + \frac{1}{2L_n} - \frac{1}{2} \mathcal{F}_n^{+,\alpha}(D_i^\alpha), \quad (5.25)$$

$$\frac{\overline{R}^\beta(i, j)}{L_n} = 1 + \frac{1}{2L_n} - \frac{1}{2} \mathcal{F}_n^{-,\beta}(D_j^\beta). \quad (5.26)$$

With these identities we obtain that,

$$\left| \frac{1}{L_n} \sum_{i \rightarrow j} \frac{\overline{R}_\alpha(i, j)}{L_n} \frac{\overline{R}^\beta(i, j)}{L_n} - \frac{1}{4L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+,\alpha}(D_i^\alpha) \mathcal{F}_n^{-,\beta}(D_j^\beta) \right|$$

$$\leq \left| 1 - \frac{1}{2L_n} \sum_{i \rightarrow j} \left(\mathcal{F}_n^{+, \alpha} (D_i^\alpha) + \mathcal{F}_n^{-, \beta} (D_j^\beta) \right) \right| \quad (5.27)$$

$$+ \left| \frac{1}{2L_n^2} \sum_{i \rightarrow j} \left(\mathcal{F}_n^{+, \alpha} (D_i^\alpha) + \mathcal{F}_n^{-, \beta} (D_j^\beta) \right) \right|. \quad (5.28)$$

Now, recall that

$$\sum_{i \rightarrow j} \bar{R}_\alpha(i, j) = \frac{L_n(L_n + 1)}{2} = \sum_{i \rightarrow j} \bar{R}^\beta(i, j),$$

so that by (5.25) and (5.26),

$$\frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha} (D_i^\alpha) = 1 = \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{-, \beta} (D_j^\beta), \quad (5.29)$$

which implies that (5.27) equals zero, while (5.28) equals $1/L_n$. Since this implies that

$$\bar{\mathcal{R}}_n^{\alpha, \beta} = \left\{ \frac{4}{L_n} > Kn^{-\delta} \right\},$$

the result now follows by Markov's inequality,

$$\begin{aligned} \mathbb{P} \left(\bar{\mathcal{R}}_n^{\alpha, \beta} \right) &\leq \mathbb{P} \left(\frac{2}{L_n} > Kn^{-\delta}, \Omega_n \right) + \mathbb{P} (\Omega_n^c) \\ &\leq \frac{2n^\delta}{K} \mathbb{E} \left[\frac{1}{L_n} \mathbb{1}_{\Omega_n} \right] + \mathbb{P} (\Omega_n^c) \\ &\leq \frac{2n^\delta}{K(\nu n - n^{1-\varepsilon})} + \mathbb{P} (\Omega_n^c) \\ &\leq O \left(n^{-1+\delta} + \mathbb{P} (\Omega_n^c) \right). \end{aligned}$$

□

Proof of Proposition 5.6. First we observe that

$$\frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) = \frac{12}{L_n^3} \sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j) \bar{R}^\beta(i \rightarrow j) - \frac{3L_n(L_n + 1)^2}{L_n^3}. \quad (5.30)$$

Therefore

$$\begin{aligned} &\left| \rho_\alpha^\beta(G_n) - \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) \right| \\ &= \left| \frac{12}{L_n^3} \sum_{i \rightarrow j} R_\alpha(i \rightarrow j) R^\beta(i \rightarrow j) - \bar{R}_\alpha(i \rightarrow j) \bar{R}^\beta(i \rightarrow j) \right| \\ &\quad + \left| \frac{3L_n(L_n + 1)^2}{L_n^3 - L_n} - \frac{3L_n(L_n + 1)^2}{L_n^3} \right| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{12}{L_n^3} \left| \sum_{i \rightarrow j} \bar{R}^\beta(i \rightarrow j) (R_\alpha(i \rightarrow j) - \bar{R}_\alpha(i \rightarrow j)) \right| \\
&\quad + \frac{12}{L_n^3} \left| \sum_{i \rightarrow j} R_\alpha(i \rightarrow j) (R^\beta(i \rightarrow j) - \bar{R}^\beta(i \rightarrow j)) \right| \\
&\quad + \frac{3(L_n + 1)^2}{L_n^3(L_n - 1)}.
\end{aligned}$$

Next we compute that

$$\begin{aligned}
\mathbb{P} \left(\frac{3(L_n + 1)^2}{L_n^3(L_n - 1)} > \frac{Kn^{-\delta}}{3}, \Omega_n \right) &\leq \frac{9n^\delta}{K} \mathbb{E} \left[\frac{(L_n + 1)^2}{L_n^3(L_n - 1)} \mathbb{1}_{\Omega_n} \right] \\
&\leq \frac{9n^\delta}{K} \mathbb{E} \left[\frac{(\nu n + n^{1-\varepsilon} + 1)^2}{(\nu n - n^{1-\varepsilon})^3(\nu n + n^{1-\varepsilon} - 1)} \mathbb{1}_{\Omega_n} \right] \\
&\leq O(n^{-2+\delta}).
\end{aligned}$$

Therefore, if we define the events

$$A_n^{\alpha, \beta} = \left\{ \frac{12}{L_n^3} \left| \sum_{i \rightarrow j} \bar{R}^\beta(i \rightarrow j) (R_\alpha(i \rightarrow j) - \bar{R}_\alpha(i \rightarrow j)) \right| \leq \frac{Kn^{-\delta}}{3} \right\}$$

and

$$B_n^{\alpha, \beta} = \left\{ \frac{12}{L_n^3} \left| \sum_{i \rightarrow j} R_\alpha(i \rightarrow j) (R^\beta(i \rightarrow j) - \bar{R}^\beta(i \rightarrow j)) \right| \leq \frac{Kn^{-\delta}}{3} \right\},$$

we have

$$\begin{aligned}
&\mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) - \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) \right| > Kn^{-\delta} \right) \\
&\leq \mathbb{P}(A_n^{\alpha, \beta}, \Omega_n) + \mathbb{P}(B_n^{\alpha, \beta}, \Omega_n) + \mathbb{P} \left(\frac{1}{L_n^2} > \frac{Kn^{-\delta}}{3}, \Omega_n \right) + \mathbb{P}(\Omega_n^c).
\end{aligned}$$

Hence, it suffices to show that, as $n \rightarrow \infty$,

$$\mathbb{P}(A_n^{\alpha, \beta}, \Omega_n) \geq 1 - O(n^{-1+2\delta})$$

and

$$\mathbb{P}(B_n^{\alpha, \beta}, \Omega_n) \geq 1 - O(n^{-1+2\delta}).$$

We will prove the first statement, since the proof of the second statement is similar. Let $i \rightarrow j$ and $s \rightarrow t$ be two edges and define the random variable $Y_{s \rightarrow t}(i \rightarrow j)$ by

$$Y_{s \rightarrow t}(i \rightarrow j) = \mathbb{1}_{\{D_s^\alpha = D_i^\alpha\}} \left(\mathbb{1}_{\{U_{s \rightarrow t} > U_{i \rightarrow j}\}} - \frac{1}{2} \right).$$

Then we have

$$R^\alpha(i \rightarrow j) = \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha > D_t^\alpha\}} + \frac{1}{2} + \sum_{s \rightarrow t \neq i \rightarrow j} \mathbb{1}_{\{D_s^\alpha = D_t^\alpha\}} \mathbb{1}_{\{U_{s \rightarrow t} > U_{i \rightarrow j}\}},$$

so that

$$\begin{aligned} & \left| \frac{1}{L_n^3} \sum_{i \rightarrow j} \bar{R}^\beta(i \rightarrow j) \left(R^\alpha(i \rightarrow j) - \bar{R}^\alpha(i \rightarrow j) \right) \right| \\ &= \frac{1}{L_n^3} \left| \sum_{i \rightarrow j} \sum_{s \rightarrow t \neq i \rightarrow j} \bar{R}^\beta(i \rightarrow j) Y_{s \rightarrow t}(i \rightarrow j) \right|. \end{aligned}$$

Now observe that since $U_{i \rightarrow j}$ and $U_{s \rightarrow t}$ are independent for all $s \rightarrow t \neq i \rightarrow j$ and the $U_{i \rightarrow j}$ are independent of the bi-degree sequence and wiring of the graph, it follows that

$$\mathbb{E}_n [Y_{s \rightarrow t}(i \rightarrow j) Y_{s' \rightarrow t'}(i' \rightarrow j')] = 0, \quad (5.31)$$

whenever the four edges $i \rightarrow j$, $i' \rightarrow j'$, $s \rightarrow t$ and $s' \rightarrow t'$ are all distinct. In particular

$$\mathbb{E} \left[\sum_{i \rightarrow j} \sum_{s \rightarrow t \neq i \rightarrow j} \bar{R}^\beta(i \rightarrow j) Y_{s \rightarrow t}(i \rightarrow j) \right] = 0.$$

Moreover, since $\bar{R}^\beta(i \rightarrow j) \leq L_n$, we have that

$$\begin{aligned} & \mathbb{E} \left[\left| \left(\frac{1}{L_n^3} \sum_{i \rightarrow j} \sum_{s \rightarrow t \neq i \rightarrow j} \bar{R}^\beta(i \rightarrow j) Y_{s \rightarrow t}(i \rightarrow j) \right)^2 \right| \Omega_n \right] \\ & \leq \mathbb{E} \left[\left| \left(\frac{1}{L_n^2} \sum_{i \rightarrow j} \sum_{s \rightarrow t \neq i \rightarrow j} Y_{s \rightarrow t}(i \rightarrow j) \right)^2 \right| \Omega_n \right] \\ & = \mathbb{E} \left[\frac{1}{L_n^4} \sum_{i \rightarrow j} \sum_{i' \rightarrow j'} \sum_{s \rightarrow t \neq i \rightarrow j} \sum_{s' \rightarrow t' \neq i' \rightarrow j'} Y_{s \rightarrow t}(i \rightarrow j) Y_{s' \rightarrow t'}(i' \rightarrow j') \right| \Omega_n \right]. \quad (5.32) \end{aligned}$$

To analyze the last term we consider the different cases depending on the edges $i \rightarrow j$, $i' \rightarrow j'$, $s \rightarrow t$ and $s' \rightarrow t'$. The case where they are all distinct equals zero, due to (5.31). The other cases give at least one less summation so that the expression can be bounded from above by L_n^3 . For instance, when $i \rightarrow j = i' \rightarrow j'$ we have

$$\begin{aligned} & \frac{1}{L_n^4} \sum_{i \rightarrow j} \sum_{s \rightarrow t \neq i \rightarrow j} \sum_{s' \rightarrow t' \neq i \rightarrow j} Y_{s \rightarrow t}(i \rightarrow j) Y_{s' \rightarrow t'}(i \rightarrow j) \\ & \leq \frac{1}{L_n^4} \sum_{i \rightarrow j} \sum_{s \rightarrow t \neq i \rightarrow j} \sum_{s' \rightarrow t' \neq i \rightarrow j} 1 \leq \frac{1}{L_n}, \end{aligned}$$

where we use the fact that $|Y_{s \rightarrow t}(i \rightarrow j)|, |Y_{s' \rightarrow t'}(i \rightarrow j)| \leq 1$. Since, on the event Ω_n , we have $L_n^{-1} \leq (\nu n - n^{1-\varepsilon})^{-1}$ it follows that (5.32) is $O(n^{-1})$, as $n \rightarrow \infty$. Hence, using Chebyshev's inequality, we obtain

$$\begin{aligned} & \mathbb{P} \left(\frac{1}{L_n^3} \left| \sum_{i \rightarrow j} \bar{R}^\beta(i \rightarrow j) (R_\alpha(i, j) - \bar{R}_\alpha(i, j)) \right| > K n^{-\delta} \middle| \Omega_n \right) \\ & \leq \frac{n^{2\delta}}{K^2} \mathbb{E} \left[\left(\frac{1}{L_n^3} \sum_{i \rightarrow j} \sum_{s \rightarrow t} \bar{R}^\beta(i \rightarrow j) \mathbb{1}_{\{s \rightarrow t \neq i \rightarrow j\}} Y_{s \rightarrow t}(i \rightarrow j) \right)^2 \middle| \Omega_n \right] \\ & \leq O(n^{-1+2\delta}), \end{aligned}$$

from which the result follows $\square$

5.5.2 Proof of Theorem 5.3 ii)

As mentioned in Section 5.4 we need the following result, in order to prove Theorem 5.3 ii):

Lemma 5.8. *Let $G_n \in \mathcal{G}_{n,\varepsilon}(D_\alpha, D^\beta)$, let $\alpha, \beta \in \{+, -\}$ and let S_α and S^β be defined as in (5.11). Then for any $0 < \delta < \varepsilon$ and $K > 0$,*

$$\mathbb{P} \left(\left| \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} - \mathbb{E}[S_\alpha(D_\alpha)] \right| \leq K n^{-\delta} \right) \geq 1 - O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c))$$

and

$$\mathbb{P} \left(\left| \frac{\bar{\sigma}^\beta(G_n)^2}{L_n^3} - \mathbb{E}[S^\beta(D^\beta)] \right| \leq K n^{-\delta} \right) \geq 1 - O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$,

Proof of Theorem 5.3 ii). Let $\alpha, \beta \in \{+, -\}$, $0 < \delta < \min(\varepsilon, 1/2)$ and $K > 0$ and let $\kappa = \min(1 - 2\delta, \varepsilon - \delta)$. Then we observe that by Proposition 5.6 and Theorem 5.3 i) it follows that if we define the event

$$A_n = \left\{ \left| \frac{3\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} \bar{\rho}_\alpha^\beta(G_n) - \rho(D_\alpha, D^\beta) \right| > K n^{-\delta} \right\},$$

then

$$\mathbb{P}(A_n) \leq O(n^{-\kappa} + \mathbb{P}(\Omega_n^c) + \mathbb{P}(\mathcal{J}_n^{\alpha,\beta})),$$

as $n \rightarrow \infty$.

Therefore, if we define the event

$$B_n = \left\{ \left| \frac{\bar{\sigma}_\alpha(G_n)\bar{\sigma}^\beta(G_n)}{L_n^3} - \sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]} \right| > K n^{-\delta} \right\},$$

then, in order to prove Theorem 5.3 ii), it is enough to show that

$$\mathbb{P}(B_n) \leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)). \quad (5.33)$$

Since D_α and D^β are not concentrated in a single point,

$$\sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]} > 0,$$

so that we have

$$\begin{aligned} & \left| \frac{\bar{\sigma}_\alpha(G_n) \bar{\sigma}^\beta(G_n)}{L_n^3} - \sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]} \right| \\ & \leq \sqrt{\mathbb{E}[S^\beta(D^\beta)]} \left| \frac{\bar{\sigma}_\alpha(G_n)}{L_n^{3/2}} - \sqrt{\mathbb{E}[S_\alpha(D_\alpha)]} \right| \\ & \quad + \left| \frac{\bar{\sigma}^\beta(G_n)}{L_n^{3/2}} \right| \left| \frac{\bar{\sigma}_\alpha(G_n)}{L_n^{3/2}} - \sqrt{\mathbb{E}[S_\alpha(D_\alpha)]} \right| \\ & \leq \frac{\left| \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} - \mathbb{E}[S_\alpha(D_\alpha)] \right|}{\sqrt{\mathbb{E}[S_\alpha(D_\alpha)]}} + \frac{\left| \frac{\bar{\sigma}^\beta(G_n)^2}{L_n^3} - \mathbb{E}[S^\beta(D^\beta)] \right|}{\sqrt{\mathbb{E}[S^\beta(D^\beta)]}} \\ & \quad + \frac{\left| \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} - \mathbb{E}[S_\alpha(D_\alpha)] \right| \left| \frac{\bar{\sigma}^\beta(G_n)^2}{L_n^3} - \mathbb{E}[S^\beta(D^\beta)] \right|}{\sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]}}. \end{aligned}$$

Next we use that

$$\frac{\bar{\sigma}^\beta(G_n)^2}{L_n^3}, \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3}, \mathbb{E}[S_\alpha(D_\alpha)], \mathbb{E}[S^\beta(D^\beta)] \leq 2,$$

to get

$$\begin{aligned} & \left| \frac{\bar{\sigma}_\alpha(G_n) \bar{\sigma}^\beta(G_n)}{L_n^3} - \sqrt{\mathbb{E}[S_\alpha(D_\alpha)] \mathbb{E}[S^\beta(D^\beta)]} \right| \\ & \leq 2 \max \left\{ \frac{\left| \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} - \mathbb{E}[S_\alpha(D_\alpha)] \right|}{\sqrt{\mathbb{E}[S_\alpha(D_\alpha)]}}, \frac{\left| \frac{\bar{\sigma}^\beta(G_n)^2}{L_n^3} - \mathbb{E}[S^\beta(D^\beta)] \right|}{\sqrt{\mathbb{E}[S^\beta(D^\beta)]}} \right\}. \end{aligned}$$

It now follows, using Lemma 5.8 and the union bound, that, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{P}(B_n) & \leq \mathbb{P} \left(\left| \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} - \mathbb{E}[S_\alpha(D_\alpha)] \right| > \frac{K \sqrt{\mathbb{E}[S_\alpha(D_\alpha)]} n^{-\delta}}{2} \right) \\ & \quad + \mathbb{P} \left(\left| \frac{\bar{\sigma}^\beta(G_n)^2}{L_n^3} - \mathbb{E}[S^\beta(D^\beta)] \right| > \frac{K \sqrt{\mathbb{E}[S^\beta(D^\beta)]} n^{-\delta}}{2} \right) \\ & \leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)), \end{aligned}$$

which proves (5.33) and hence the result. $\square$

All that is left now, is to prove Lemma 5.8.

Proof of Lemma 5.8. Again, since the proofs for both statements are similar, we will only prove the first. From (5.25), and the calculations following this definition, it follows that

$$\left| \frac{4}{L_n^3} \sum_{i \rightarrow j} \bar{R}_\alpha(i \rightarrow j)^2 - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha)^2 \right|$$

$$\begin{aligned}
 &\leq \left| 1 - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \right| + \frac{1}{L_n^2} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha) \\
 &= \frac{3}{L_n - 1}.
 \end{aligned}$$

Here we used (5.29) for the last line. Now observe that by Lemma 2.7

$$\begin{aligned}
 \mathbb{E} [\mathcal{F}^{+, \alpha}(D^\alpha)^2] &= \mathbb{E} [F^{+, \alpha}(D_\alpha)^2 + F^{+, \alpha}(D_\alpha)F^{+, \alpha}(D_\alpha - 1) + F^{+, \alpha}(D_\alpha - 1)] \\
 &\quad + \mathbb{E} [F^{+, \alpha}(D_\alpha)F^{+, \alpha}(D_\alpha - 1)] \\
 &= 1 + \mathbb{E} [S_\alpha(D_\alpha)].
 \end{aligned}$$

Therefore

$$\begin{aligned}
 \left| \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} - \mathbb{E} [S_\alpha(D_\alpha)] \right| &\leq \left| \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha)^2 - \mathbb{E} [\mathcal{F}^{+, \alpha}(D^\alpha)^2] \right| \\
 &\quad + \frac{1}{L_n} + \left| \frac{L_n(L_n + 1)^2}{L_n^2 + L_n} - 1 \right| \\
 &\leq \left| \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha)^2 - \mathbb{E} [\mathcal{F}^{+, \alpha}(D^\alpha)^2] \right| + \frac{2}{L_n}.
 \end{aligned}$$

Since for the first term in the last line we have that

$$\begin{aligned}
 &\left| \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha)^2 - \mathbb{E} [\mathcal{F}^{+, \alpha}(D^\alpha)^2] \right| \\
 &\leq \sum_{k=0}^{\infty} |\mathcal{F}_n^{+, \alpha}(k)^2 f_n^{+, \alpha}(k) - \mathcal{F}^{+, \alpha}(k)^2 f^{+, \alpha}(k)| \\
 &\leq \sum_{k=0}^{\infty} \mathcal{F}^{+, \alpha}(k)^2 |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| + \sum_{k=0}^{\infty} |\mathcal{F}_n^{+, \alpha}(k)^2 - \mathcal{F}^{+, \alpha}(k)^2| f^{+, \alpha}(k) \\
 &\leq 4 \sum_{k=0}^{\infty} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| + \sup_k |\mathcal{F}_n^{+, \alpha}(k)^2 - \mathcal{F}^{+, \alpha}(k)^2| \\
 &\leq 4 \sum_{k=0}^{\infty} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| + 4 \sup_k |\mathcal{F}_n^{+, \alpha}(k) - \mathcal{F}^{+, \alpha}(k)| \\
 &\leq 8 \sum_{k=0}^{\infty} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)|,
 \end{aligned}$$

it follows that

$$\begin{aligned}
 &\mathbb{P} \left(\left| \frac{\bar{\sigma}_\alpha(G_n)^2}{L_n^3} - \mathbb{E} [S_\alpha(D_\alpha)] \right| > K n^{-\delta} \right) \\
 &\leq \mathbb{P} \left(\sum_{k=0}^{\infty} |f_n^{+, \alpha}(k) - f^{+, \alpha}(k)| > \frac{K n^{-\delta}}{16}, \Omega_n \right)
 \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P} \left(\frac{3}{L_n - 1} > \frac{Kn^{-\delta}}{2}, \Omega_n \right) + \mathbb{P}(\Omega_n^c) \\
& \leq O \left(n^{-\varepsilon+\delta} + n^{-1+\delta} + \mathbb{P}(\Omega_n^c) \right),
\end{aligned}$$

which proves the result. $\square$

5.5.3 Proof of Theorem 5.3 iii)

We first establish the asymptotic relation between $\tau_\alpha^\beta(G_n)$ and

$$\frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) - 1.$$

Proposition 5.9. *Let $G_n \in \mathcal{G}_{\eta, \varepsilon}$ and let $\alpha, \beta \in \{+, -\}$. Then, for $0 < \delta < 1$ and $K > 0$,*

$$\mathbb{P} \left(\left| \tau_\alpha^\beta(G_n) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) + 1 \right| > Kn^{-\delta} \right) \leq O \left(n^{-1+\delta} + \mathbb{P}(\Omega_n^c) \right),$$

as $n \rightarrow \infty$.

Proof. Let us first consider the number of concordant pairs, $\mathfrak{C}_\alpha^\beta(G_n)$, among $(\mathbf{D}_i^\alpha, \mathbf{D}_j^\beta)_{i \rightarrow j}$. Then

$$\begin{aligned}
\frac{2}{L_n^2} \mathfrak{C}_\alpha^\beta(G_n) &= \frac{1}{L_n^2} \sum_{i \rightarrow j} \sum_{s \rightarrow t} \mathbb{1}_{\{D_s^\alpha < D_i^\alpha\}} \mathbb{1}_{\{D_t^\beta < D_j^\beta\}} + \mathbb{1}_{\{D_s^\alpha > D_i^\alpha\}} \mathbb{1}_{\{D_t^\beta > D_j^\beta\}} \\
&= \frac{1}{L_n} \sum_{i \rightarrow j} H_n^{\alpha, \beta}(D_i^\alpha - 1, D_j^\beta - 1) \\
&\quad + \frac{1}{L_n^2} \sum_{i \rightarrow j} \sum_{s \rightarrow t} \left(1 - \mathbb{1}_{\{D_s^\alpha \leq D_i^\alpha\}} \right) \left(1 - \mathbb{1}_{\{D_t^\beta \leq D_j^\beta\}} \right) \\
&= \frac{1}{L_n} \sum_{i \rightarrow j} H_n^{\alpha, \beta}(D_i^\alpha - 1, D_j^\beta - 1) + \frac{1}{L_n} \sum_{i \rightarrow j} H_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) \\
&\quad + 1 - \frac{1}{L_n} \sum_{i \rightarrow j} F_n^{+, \alpha}(D_i^\alpha) - \frac{1}{L_n} \sum_{i \rightarrow j} F_n^{-, \beta}(D_j^\beta).
\end{aligned}$$

Here the factor 2 comes from the fact that we count each pair of distinct edges $i \rightarrow j$ and $s \rightarrow t$ twice. In a similar fashion we obtain that

$$\begin{aligned}
\frac{2}{L_n^2} \mathfrak{D}_\alpha^\beta(G_n) &= \frac{1}{L_n} \sum_{i \rightarrow j} H_n^{\alpha, \beta}(D_i^\alpha - 1, D_j^\beta) + \frac{1}{L_n} \sum_{i \rightarrow j} H_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta - 1) \\
&\quad - \frac{1}{L_n} \sum_{i \rightarrow j} F_n^{+, \alpha}(D_i^\alpha - 1) - \frac{1}{L_n} \sum_{i \rightarrow j} F_n^{-, \beta}(D_j^\beta - 1).
\end{aligned}$$

Therefore, using (5.29),

$$\frac{2(\mathfrak{C}_\alpha^\beta(G_n) + \mathfrak{D}_\alpha^\beta(G_n))}{L_n^2} = \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{+, \alpha}(D_i^\alpha)$$

$$\begin{aligned}
 & + 1 - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^{-, \beta}(D_j^\beta) \\
 & = \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) - 1.
 \end{aligned}$$

The result now follows since

$$\begin{aligned}
 & \mathbb{P} \left(\left| \tau_\alpha^\beta(G_n) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) + 1 \right| > K n^{-\delta} \right) \\
 & = \mathbb{P} \left(\left| \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) \right| \left| \frac{1}{1 - \frac{1}{L_n}} - 1 \right| > K n^{-\delta} \right) \\
 & \leq \mathbb{P} \left(\left| \frac{1}{L_n - 1} \right| > K n^{-\delta}, \Omega_n \right) + \mathbb{P}(\Omega_n^c) \\
 & \leq O(n^{-1+\delta} + \mathbb{P}(\Omega_n^c)),
 \end{aligned}$$

as $n \rightarrow \infty$. □

With this result we can prove the last part of Theorem 5.3:

Proof of Theorem 5.3 iii). Let $\alpha, \beta \in \{+, -\}$, $0 < \delta < \min(\varepsilon, 1/2)$ and $K > 0$ and let $\kappa = \min(1 - 2\delta, \varepsilon - \delta)$. We first bound the probability as follows:

$$\begin{aligned}
 & \mathbb{P}(|\tau_\alpha^\beta(G_n) - \tau(D_\alpha, D^\beta)| > n^{-\delta}) \\
 & \leq \mathbb{P} \left(\left| \tau_\alpha^\beta(G_n) - \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) + 1 \right| > \frac{n^{-\delta}}{2} \right) \\
 & \quad + \mathbb{P} \left(\left| \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) - \mathbb{E}[\mathcal{H}^{\alpha, \beta}(D^\alpha, D^\beta)] \right| > \frac{n^{-\delta}}{2} \right).
 \end{aligned}$$

From Proposition 5.9 it follows that the first term is $O(n^{-1+\delta} + \mathbb{P}(\Omega_n^c))$, as $n \rightarrow \infty$. Hence, it is enough to show that the second term is $O(\mathbb{P}(\mathcal{J}_n^{\alpha, \beta}))$. For this, note that, $\mathcal{H}^{\alpha, \beta}(k, \ell) \leq 4$ for all $k, \ell \geq 0$, so that

$$\begin{aligned}
 \sup_{k, \ell} |\mathcal{H}_n^{\alpha, \beta}(k, \ell) - \mathcal{H}^{\alpha, \beta}(k, \ell)| & \leq 4 \sup_{k, \ell} |H_n^{\alpha, \beta}(k, \ell) - H^{\alpha, \beta}(k, \ell)| \\
 & \leq 4 \sum_{k, \ell=0}^{\infty} |h_n^{\alpha, \beta}(k, \ell) - h^{\alpha, \beta}(k, \ell)|.
 \end{aligned}$$

Therefore

$$\begin{aligned}
 & \mathbb{P} \left(\left| \frac{1}{L_n} \sum_{i \rightarrow j} \mathcal{H}_n^{\alpha, \beta}(D_i^\alpha, D_j^\beta) - \mathbb{E}[\mathcal{H}^{\alpha, \beta}(D^\alpha, D^\beta)] \right| > \frac{n^{-\delta}}{2} \right) \\
 & = \mathbb{P} \left(\left| \sum_{k, \ell=0}^{\infty} \mathcal{H}_n^{\alpha, \beta}(k, \ell) h_n^{\alpha, \beta}(k, \ell) - \mathcal{H}^{\alpha, \beta}(k, \ell) h^{\alpha, \beta}(k, \ell) \right| > \frac{n^{-\delta}}{2} \right)
 \end{aligned}$$

$$\begin{aligned}
 &\leq \mathbb{P} \left(\left| \sum_{k,\ell=0}^{\infty} \mathcal{H}^{\alpha,\beta}(k,\ell) \left(h_n^{\alpha,\beta}(k,\ell) - h^{\alpha,\beta}(k,\ell) \right) \right| > \frac{n^{-\delta}}{4} \right) \\
 &\quad + \mathbb{P} \left(\sup_{k,\ell} |\mathcal{H}_n^{\alpha,\beta}(k,\ell) - \mathcal{H}^{\alpha,\beta}(k,\ell)| > \frac{n^{-\delta}}{4} \right) \\
 &\leq 2\mathbb{P} \left(\sum_{k,\ell=0}^{\infty} |h_n^{\alpha,\beta}(k,\ell) - h^{\alpha,\beta}(k,\ell)| > \frac{n^{-\delta}}{16} \right) \\
 &\leq O \left(\mathbb{P} \left(\mathcal{J}_n^{\alpha,\beta} \right) \right),
 \end{aligned}$$

as $n \rightarrow \infty$, which finishes the proof. $\square$

5.6 Convergence of rank-correlation measures under weaker assumptions

In this section we will briefly discuss convergence results for rank-correlation measures where both the bi-degree sequence and the graph satisfy weaker assumptions than we considered in the previous sections. There the results use the scaling $n^{-\varepsilon}$ from both Assumption 2.2 and 5.1 to obtain a scaling in both the concentration and the speed of convergence. However, we can still get convergence in probability of the rank-correlations when we only assume convergence in probability of the empirical degree and size-biased degree distribution and for the joint degree density $h_n^{\alpha,\beta}$. Although the results will be less strong than those in Theorem 5.3, they are still useful since, as we will see in a later chapter, it might not always be possible to prove that Assumption 5.1 is satisfied:

We will consider bi-degree sequences that satisfy the following assumption

Assumption 5.2. There exist probability mass functions f^+ and f^- , on the non-negative integers, that satisfy

$$\nu = \sum_{k=0}^{\infty} k f^+ = \sum_{k=0}^{\infty} k f^- < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} k^{1+\eta} f^{\pm} < \infty \quad \text{for some } \eta > 0,$$

such that, for any $k \geq 0$,

$$f_n^{\pm}(k) \xrightarrow{\mathbb{P}} f^{\pm}(k), \quad \text{and} \quad \nu_n \xrightarrow{\mathbb{P}} \nu,$$

as $n \rightarrow \infty$.

Observe that this assumption resembles Assumption 2.1, with the exception that we do not require convergence of the joint distribution h_n . In addition we do require, similar to Assumption 2.2, that the $1 + \eta$ moments exists. Next, we state the assumption on the joint degree probability mass function.

Assumption 5.3. There exists random variables D_{α} and D^{β} , not concentrated in a single point, with joint probability density $h^{\alpha,\beta}$, such that for any bounded

function ϕ and $\delta > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \sum_{k, \ell=0}^n \phi(k, \ell) \left(h_n^{\alpha, \beta}(k, \ell) - h^{\alpha, \beta}(k, \ell) \right) \right| \leq \delta \right) = 1. \quad (5.34)$$

Similar to the stronger assumption, we write $G_n \in \mathcal{G}(D_\alpha, D^\beta)$ when G_n has a bi-degree sequence that satisfies Assumption 5.2 while G_n satisfies Assumption 5.3, where the random variables D_α and D^β have, respectively, probability densities $f^{+, \alpha}$ and $f^{-, \beta}$.

Observe that if we denote by X_n^α and Y_n^β , respectively, the α -degree of the source and β -degree of the target of an edge sampled uniformly at random from all edges in G_n , then (5.34) is equivalent to

$$\mathbb{E}_n [\phi(X_n^\alpha, Y_n^\beta)] \xrightarrow{\mathbb{P}} \mathbb{E} [\phi(\mathcal{D}^\alpha, \mathcal{D}^\beta)],$$

as $n \rightarrow \infty$. The undirected version of this assumption is used in [82] to prove that Spearman's rho on undirected graphs converges.

Note that Assumption 5.3 implies that, for any $k \geq 0$,

$$f_n^{+, \alpha}(k) \xrightarrow{\mathbb{P}} f^{+, \alpha}(k) \quad \text{and} \quad f_n^{-, \beta}(k) \xrightarrow{\mathbb{P}} f^{-, \beta}(k),$$

as $n \rightarrow \infty$. In addition, since all probability densities are defined on the integers, the convergence in probability holds uniformly. For example

$$\sup_{k \geq 0} |F_n^\pm(k) - F^\pm(k)| \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$. Finally, we note that Assumption 5.3 is equivalent to the statement that, for any $k, \ell \geq 0$,

$$h_n^{\alpha, \beta}(k, \ell) \xrightarrow{\mathbb{P}} h^{\alpha, \beta}(k, \ell),$$

as $n \rightarrow \infty$.

The proofs in Section 5.5 can now be adjusted so that all results, where we replace the scaled bounds $Kn^{-\delta}$ by fixed bounds $\delta > 0$, hold under Assumption 5.2 and 5.3. We then obtain convergence in probability of ρ_α^β , $\bar{\rho}_\alpha^\beta$ and τ_α^β . Since the rank-correlation measures are uniformly bounded, it follows that the expectations also converge. These results are summarized in the following theorem.

Theorem 5.10. *Let $\alpha, \beta \in \{+, -\}$ and $G_n \in \mathcal{G}(D_\alpha, D^\beta)$. Then,*

$$\rho_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \rho(D^\alpha, D^\beta), \quad \bar{\rho}_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \bar{\rho}(D^\alpha, D^\beta) \quad \text{and} \quad \tau_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \tau(D^\alpha, D^\beta),$$

as $n \rightarrow \infty$. Moreover, also the means converge, i.e.,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} [\rho(G_n)_\alpha^\beta] &= \rho(D^\alpha, D^\beta), \\ \lim_{n \rightarrow \infty} \mathbb{E} [\bar{\rho}(G_n)_\alpha^\beta] &= \bar{\rho}(D^\alpha, D^\beta) \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} \mathbb{E} [\tau(G_n)_\alpha^\beta] = \tau(D^\alpha, D^\beta).$$

Chapter 6

Directed graphs with neutral mixing of degrees and given degree distributions

Based on [83]:

van der Hoorn, P. and Litvak, N.

Convergence of rank based degree-degree correlations in random directed networks

Moscow Journal of Combinatorics and Number Theory 4.4 (2014), pp 45-83.

6.1 Introduction

In this chapter we will show that the directed configuration model can be used for generating simple directed graphs for which the rank-correlation measures, ρ_α^β , $\bar{\rho}_\alpha^\beta$ and τ_α^β as given in, respectively, Definition 4.3, 4.4 and 4.5, converge to zero, while the empirical degree distributions converge to certain prescribed limits.

We show that in both the repeated and erased version of the directed configuration model all three rank-correlation measures converge in probability to zero as the size of the graph increases. This result is not very surprising for the general configuration model, since here nodes are connected at random. However, in the repeated model we condition on the the graph being simple, and this gives a restriction on which edges can be present in the graph. Because in the erased configuration model we remove edges, we might expect the degree-degree correlations to be affected by this.

The rest of this chapter is structured as follows. We start with analyzing rank-correlation measures in the general configuration model, for generating graphs with prescribed bi-degree sequence, in Section 6.2. We then turn to the setting where we generate simple graphs with prescribed limits of the empirical degree distributions. Results for the repeated version are discussed in Section 6.3, while the erased configuration model is analyzed in Section 6.4. We end the chapter with numerical experiments for the rank-correlations in all three versions of the configuration model in Section 6.5.

6.2 Generating graphs with prescribed bi-degree sequence

Given a bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$, the lists of outbound and inbound stubs $\mathbf{S}_{L_n}$ and $\mathbf{T}_{L_n}$, respectively, are defined as:

$$\mathbf{S}_{L_n} = \{\underbrace{1, \dots, 1}_{D_1^+}, \dots, \underbrace{i, \dots, i}_{D_i^+}, \dots, \underbrace{n, \dots, n}_{D_n^+}\} \quad (6.1)$$

$$\mathbf{T}_{L_n} = \{\underbrace{1, \dots, 1}_{D_1^-}, \dots, \underbrace{j, \dots, j}_{D_j^-}, \dots, \underbrace{n, \dots, n}_{D_n^-}\}, \quad (6.2)$$

The general version of the directed configuration model constructs a graph by randomly pairing the outbound and inbound stubs.

We will write $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$, when G_n is generated by the directed configuration model, Algorithm 3, from bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$. Note that such a graph G_n will, in general, be a multi-graph.

We will focus on the setting where the bi-degree sequence is generated by the directed IID algorithm, as described in Chapter 3. We will write $G_n = \text{CM}(\mathcal{D}^+, \mathcal{D}^-)$, when the G_n is generated by the directed configuration model and $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$.

Algorithm 3 Directed Configuration Model

Input: a bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$.
 Extract the sequences $\mathbf{S}_{L_n}$ and $\mathbf{T}_{L_n}$ as defined in, respectively, (6.1) and (6.2).
 Let σ be a random permutation of the set $\{1, 2, \dots, L_n\}$.
 Let $(G_{ij})_{1 \leq i, j \leq n}$ to be an all zero $n \times n$ matrix.
for $1 \leq t \leq L_n$ **do**
 Set $i = S_t$ and $j = T_{\sigma(t)}$.
 $G_{ij} = G_{ij} + 1$
end for
 Output: G_n .

6.2.1 Neutral mixing of degrees in configuration graphs

Since, in the general configuration model, outbound and inbound stubs are randomly paired, there is no preference in connecting nodes with a specific out- or in-degree. Hence, we should expect neutral mixing of the degrees in graphs generated by this model. Our main result for the general configuration model confirms this intuition by stating that, when the bi-degree sequence is generated by the IID algorithm, the three rank-correlation measures converge to zero. For this recall from Chapter 3 (Theorem 3.2) that when $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.2 then $\text{IID}(\mathcal{D}^+, \mathcal{D}^-) \in \mathcal{D}_{\eta, \varepsilon}$ for any $\varepsilon < \eta/(8 + 4\eta)$ and η from Assumption 3.2.

Theorem 6.1. *Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be integer-valued random variables that satisfy Assumption 3.2 for some $0 < \eta \leq 1$. Let $\alpha, \beta \in \{+, -\}$, $G_n = \mathbf{CM}(\mathcal{D}^+, \mathcal{D}^-)$ and define $m = \eta/(8 + 4\eta)$. Then, for any $0 < \delta < m$, as $n \rightarrow \infty$,*

- i) $\mathbb{P}(|\rho_\alpha^\beta(G_n)| > n^{-\delta}) \leq O(n^{-\kappa}),$
- ii) $\mathbb{P}(|\bar{\rho}_\alpha^\beta(G_n)| > n^{-\delta}) \leq O(n^{-\kappa}),$
- iii) $\mathbb{P}(|\tau_\alpha^\beta(G_n)| > n^{-\delta}) \leq O(n^{-\kappa}),$

where

$$\kappa = \min \left\{ \delta, \frac{m - \delta}{2}, \frac{\eta}{1 + \eta} - \delta, \frac{1}{2} - \frac{\delta(1 + \eta)}{\eta} - \frac{4m}{\eta} \right\}.$$

Note that by the choice of m in Theorem 6.1 it follows that

$$\frac{1}{2} - \frac{\delta(1 + \eta)}{\eta} - \frac{4m}{\eta} > \frac{1}{2} - \frac{m(2 + \eta)}{\eta} - \frac{2m}{\eta} = \frac{\eta - m(4 + \eta)}{2\eta} > \frac{1}{4},$$

so that $\kappa > 0$.

In [83] we showed when the bi-degree sequence satisfies Assumption 2.1, then for the directed configuration model,

$$\rho_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0, \quad \bar{\rho}_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \tau_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$. For the undirected model, under similar assumptions, it was proven in [82] that

$$\rho(G_n) \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$. Theorem 6.1 strengthens our results in [83] by giving a scaling for both the concentration of the rank-correlation measures, as well as for the speed of convergence, under the slightly stronger Assumption 2.2. However, since bi-degree sequences generated by the IID algorithm satisfy both Assumption 2.1 and 2.2, we obtain a direct improvement in this setting.

In addition, we emphasize that since rank-correlation measures on undirected graphs are computed on the directed version of the graph, cf. [82], Theorem 6.1 also improves upon the result for the undirected configuration model, when the degree sequence is created by sampling from a given degree distribution.

Theorem 6.1 will follow from Theorem 5.3 and the following proposition.

Proposition 6.2. *Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be integer-valued random variables that satisfy Assumption 3.2 for some $0 < \eta \leq 1$. Let $\alpha, \beta \in \{+, -\}$, $G_n = \mathbf{CM}(\mathcal{D}^+, \mathcal{D}^-)$ and define $m = \eta/(8 + 4\eta)$. Then, for any $0 < \delta < m$, as $n \rightarrow \infty$,*

$$\mathbf{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-) \in \mathcal{G}_{\eta, \delta}(D_\alpha, D^\beta), \quad (6.3)$$

where D_α and D^β are independent random variables. Moreover, if we define, for any $K > 0$, the event

$$\Gamma_n^{\alpha, \beta} = \left\{ \sum_{k, \ell=0}^{\infty} |h_n^{\alpha, \beta}(k, \ell) - f^{+, \alpha}(k) f^{-, \beta}(\ell)| \leq K n^{-\delta} \right\},$$

we have that for any $\delta < \lambda < m$

$$\mathbb{P}(\Gamma_n^{\alpha, \beta}) \geq 1 - O\left(n^{-\delta} + n^{-\lambda+\delta} + n^{-1/2+\delta+4\lambda/\eta}\right), \quad (6.4)$$

as $n \rightarrow \infty$.

We will first prove Theorem 6.1 using Proposition 6.2. The proof of Proposition 6.2 can be found in Section 6.2.2.

Proof of Theorem 6.1. Fix $0 < \delta < m$, $\alpha, \beta \in \{+, -\}$ and let $\lambda = (m + \delta)/2$. Then, by the first part of Proposition 6.2, D_α and D^β are independent which implies that $\rho(D_\alpha, D^\beta) = 0$ and $h^{\alpha, \beta}(k, \ell) = f^{+, \alpha}(k) f^{-, \beta}(\ell)$. Moreover, since $\delta < \lambda < m$, it follows from the second part of Proposition 6.2, where we replaced κ by λ , that if we define the event

$$\mathcal{J}_n^{\alpha, \beta} = \left\{ \sum_{k, \ell=0}^{\infty} |h_n^{\alpha, \beta}(k, \ell) - f^{+, \alpha}(k) f^{-, \beta}(\ell)| > \frac{n^{-\delta}}{12} \right\},$$

then

$$\mathbb{P}(\mathcal{J}_n^{\alpha, \beta}) \leq O\left(n^{-\lambda+\delta} + n^{-1/2+\delta+4\lambda/\eta} + \mathbb{P}(\Omega_n^c)\right).$$

Finally, note that

$$\frac{1}{2} - \delta - \frac{4\lambda}{\eta} < 1 - 2\delta,$$

so that

$$O(n^{-1+2\delta}) \leq O\left(n^{-1/2+\delta+4\lambda/\eta}\right),$$

as $n \rightarrow \infty$. Theorem 5.3 now implies that, for instance,

$$\begin{aligned} \mathbb{P}(|\rho_\alpha^\beta(G_n)| > n^{-\delta}) &\leq O\left(n^{-\lambda+\delta} + n^{-1/2+\delta+4\lambda/\eta} + \mathbb{P}(\Omega_n^c)\right) \\ &= O\left(n^{-\kappa} + \mathbb{P}(\Omega_n^c)\right), \end{aligned}$$

as $n \rightarrow \infty$. For the final step we note that by Theorem 3.2

$$\mathbb{P}(\Omega_n^c) \leq O\left(n^{-\delta} + n^{-\frac{\eta}{1+\eta}+\delta}\right) = O\left(n^{-\kappa}\right),$$

as $n \rightarrow \infty$. The results for $\bar{\rho}_\alpha^\beta(G_n)$ and $\tau_\alpha^\beta(G_n)$ follow, using similar arguments. $\square$

The remainder of this section will be devoted to the proof of Proposition 6.2.

6.2.2 Proof of Proposition 6.2

It is helpful to first understand the result in Proposition 6.2 in a little more detail. For this note that by definition of the model we have that, for any $1 \leq i, j \leq n$,

$$\mathbb{E}_n[G_{ij}] = \frac{D_i^+ D_j^-}{L_n},$$

so that for any $k, \ell \geq 0$, by definition of $h_n^{\alpha,\beta}$,

$$\begin{aligned} \mathbb{E}_n[h_n^{\alpha,\beta}(k, \ell)] &= \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \mathbb{E}_n[G_{ij}] \\ &= \left(\frac{1}{L_n} \sum_{i=1}^n D_i^+ \mathbb{1}_{\{D_i^\alpha=k\}} \right) \left(\frac{1}{L_n} \sum_{j=1}^n D_j^- \mathbb{1}_{\{D_j^\beta=\ell\}} \right) \\ &= f_n^{+,\alpha}(k) f_n^{-,\beta}(\ell). \end{aligned}$$

Under Assumption 2.2 the empirical size-biased degree probability mass functions $f_n^{+,\alpha}$ and $f_n^{-,\beta}$ converge to, respectively $f^{+,\alpha}$ and $f^{-,\beta}$. Therefore, we need to prove that the variance of $|h_n^{\alpha,\beta}(k, \ell) - \mathbb{E}_n[h_n^{\alpha,\beta}(k, \ell)]|$ tends to zero. This result is stated in the following lemma. The proof contains many separate computations and is therefore placed at the end of this section.

Lemma 6.3. *Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be integer-valued random variables that satisfy (3.3), $\alpha, \beta \in \{+, -\}$ and $G_n = \mathbf{CM}(\mathcal{D}^+, \mathcal{D}^-)$. Then, for any $k, \ell \geq 0$*

$$\mathbb{E} \left[\left(\frac{1}{L_n} \sum_{i,j=1}^n \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \left(G_{ij} - \frac{D_i^+ D_j^-}{L_n} \right) \right)^2 \middle| \Omega_n \right] \leq O(n^{-1}),$$

as $n \rightarrow \infty$.

Using Lemma 6.3 we can prove Proposition 6.2

Proof of Proposition 6.2. Observe that it is enough to prove the second statement, since this implies the first. For this we define the events

$$A_n^{\alpha,\beta} = \left\{ \sum_{k,\ell=0}^{\infty} |h_n^{\alpha,\beta}(k,\ell) - f_n^{+,\alpha}(k)f_n^{-,\beta}(\ell)| > Kn^{-\delta} \right\},$$

$$B_n^{\alpha,\beta} = \left\{ \sum_{k,\ell=0}^{\infty} |f_n^{+,\alpha}(k)f_n^{-,\beta}(\ell) - f^{+,\alpha}(k)f^{-,\beta}(\ell)| > Kn^{-\delta} \right\}.$$

Then, by Theorem 3.2, $(\mathbf{D}_n^+, \mathbf{D}_n^-) \in \mathcal{D}_{\eta,\lambda}$ and, moreover,

$$\begin{aligned} \mathbb{P}(B_n^{\alpha,\beta}) &\leq \mathbb{P}\left(\sum_{k=0}^{\infty} |f_n^{+,\alpha}(k) - f^{+,\alpha}(k)| > \frac{Kn^{-\delta}}{2}\right) \\ &\quad + \mathbb{P}\left(\sum_{\ell=0}^{\infty} |f_n^{-,\beta}(\ell) - f^{-,\beta}(\ell)| > \frac{Kn^{-\delta}}{2}\right) \\ &\leq O(n^{-\lambda+\delta} + \mathbb{P}(\Omega_n^c)) \\ &= O(n^{-\lambda+\delta} + n^{-\delta} + n^{-\frac{\eta}{1+\eta}+\delta}), \end{aligned}$$

as $n \rightarrow \infty$. Therefore, since

$$1 - \mathbb{P}(\Gamma_n^{\alpha,\beta}) \leq \mathbb{P}(A_n^{\alpha,\beta}) + \mathbb{P}(B_n^{\alpha,\beta}),$$

it is enough to show that, as $n \rightarrow \infty$,

$$\mathbb{P}(A_n^{\alpha,\beta} | \Omega_n) \leq O(n^{-\lambda+\delta} + n^{-1/2+\delta+4\lambda/\eta}). \quad (6.5)$$

To accomplish this, we define the random variable

$$X_{ij}^{\alpha,\beta}(k,\ell) = \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \left(\frac{G_{ij}}{L_n} - \frac{D_i^+ D_j^-}{L_n^2} \right)$$

and observe that

$$|h_n^{\alpha,\beta}(k,\ell) - f_n^{+,\alpha}(k)f_n^{-,\beta}(\ell)| = \left| \sum_{i,j=1}^n X_{ij}^{\alpha,\beta}(k,\ell) \right|.$$

Next, let $\omega_n = \lfloor n^{2\lambda/\eta} \rfloor$, and write

$$\mathbb{P}(A_n^{\alpha,\beta} | \Omega_n) \leq Kn^\delta \sum_{k=0}^{\omega_n} \sum_{\ell=0}^{\omega_n} \mathbb{E} \left[\left\| \sum_{i,j=1}^n X_{ij}^{\alpha,\beta}(k,\ell) \right\|^2 \middle| \Omega_n \right]^{1/2} \quad (6.6)$$

$$+ Kn^\delta \sum_{k > \omega_n} \sum_{\ell=0}^{\infty} \mathbb{E} \left[\left\| \sum_{i,j=1}^n X_{ij}^{\alpha,\beta}(k,\ell) \right\|^2 \middle| \Omega_n \right] \quad (6.7)$$

$$+ Kn^\delta \sum_{k=0}^{\omega_n} \sum_{\ell > \omega_n}^{\infty} \mathbb{E} \left[\left\| \sum_{i,j=1}^n X_{ij}^{\alpha,\beta}(k,\ell) \right\| \Omega_n \right]. \quad (6.8)$$

We will first deal with (6.6). Since

$$-\frac{1}{2} + \delta + \frac{4\lambda}{\eta} < \frac{\lambda(8+2\eta) - \eta}{2\eta} < 0,$$

it now follows from Lemma 6.3 that

$$\begin{aligned} & Kn^\delta \sum_{k=0}^{\omega_n} \sum_{\ell=0}^{\omega_n} \mathbb{E} \left[\left\| \sum_{i,j=1}^n X_{ij}^{\alpha,\beta}(k,\ell) \right\|^2 \Omega_n \right]^{1/2} \\ & \leq O\left(n^{\delta-1/2}\omega_n^2\right) \leq O\left(n^{-1/2+\delta+4\lambda/\eta}\right), \end{aligned}$$

as $n \rightarrow \infty$.

Next we consider (6.7). Using that

$$\begin{aligned} \mathbb{E} \left[\mathbb{E}_n \left[\frac{G_{ij}}{L_n} \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \right] \right] &= \mathbb{E} \left[\frac{\mathbb{E}_n[G_{ij}]}{L_n} \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \right] \\ &= \mathbb{E} \left[\frac{D_i^+ D_j^-}{L_n^2} \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \right], \end{aligned}$$

we have

$$\begin{aligned} & \sum_{k > a_n} Kn^\delta \sum_{\ell=0}^{\infty} \mathbb{E} \left[\left\| \sum_{i,j=1}^n X_{ij}^{\alpha,\beta}(k,\ell) \right\| \Omega_n \right] \\ & \leq 2Kn^\delta \sum_{k > a_n} \sum_{\ell=0}^{\infty} \sum_{i,j=1}^n \mathbb{E} \left[\left| \frac{D_i^+ D_j^-}{L_n^2} \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \right| \Omega_n \right] \\ & \leq \frac{2Kn^\delta}{\mathbb{P}(\Omega_n)} \sum_{i,j=1}^n \mathbb{E} \left[\frac{D_i^+ D_j^-}{(\nu n - n^{1-\varepsilon})^2} \mathbb{1}_{\{D_i^\alpha > a_n\}} \right]. \end{aligned}$$

To analyze this term we will use Lemma 3.7, with λ as the scaling constant and ω_n as the increasing sequence. To keep the notation clean we notice that $\omega_n^{-1-\eta} \leq O(\omega_n^{-\eta})$, as $n \rightarrow \infty$. Therefore we can use the bound for $\alpha = +$ for both cases $\alpha \in \{+, -\}$. With these observations Lemma 3.7 now states that, as $n \rightarrow \infty$,

$$\sum_{i,j=1}^n \mathbb{E} \left[D_i^+ D_j^- \mathbb{1}_{\{D_i^\alpha > \omega_n\}} \right] \leq O(n^{2-\kappa} + n^2 \omega_n^{-\eta}).$$

Hence,

$$\sum_{k > a_n} Kn^\delta \sum_{\ell=0}^{\infty} \mathbb{E} \left[\left\| \sum_{i,j=1}^n X_{ij}^{\alpha,\beta}(k,\ell) \right\| \Omega_n \right] \leq O(n^{-\lambda+\delta} + n^{-2\lambda+\delta}) = O(n^{-\lambda+\delta}),$$

as $n \rightarrow \infty$.

Using similar arguments, we obtain the same bound for (6.8), which finishes the proof of (6.5). $\square$

We will now prove Lemma 6.3:

Proof of Lemma 6.3. Define, for $1 \leq i, j, s, t \leq n$,

$$Y_{ijst} = \mathbb{E}_n [G_{ij} G_{st}] - \frac{D_i^+ D_j^- D_s^+ D_t^-}{L_n^2}.$$

Then we compute that

$$\begin{aligned} \mathbb{E}_n \left[\left(\frac{1}{L_n} \sum_{i,j=1}^n \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \left(G_{ij} - \frac{D_i^+ D_j^-}{L_n} \right) \right)^2 \right] \\ = \frac{1}{L_n^2} \sum_{i,j=1}^n \sum_{s,t=1}^n \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \mathbb{1}_{\{D_s^\alpha=k\}} \mathbb{1}_{\{D_t^\beta=\ell\}} Y_{ijst} \\ \leq \frac{1}{L_n^2} \sum_{i,j=1}^n \sum_{s,t=1}^n Y_{ijst}. \end{aligned} \quad (6.9)$$

We will show that, for certain constants $C_1, C_2, C_3 > 0$,

$$\frac{1}{L_n^2} \sum_{i,j=1}^n \sum_{s,t=1}^n Y_{ijst} \leq \frac{C_1}{L_n - 1} + \frac{C_2}{L_n} + \frac{C_3 n}{L_n^2}, \quad (6.10)$$

by analyzing the contribution of all different combinations of the indices i, j, s and t to (6.9). Such combinations include, among others, the case where $i = j = s$ and $i \neq t$. We remark that, on the event Ω_n , $L_n \geq \nu n - n^{1-\varepsilon}$, so that (6.10) implies the result of the lemma.

We will now prove (6.10). Instead of going through all cases we will consider three, since the proofs for the other cases are similar. Throughout the rest of the proof we will often use the identity

$$\frac{a}{b(b-1)} - \frac{a}{b^2} = \frac{a}{b^2(b-1)},$$

for $a, > 0$ and $b \neq 1$. We remark that since we take $n \rightarrow \infty$, we can assume that $L_n > 1$.

When all indices are distinct, we compute that

$$\begin{aligned} \mathbb{E}_n [G_{ij} G_{st}] - \frac{D_i^+ D_j^- D_s^+ D_t^-}{L_n^2} &= \frac{D_i^+ D_j^- D_s^+ D_t^-}{L_n(L_n - 1)} - \frac{D_i^+ D_j^- D_s^+ D_t^-}{L_n^2} \\ &= \frac{D_i^+ D_j^- D_s^+ D_t^-}{L_n^2(L_n - 1)}, \end{aligned}$$

so that the contribution of this term to the sum (6.9) is bounded from above by

$$\frac{1}{L_n^2} \sum_{i,j=1}^n \sum_{s,t=1}^n \frac{D_i^+ D_j^- D_s^+ D_t^-}{L_n^2(L_n - 1)} = \frac{1}{L_n - 1}.$$

Next we consider the case

$$i = j = s \quad \text{and} \quad t \neq i.$$

Since we consider the same source node i , $\mathbb{E}_n[G_{ii}G_{it}] = 0$ when $D_i^+ = 1$. Hence, we need to distinguish between the case $D_i^+ = 1$ and $D_i^+ > 1$.

$$\begin{aligned} \mathbb{E}_n[G_{ii}G_{it}] &= \frac{(D_i^+)^2 D_i^- D_t^-}{L_n^2} \\ &= \left(\frac{D_i^+(D_i^+ - 1)D_i^- D_t^-}{L_n(L_n - 1)} - \frac{(D_i^+)^2 D_i^- D_t^-}{L_n^2} \right) \mathbb{1}_{\{D_i^+ > 1\}} - \frac{D_i^- D_t^-}{L_n^2} \mathbb{1}_{\{D_i^+ = 1\}} \\ &= \left(\frac{(D_i^+)^2 D_i^- D_t^-}{L_n^2(L_n - 1)} - \frac{D_i^+ D_i^- D_t^-}{L_n(L_n - 1)} \right) \mathbb{1}_{\{D_i^+ > 1\}} - \frac{D_i^- D_t^-}{L_n^2} \mathbb{1}_{\{D_i^+ = 1\}} \\ &\leq \frac{(D_i^+)^2 D_i^- D_t^-}{L_n^2(L_n - 1)} \mathbb{1}_{\{D_i^+ > 1\}} - \frac{D_i^- D_t^-}{L_n^2} \mathbb{1}_{\{D_i^+ = 1\}} \\ &\leq \frac{D_i^+ D_t^-}{L_n - 1} \mathbb{1}_{\{D_i^+ > 1\}} - \frac{D_i^- D_t^-}{L_n^2} \mathbb{1}_{\{D_i^+ = 1\}}, \end{aligned}$$

where we have used that $D_i^+, D_i^- \leq L_n$. From this it follows that the contribution of this term to (6.9) equals,

$$\begin{aligned} \frac{1}{L_n^2} \sum_{i=1}^n \sum_{t \neq i} Y_{iit} &\leq \frac{2}{L_n^2} \sum_{i=1}^n \sum_{t \neq i} \frac{D_i^+ D_i^- D_t^-}{L_n(L_n - 1)} \mathbb{1}_{\{D_i^+ > 1\}} \\ &\quad + \frac{1}{L_n^2} \sum_{i=1}^n \sum_{t \neq i} \frac{D_i^- D_t^-}{L_n^2} \mathbb{1}_{\{D_i^+ = 1\}} \\ &\leq \frac{1}{L_n^2} \sum_{i=1}^n \frac{D_i^+ D_i^-}{L_n - 1} + \frac{1}{L_n^2} \\ &\leq \frac{1}{L_n - 1} + \frac{1}{L_n^2}. \end{aligned}$$

Finally, let us consider the case

$$i = s, \quad j = t.$$

Since here both the source and target node are the same we need to consider the different case for both D_i^+ and D_j^- . We first compute that

$$\begin{aligned} \mathbb{E}_n[G_{ij}^2] &= \frac{D_i^+(D_i^+ - 1)D_j^-(D_j^- - 1)}{L_n(L_n - 1)} \mathbb{1}_{\{D_i^+ > 1\}} \mathbb{1}_{\{D_j^- > 1\}} \\ &\quad + \frac{D_j^-}{L_n} \mathbb{1}_{\{D_i^+ = 1\}} \mathbb{1}_{\{D_j^- > 1\}} + \frac{D_i^+}{L_n} \mathbb{1}_{\{D_j^- = 1\}}. \end{aligned}$$

Therefore we have

$$\begin{aligned}
& \mathbb{E}_n [G_{ij}^2] - \frac{(D_i^+)^2 (D_j^-)^2}{L_n^2} \\
&= \left(\frac{D_i^+ (D_i^+ - 1) D_j^- (D_j^- - 1)}{L_n (L_n - 1)} - \frac{(D_i^+)^2 D_j^- D_t^-}{L_n^2} \right) \mathbb{1}_{\{D_i^+ > 1\}} \mathbb{1}_{\{D_j^- > 1\}} \\
&\quad + \frac{D_j^- (L_n - D_j^-)}{L_n^2} \mathbb{1}_{\{D_i^+ = 1\}} \mathbb{1}_{\{D_j^- > 1\}} + \frac{D_i^+ (L_n - D_i^+)}{L_n^2} \mathbb{1}_{\{D_j^- = 1\}} \\
&\leq \frac{(D_i^+)^2 (D_j^-)^2}{L_n^2 (L_n - 1)} + \frac{D_i^+ + D_j^-}{L_n} \\
&\leq \frac{D_i^+ D_j^-}{L_n - 1} + \frac{D_i^+ + D_j^-}{L_n},
\end{aligned}$$

where we have again used that $D_i^+, D_j^- \leq L_n$ for the last line. With this it follows that the contribution of the term $i = s$ and $j = t$ to (6.9) equals

$$\frac{1}{L_n} \sum_{i,j=1}^n Y_{ijij} \leq \frac{1}{L_n^2} \sum_{i,j=1}^n \frac{D_i^+ D_j^-}{L_n - 1} + \frac{D_i^+ + D_j^-}{L_n} = \frac{1}{L_n - 1} + \frac{2n}{L_n^2}.$$

Similar computations show that the contributions of all other cases to (6.9) are linear combinations of the terms

$$\frac{1}{L_n - 1}, \quad \frac{1}{L_n} \quad \text{and} \quad \frac{2}{L_n^2}.$$

From this we deduce that there exist positive constants C_1, C_2 and C_3 such that (6.10) holds. $\square$

6.3 Repeated configuration model

Let us come back to the problem of generating simple graphs for which the rank-correlation measures converge to zero and the empirical degree probability mass functions f_n^+ and f_n^- converge to certain prescribed limiting probability mass functions f^+ and f^- . Consider two integer random variables $\mathcal{D}^+$ and $\mathcal{D}^-$, that satisfy Assumption 3.2 and let $G_n = \text{CM}(\mathcal{D}^+, \mathcal{D}^-)$. Then, since Theorem 6.1 tells us that in this case the rank-correlation measures converge to zero, one suggestion would be to just repeat the general model until the resulting graph is simple. We refer to this as the *repeated configuration model*. However, this procedure might not terminate, since not every bi-degree sequence permits a simple graph. For instance, there are only two directed graphs with bi-degree sequence $\{(3, 0), (1, 1), (0, 3)\}$, see Figure 6.1, none of which are simple.

As we have mentioned in Section 1.3 we need finite variance for the repeating procedure to work. Therefore, we need the random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ to satisfy the following, stronger conditions, in addition to Assumption 3.2:

$$\mathbb{E}[(\mathcal{D}^+)^{2+\eta}] < \infty, \quad \text{and} \quad \mathbb{E}[(\mathcal{D}^-)^{2+\eta}] < \infty, \quad (6.11)$$

for some $\eta > 0$ which we take to be the largest one for which both Assumption 3.2 and (6.11) are satisfied.

Figure 6.1: The only two directed graphs on three nodes with $\mathbf{D}_3^+ = \{3, 1, 0\}$ and $\mathbf{D}_3^- = \{0, 1, 3\}$.

We now have two versions of the repeated configuration model which are described in detail below, see also the algorithm at the end of [28, Section 4.1]. We recall that a bi-degree sequence is called graphical if it admits at least one simple graph, see [28, Theorem 2.2] for necessary and sufficient conditions.

Algorithm 4 Repeated Directed Configuration Model I

Input: two integer-valued random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ that satisfy Assumption 3.2 and (6.11) and an $n \geq 1$.
Set $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$.
while not $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ is graphical **do**
 $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$.
end while
Set $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$.
while not G_n is simple **do**
 $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$.
end while
Output: G_n .

Algorithm 5 Repeated Directed Configuration Model II

Input: two integer-valued random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ that satisfy Assumption 3.2 and (6.11) and an $n \geq 1$.
Set $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$.
Set $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$.
while not G_n is simple **do**
 $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$.
 $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$.
end while
Output: G_n .

Similar to the general model, we will write $G_n = \text{RCM}(\mathcal{D}^+, \mathcal{D}^-)$ if G_n is generated by one of the two versions of the repeated configuration model with $\mathcal{D}^+$ and $\mathcal{D}^-$ as input.

We will investigate the convergence of rank-correlation measures on graphs generated by the repeated configuration model. Observe that by definition of the model, this means that if $\tilde{G}_n = \text{RCM}(\mathcal{D}^+, \mathcal{D}^-)$, we are interested in

$$\mathbb{P} \left(\left| \rho_\alpha^\beta(\tilde{G}_n) \right| > n^{-\delta} \right) = \mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) \right| > n^{-\delta} \mid \mathcal{S}_n \right),$$

for some $\delta > 0$, where $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$ and $\mathcal{S}_n$ denotes the event that G_n is simple. The main ingredient in this analysis is the probability of the event $\mathcal{S}_n$.

In [28] it is shown that when the bi-degree sequence satisfies Assumption 2.1 where, in addition, the limit random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy

$$\mathbb{E} [(\mathcal{D}^+)^2] < \infty \quad \text{and} \quad \mathbb{E} [(\mathcal{D}^-)^2] < \infty,$$

then the number of self-loops and multiple edges converge jointly in distribution to Poisson random variables. As a consequence, the probability that a graph, generated by the repeated configuration model, is simple converges to a non-negative number. Since (6.11) implies the above, we have the following.

Theorem 6.4 ([28], Theorem 4.3). *Suppose that $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.2 and (6.11), and define*

$$\lambda_1 = \frac{\mu}{\nu} \quad \text{and} \quad \lambda_2 = \frac{(\nu_2^+ - \nu)(\nu_2^- - \nu)}{2\nu^2}.$$

Then, if $G_n = \text{CM}(\mathcal{D}^+, \mathcal{D}^-)$ and $\mathcal{S}_n$ denotes the event that G_n is simple, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{S}_n) = e^{-\lambda_1 - \lambda_2} > 0.$$

Now let $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.2 and (6.11), $G_n = \text{CM}(\mathcal{D}^+, \mathcal{D}^-)$ and take $\alpha, \beta \in \{+, -\}$. Then, since $\mathbb{P}(\mathcal{S}_n)$ is bounded away from zero by Theorem 6.4, it follows from Theorem 6.1 that

$$\begin{aligned} \mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) \right| > n^{-\delta} \mid \mathcal{S}_n \right) &\leq \frac{1}{\mathbb{P}(\mathcal{S}_n)} \mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) \right| > n^{-\delta} \right) \\ &\leq O \left(n^{-\kappa} + \mathbb{P}(\Omega_n^c) \right), \end{aligned}$$

as $n \rightarrow \infty$, with δ and κ as defined in Theorem 6.1. Using similar argument for $\bar{\rho}_\alpha^\beta(G_n)$ and $\tau_\alpha^\beta(G_n)$ we obtain the following result:

Theorem 6.5. *Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be integer-valued random variables that satisfy Assumption (3.2) and (6.11), for some $0 < \eta \leq 1$. Let $\alpha, \beta \in \{+, -\}$, $G_n = \text{RCM}(\mathcal{D}^+, \mathcal{D}^-)$ and define $m = \eta/(8 + 4\eta)$. Then, for any $0 < \delta < m$, as $n \rightarrow \infty$,*

- i) $\mathbb{P} \left(\left| \rho_\alpha^\beta(G_n) \right| > n^{-\delta} \right) \leq O \left(n^{-\kappa} \right),$
- ii) $\mathbb{P} \left(\left| \bar{\rho}_\alpha^\beta(G_n) \right| > n^{-\delta} \right) \leq O \left(n^{-\kappa} \right),$
- iii) $\mathbb{P} \left(\left| \tau_\alpha^\beta(G_n) \right| > n^{-\delta} \right) \leq O \left(n^{-\kappa} \right),$

where

$$\kappa = \min \left\{ \delta, \frac{m - \delta}{2}, \frac{\eta}{1 + \eta} - \delta, \frac{1}{2} - \frac{\delta(1 + \eta)}{\eta} - \frac{4m}{\eta} \right\}.$$

6.4 Erased configuration model

As we have mentioned in Section 1.3, the erased configuration model first creates a graph using the general model. Then, all self-loops are removed and all multiple edges from nodes i to j are replaced by one edge $i \rightarrow j$. In this way the resulting graph will always be simple. The model is described in detail below.

Algorithm 6 Erased Directed Configuration Model

Input: two integer valued random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ that satisfy Assumption (3.2) and a $n \geq 1$.
Set $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$.
Set $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$.
Let $(\hat{G}_{ij})_{1 \leq i, j \leq n}$ to be an all zero $n \times n$ matrix.
for $1 \leq i, j \leq n$ **do**
 if $i \neq j$ **and** $G_{ij} \geq 1$ **then**
 $\hat{G}_{ij} = 1$
 end if
end for
Output: $\hat{G}_n$.

When $\hat{G}_n$ is generated by the erased configuration model, using the random variables $\mathcal{D}^+$ and $\mathcal{D}^-$, we will write $\hat{G}_n = \text{ECM}(\mathcal{D}^+, \mathcal{D}^-)$. In this section we will always use $\hat{G}_n$ for graphs generated by the erased configuration model and use G_n to refer to the graph, generated by the configuration model, before the removal step.

6.4.1 Altered empirical distributions and the average number of removed edges

We emphasize that, since removing edges changes the degrees of nodes, the bi-degree sequences of the graphs $\hat{G}_n$ and G_n will, in general, be different. Therefore we will denote the bi-degree sequence of $\hat{G}_n$ by $(\hat{\mathbf{D}}_n^+, \hat{\mathbf{D}}_n^-)$ and use the hats to distinguish between the properties of the erased configuration graph $\hat{G}_n$ and the corresponding, possible multi-graph, G_n generated by the general configuration model. For instance,

$$\hat{L}_n = \sum_{i=1}^n \hat{D}_i^+ = \sum_{i=1}^n \hat{D}_i^-,$$

while

$$\hat{f}_n^{\alpha, \beta}(k) = \frac{1}{\hat{L}_n} \sum_{i \rightarrow j} \hat{D}_i^\alpha \mathbb{1}_{\{\hat{D}_j = k\}},$$

denote the size biased degree probability mass functions. In addition we define

$$E_i^+ = D_i^+ - \hat{D}_i^+ \quad \text{and} \quad E_i^- = D_i^- - \hat{D}_i^- \quad (6.12)$$

to be the number of remove outbound and inbound stubs of node i , respectively, and let

$$E_{ij} = G_{ij} - \hat{G}_{ij} \quad (6.13)$$

denote the number of removed edges between nodes i and j . Observe that with these definitions we have, for instance, that

$$L_n - \widehat{L}_n = \sum_{i,j=1}^n E_{ij}, \quad E_i^+ = \sum_{j=1}^n E_{ij} \quad \text{and} \quad E_j^- = \sum_{i=1}^n E_{ij}.$$

In [28] it is proven that

$$\lim_{n \rightarrow \infty} \mathbb{P}(E_1^+ = 0) = 1 = \lim_{n \rightarrow \infty} \mathbb{P}(E_1^- = 0). \quad (6.14)$$

Since in this case $|E_1^+| \leq D_1^+$, $\mathbb{E}[D_1^+] \leq \mathbb{E}[\mathcal{D}^+ + 1] < \infty$ for all n and node labels are exchangeable, dominated convergence implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[E_{ij}] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbb{E}[E_t^+] = 0. \quad (6.15)$$

We will come back to the speed of convergence of this term in Chapter 8.

6.4.2 Convergence of rank-correlations in the erased configuration model

We have the following convergence result.

Theorem 6.6. *Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be integer-valued random variables that satisfy Assumption (3.2), $\alpha, \beta \in \{+, -\}$ and $\widehat{G}_n = \text{ECM}(\mathcal{D}^+, \mathcal{D}^-)$. Then,*

$$\rho_\alpha^\beta(\widehat{G}_n) \xrightarrow{\mathbb{P}} 0, \quad \bar{\rho}_\alpha^\beta(\widehat{G}_n) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \tau_\alpha^\beta(\widehat{G}_n) \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$.

Observe that these results are weaker than those in Theorem 6.5, which include a scaling for both the concentration and speed of convergence. The main problem with obtaining a stronger results is proving that $\text{ECM}(\mathcal{D}^+, \mathcal{D}^-) \in \mathcal{G}_{\eta, \varepsilon}(D_\alpha, D^\beta)$. We will therefore use the results from Section 5.6 to prove Theorem 6.6.

For the strategy of the proof, note that if the original bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfies Assumption 5.2 then, by [28, Proposition 4.4], so does the altered bi-degree sequence $(\widehat{\mathbf{D}}_n^+, \widehat{\mathbf{D}}_n^-)$. In addition, we can prove weak convergence of $\widehat{h}_n^{\alpha, \beta}(k, \ell)$, i.e. $\widehat{G}_n \in \mathcal{G}(D_\alpha, D^\beta)$, so that we can use the weak convergence theorem, Theorem 5.10 to obtain Theorem 6.6.

Proof of Theorem 6.6. Let $\widehat{G}_n = \text{ECM}(\mathcal{D}^+, \mathcal{D}^-)$ and fix $\alpha, \beta \in \{+, -\}$. From the above remarks it follows that it is enough to show that $\widehat{G}_n \in \mathcal{G}(D_\alpha, D^\beta)$, which is equivalent to proving that for any $k, \ell \geq 0$

$$\widehat{h}_n^{\alpha, \beta}(k, \ell) \xrightarrow{\mathbb{P}} f^{+, \alpha}(k) f^{-, \beta}(\ell),$$

as $n \rightarrow \infty$. Since, by Proposition 6.2, it follows that for any $k, \ell \geq 0$,

$$h_n^{\alpha, \beta}(k, \ell) \xrightarrow{\mathbb{P}} f^{+, \alpha}(k) f^{-, \beta}(\ell),$$

it is enough to show that, as $n \rightarrow \infty$,

$$\left| \widehat{h}_n^{\alpha,\beta}(k, \ell) - h_n^{\alpha,\beta}(k, \ell) \right| \xrightarrow{\mathbb{P}} 0, \quad (6.16)$$

for every $k, \ell \geq 0$.

Let us now fix $k, \ell \geq 0$. Then,

$$\begin{aligned} & \left| \widehat{h}_n^{\alpha,\beta}(k, \ell) - h_n^{\alpha,\beta}(k, \ell) \right| \\ &= \left| \frac{1}{\widehat{L}_n} \sum_{i,j=1}^n \mathbb{1}_{\{\widehat{D}_i^\alpha=k\}} \mathbb{1}_{\{\widehat{D}_j^\beta=\ell\}} \widehat{G}_{ij} - \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} G_{ij} \right| \\ &\leq \frac{1}{L_n} \sum_{i,j=1}^n \left| \widehat{G}_{ij} - G_{ij} \right| + \left| \frac{1}{\widehat{L}_n} - \frac{1}{L_n} \right| \sum_{i,j=1}^n \widehat{G}_{ij} \end{aligned} \quad (6.17)$$

$$+ \left| \frac{1}{L_n} \sum_{i,j=1}^n \left(\mathbb{1}_{\{\widehat{D}_i^\alpha=k\}} \mathbb{1}_{\{\widehat{D}_j^\beta=\ell\}} - \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \right) \widehat{G}_{ij} \right|. \quad (6.18)$$

For (6.17) we have,

$$\begin{aligned} \frac{1}{L_n} \sum_{i,j=1}^n \left| \widehat{G}_{ij} - G_{ij} \right| + \left| \frac{1}{\widehat{L}_n} - \frac{1}{L_n} \right| \sum_{i,j=1}^n \widehat{G}_{ij} &\leq \frac{1}{L_n} \sum_{i,j=1}^n E_{ij} + \left| 1 - \frac{\widehat{L}_n}{L_n} \right| \\ &\leq \frac{1}{L_n} \sum_{i,j=1}^n E_{ij} + \frac{1}{L_n} \left| L_n - \widehat{L}_n \right| \\ &\leq \frac{2}{L_n} \sum_{i,j=1}^n E_{ij}. \end{aligned} \quad (6.19)$$

In order to deal with (6.18), we first observe that

$$\left| \mathbb{1}_{\{\widehat{D}_i^\alpha=k\}} - \mathbb{1}_{\{D_i^\alpha=k\}} \right| \leq \mathbb{1}_{\{E_i^\alpha>0\}} \quad \text{and} \quad \left| \mathbb{1}_{\{\widehat{D}_j^\beta=\ell\}} - \mathbb{1}_{\{D_j^\beta=\ell\}} \right| \leq \mathbb{1}_{\{E_j^\beta>0\}}.$$

Therefore

$$\begin{aligned} & \left| \frac{1}{L_n} \sum_{i,j=1}^n \left(\mathbb{1}_{\{\widehat{D}_i^\alpha=k\}} \mathbb{1}_{\{\widehat{D}_j^\beta=\ell\}} - \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \right) \widehat{G}_{ij} \right| \\ &\leq \left| \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{1}_{\{\widehat{D}_j^\beta=\ell\}} \left(\mathbb{1}_{\{\widehat{D}_i^\alpha=k\}} - \mathbb{1}_{\{D_i^\alpha=k\}} \right) \widehat{G}_{ij} \right| \\ &\quad + \left| \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{1}_{\{\widehat{D}_i^\alpha=k\}} \left(\mathbb{1}_{\{\widehat{D}_j^\beta=\ell\}} - \mathbb{1}_{\{D_j^\beta=\ell\}} \right) \widehat{G}_{ij} \right| \\ &\leq \frac{1}{L_n} \sum_{i,j=1}^n \widehat{G}_{ij} \mathbb{1}_{\{E_i^\alpha>0\}} + \frac{1}{L_n} \sum_{i,j=1}^n \widehat{G}_{ij} \mathbb{1}_{\{E_j^\beta>0\}} \end{aligned}$$

$$\leq \frac{1}{L_n} \sum_{i=1}^n \widehat{D}_i^+ \mathbb{1}_{\{E_i^\alpha > 0\}} + \frac{1}{L_n} \sum_{j=1}^n \widehat{D}_j^- \mathbb{1}_{\{E_j^\beta > 0\}} \quad (6.20)$$

Now take $p = 1 + \eta$ and $q = (1 + \eta)/\eta$. Then, using (6.19), (6.20), Markov's inequality and finally Hölder's inequality, we obtain, for any $\delta > 0$,

$$\begin{aligned} & \mathbb{P} \left(\left| \widehat{h}_n^{\alpha, \beta}(k, \ell) - h_n^{\alpha, \beta}(k, \ell) \right| > \delta \right) \\ & \leq \mathbb{P} \left(\left| \widehat{h}_n^{\alpha, \beta}(k, \ell) - h_n^{\alpha, \beta}(k, \ell) \right| > \delta, |L_n - \nu n| < \frac{\nu n}{2} \right) + \mathbb{P} \left(\left| \frac{L_n}{n} - \nu \right| > \frac{\nu}{2} \right) \\ & \leq \mathbb{P} \left(\frac{4}{\nu n} \sum_{i,j=1}^n E_{ij} > \delta \right) + \mathbb{P} \left(\frac{2}{\nu n} \sum_{i=1}^n \widehat{D}_i^+ \mathbb{1}_{\{E_i^\alpha > 0\}} > \delta \right) \\ & \quad + \mathbb{P} \left(\frac{2}{\nu n} \sum_{j=1}^n \widehat{D}_j^- \mathbb{1}_{\{E_j^\beta > 0\}} > \delta \right) + \mathbb{P} \left(\left| \frac{L_n}{n} - \nu \right| > \frac{\nu}{2} \right) \\ & \leq \frac{4}{\delta \nu n} \sum_{i,j=1}^n \mathbb{E}[E_{ij}] + \frac{2}{\delta \nu n} \mathbb{E} \left[(D_1^+)^{1+\eta} \right]^{1/p} \mathbb{P}(E_1^\alpha > 0)^{1/q} \\ & \quad + \frac{2}{\delta \nu n} \mathbb{E} \left[(D_1^-)^{1+\eta} \right]^{1/p} \mathbb{P}(E_1^\beta > 0)^{1/q} + \mathbb{P} \left(\left| \frac{L_n}{n} - \nu \right| > \frac{\nu}{2} \right). \end{aligned}$$

The first term converges to zero as $n \rightarrow \infty$ by (6.15), while this holds for the second and third term due to (6.14). Finally, the last term converges to zero since by Theorem 3.2 $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfies Assumption 2.2 and hence Assumption 5.2. $\square$

6.5 Numerical evaluations of rank-correlation measures

Now that we have established theoretical results on the convergence of rank-correlations measures in all three versions of the directed configuration model we will compare them using numerical experiments.

6.5.1 Setup of the numerical experiments

We use mixed Poisson random variables with Pareto rates as input for the IID algorithm. To be more precise, we choose two independent Pareto random variables $\mathcal{P}^+$ and $\mathcal{P}^-$, both with scale parameter 1 and shape parameter γ . Then we let $\mathcal{D}^+$ and $\mathcal{D}^-$ be integer-valued random variables with conditional distributions

$$\mathbb{P}(\mathcal{D}^\pm = k | \mathcal{P}^\pm = p) = \frac{p^k e^{-p}}{k!}.$$

It can be verified that, see e.g. Exercise 6.11 and 6.12 in [79], that $\mathbb{E}[\mathcal{D}^\pm] = \mathbb{E}[\mathcal{P}^\pm]$ and

$$c_1 k^{-\gamma} \leq \mathbb{P}(\mathcal{D}^\pm > k) \leq c_2 k^{-\gamma},$$

Figure 6.2: Plot of the empirical inverse cdf of the absolute values of the rank-correlation measures for the Out-In and In-In degree-degree correlation type, in graphs of different sizes, generated by the configuration model.

for some constants $c_1, c_2 > 0$ and all $k \geq 1$.

For each model we have generated 10^3 graphs of size 10^4 , 10^5 and 10^6 . For both the general and erased configuration model we have picked $\gamma = 1.5$, which implies that the random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ have infinite variance. For the repeated configuration model we have used $\gamma = 3.5$. We have not used $\gamma = 2.5$ since in this case the model needed more than 100 attempts for generating *one* simple graph, which slowed down computations significantly. This confirms our remark in Section 1.3, where we have expressed the concern that computations with the repeated model might not be efficient.

After we have generated the graphs we have computed the three rank-correlation measures ρ , $\bar{\rho}$ and τ , for all four degree-degree correlations types. This gives us, for each combination of measure and correlation type, a vector of 10^3 independent measurements. We then computed the empirical inverse cumulative distribution function of the absolute values $|\rho|$, $|\bar{\rho}|$ and $|\tau|$. Figures 6.2-6.4 show the results for the Out-In and In-In degree-degree correlation types. We did not include all combinations because this would take too much space and the interesting things can already be observed with this smaller set.

6.5.2 Numerical results

From Figure 6.2 we see that ρ and $\bar{\rho}$ behave similarly, both for the Out-In and In-In degree-degree correlations. In addition we observe that τ is more tightly concentrated around zero.

Figure 6.3: Plot of the empirical inverse cdf of the absolute values of the rank-correlation measures for the Out-In and In-In degree-degree correlation type, in graphs of different sizes, generated by the erased configuration model.

Figure 6.4: Plot of the empirical inverse cdf of the absolute values of the rank-correlation measures for the Out-In and In-Out degree-degree correlation type, in graphs of different sizes, generated by the repeated configuration model.

When we consider the results for the erased configuration model, as displayed in Figure 6.3, we observe the same relative behavior between the three measures ρ , $\bar{\rho}$ and τ . However, when we compare the plots in Figure 6.3 with those in Figure 6.2, it is clear that the rank-correlation measures are much more concentrated around zero for configuration model graphs than for erased configuration model graphs of the same size. Here we see the first impact of the erasing procedure. Note that the difference in concentration is especially large for the Out-In degree-degree correlation. We will see in Chapter 9 that the behavior of this correlation type is significantly different from the other three, in the erased configuration model where the degree distributions have infinite variance.

Finally, we see that the results for ρ and $\bar{\rho}$ in the repeated configuration model, Figure 6.4 (a),(b),(d) and (e), strongly resemble those for the general configuration model, Figure 6.2 (a),(b),(d) and (e). For Kendall's tau we observe that the results in the repeated model are more inline with those for Spearman's rho. This is different for the general model where we saw that τ was more concentrated. This could be due to the fact that the graphs are simple, although we do not have a definitive answer for this.

In general, the simulation results show that in all three models the rank-correlation measures we introduced in Chapter 4 converge to zero fast in practice. For instance, when $n = 10^6$ all measures are with large probability, contained inside $[-0.05, 0.05]$, even when the degree distributions have infinite variance. We end by remarking that computations for the repeated model, with $\gamma = 3.5$, took significantly longer than for the erased model, with $\gamma = 1.5$. This is a strong argument against using the repeated model, in practice, for generating simple graphs with neutral mixing. We hence advice to use the erased configuration model for this purpose, even when the degree distributions have finite variance.

Chapter 7

Maximal disassortative undirected graphs, with given degree distribution

Based on [88]:

van der Hoorn, P., Ostroumova Prokhorenkova, L. and Samosvat E.

Generating maximally disassortative graphs with given degree distribution.

arXiv:1607.01742

1

7.1 Introduction

Degree-degree correlations fall into three typical categories; neutral, assortative and disassortative mixing. Graphs with neutral mixing can be constructed using the configuration model, see Chapter 6. The next natural question is how to generate graphs that fall into the latter two categories. In this chapter we will consider the problem of generating undirected graphs with disassortative mixing of degrees. We introduce a greedy algorithm for generating graphs with a given degree distribution that are maximally disassortative, with respect to Spearman's rho. With our algorithm we are able to characterize the joint degree distribution h_n^* of maximally disassortative graphs, in terms of the size-biased degree distribution f_n^* . As a result, when the degree sequence satisfies the standard regularity assumption, we obtain an expression for the limiting joint degree distribution h^* in terms of the limit of f_n^* .

In addition we prove that the tail of the distribution does not affect the minimum value of Spearman's rho. We use this to construct degree distributions with prescribed tail behavior, such that Spearman's rho on any graph with this degree distribution is bounded from below by a value that is arbitrary close to zero. We complement our theoretical results with numerical experiments that show the performance of our algorithm, in terms of the concentration of Spearman's rho around it's limit, and illustrate how the degree distribution influences the minimal value of this measure. Based on the results we formulate a conjecture which relates the minimal value of Spearman's rho on maximally disassortative graphs to the shape of some degree distributions.

This chapter is structured as follows. We will briefly address Spearman's rho in undirected graphs in Section 7.2. Our algorithm is described in Section 7.3. We first derive expressions for the joint degree density in Section 7.4 and then prove several properties of our algorithm in Section 7.5. We analyze the behavior of Spearman's rho on maximally disassortative graphs in Section 7.6 and prove that it is independent of the tail of the distribution in Section 7.7. The numerical experiments are presented in Section 7.8 and we end in Section 7.8.3 with some insights and our conjecture.

7.2 Spearman in undirected graphs

In this chapter we will work with undirected graphs. We refer the reader to Section 2.2.2 and Section 2.3 for the definitions related to undirected networks.

We will now summarize some results related to Spearman's rho in the undirected setting. For the definition of this measure we turn the undirected graph into a directed one by replacing each edge $i - j$ with two edges, $i \rightarrow j$ and $j \rightarrow i$, see [82]. We refer to this graph as the bi-directed version of the original graph. We will not distinguish between this directed version and the original undirected graph G_n . That is, we will write $i \rightarrow j \in G_n$ to mean that $i \rightarrow j$ is present in the bi-directed version of G_n , which is equivalent to stating that $i - j \in G_n$. We recall that we denote by L_n the sum over all degrees, so that L_n is *twice* the number of undirected edges and equal to the corresponding number of directed edges in G_n .

Now let $(\mathbf{U}_{i \rightarrow j}, \mathbf{W}_{i \rightarrow j})$ be a vector of independent uniform random variables $U_{i \rightarrow j}$ and $W_{i \rightarrow j}$ on $(0, 1)$, for each edge $i \rightarrow j \in G_n$, and define the ranking functions $R_*(i \rightarrow j)$ and $R^*(i \rightarrow j)$ by

$$\begin{aligned} R_*(i \rightarrow j) &= \sum_{s \rightarrow t \in G} \mathbb{1}_{\{D_s + U_{s \rightarrow t} \geq D_i + U_{i \rightarrow j}\}}, \\ R^*(i \rightarrow j) &= \sum_{s \rightarrow t \in G} \mathbb{1}_{\{D_t + W_{s \rightarrow t} \geq D_j + W_{i \rightarrow j}\}}. \end{aligned}$$

Then Spearman's rho is defined as, see [82],

$$\rho(G_n) = \frac{12 \sum_{i \rightarrow j \in G} R_*(i \rightarrow j) R^*(i \rightarrow j) - 3L_n(L_n + 1)^2}{L_n^3 - L_n}. \quad (7.1)$$

Observe that this definition is very similar to the definition of Spearman's rho $\rho_\alpha^\beta(G_n)$ for directed graphs, see Definition 4.3. Consequently, many results from Chapter 5 have a straightforward translation to the undirected setting. For instance, define

$$\mathcal{F}_n^*(k) = F_n^*(k) + F_n^*(k - 1) \quad (7.2)$$

and

$$\tilde{\rho}(G_n) = \frac{3}{2L_n} \sum_{i \rightarrow j} \mathcal{F}_n^*(D_i) \mathcal{F}_n^*(D_j) - 3. \quad (7.3)$$

Then we have the following version of Proposition 5.5:

Proposition 7.1. *Let $G_n \in \mathcal{G}_{\eta, \varepsilon}$ be an undirected graph, $0 < \delta < \min(\varepsilon, 1/2)$ and $K > 0$, and define the event*

$$A_n = \left\{ \left| \rho(G_n) - \frac{3}{L_n} \sum_{i \rightarrow j} \mathcal{F}_n^*(D_i) \mathcal{F}_n^*(D_j) + 3 \right| \leq K n^{-\delta} \right\}.$$

Then, as $n \rightarrow \infty$,

$$\mathbb{P}(A_n) \geq 1 - O(n^{-1+2\delta} + n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n)).$$

We now define the following assumption, see e.g. Assumption 5.1.

Assumption 7.1. There exist integer-valued random variables D_* and D^* with joint probability mass function $h^*(k, \ell)$ and $\varepsilon > 0$, such that, for any $K > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{k, \ell=0}^{\infty} |h_n^*(k, \ell) - h^*(k, \ell)| \leq K n^{-\varepsilon} \right) = 1.$$

Similar to the directed setting, we will write $G_n \in \mathcal{G}_{\eta, \varepsilon}(D_*, D^*)$ if $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$ and G_n satisfies Assumption 7.1, where the random variables D_* and D^* both have probability density f^* and joint probability mass function h^* . With this we then have the following undirected version of Theorem 5.3:

Theorem 7.2. *Let $G_n \in \mathcal{G}_{\eta, \varepsilon}(D_*, D^*)$ be an undirected graph. Then, for any $0 < \delta < \min(\varepsilon, 1/2)$ and $K > 0$, as $n \rightarrow \infty$,*

$$\mathbb{P}(|\rho(G_n) - \rho(D_*, D^*)| > n^{-\delta}) = O(n^{-1+2\delta} + n^{-\varepsilon+\delta} + \mathbb{P}(\mathcal{J}_n) + \mathbb{P}(\Omega_n^c)),$$

where

$$\mathcal{J}_n = \left\{ \left| \sum_{k, \ell=0}^{\infty} h_n^*(k, \ell) - h^*(k, \ell) \right| > \frac{n^{-\delta}}{16} \right\}.$$

7.3 The disassortative graph algorithm (DGA)

Given a degree sequence $\mathbf{D}_n$, let $\mathcal{G}(\mathbf{D}_n)$ denote the space of all graphs with this degree sequence and consider the following optimization problem

$$\min_{G \in \mathcal{G}(\mathbf{D}_n)} \frac{1}{L_n} \sum_{i \rightarrow j \in G} \mathcal{F}_n^*(D_i) \mathcal{F}_n^*(D_j). \quad (7.4)$$

To see the relation with minimizing $\rho(G_n)$, we note that if $\tilde{G}$ solves (7.4), then $\tilde{\rho}(\tilde{G})$ is minimal with respect to $\mathcal{G}(\mathbf{D}_n)$ and hence, by Proposition 7.1, this holds asymptotically for $\rho(G_n)$. We will describe an algorithm, called the **Disassortative Graph Algorithm (DGA)**, that solves (7.4).

Any degree sequence $\mathbf{D}_n$ can be represented by a list of stubs, where for each node i we have D_i stubs labeled i . A graph with the degree sequence $\mathbf{D}_n$ is then completely determined by the pairing of the stubs. In order to describe the algorithm, let N_k denote the number of nodes with degree k and let z_n be the unique integer satisfying

$$\sum_{t=1}^{z_n} tN_t \geq \frac{L_n}{2} \quad \text{and} \quad \sum_{t=1}^{z_n-1} tN_t < \frac{L_n}{2}. \quad (7.5)$$

Since $L_n/2$ is the number of undirected edges for a graph with degree sequence $\mathbf{D}_n$, it follows from (7.5) that there are less than $L_n/2$ stubs belonging to nodes of degree strictly smaller than z_n . Hence we can always connect all nodes with degree strictly smaller than z_n to nodes with degree strictly larger than z_n .

The idea of the **Disassortative Graph Algorithm** is to use z_n to divide the stubs in two columns. In the left column S_n we add the stubs belonging to nodes with high degree ($D_i \geq z_n$), in descending order. The right column T_n will be filled with stubs that belong to nodes with small degree ($D_i \leq z_n$) in ascending order. After this ordering we start pairing stubs from the left column to stubs in the right column, until we reach the first pair (i, j) for which $D_i = z_n = D_j$. We are now left with stubs belonging to nodes with degree z_n , hence the value of Spearman's rho (7.3) will not be influenced by the specific way in which we connect them. This means that we can, in principle, use any algorithm to connect these median degree nodes. We will use the configuration model, see e.g. [79] or Section 6.2 for the directed version. For the detailed description see Algorithm 7.

We denote by G_n^* the induced sub-graph that has been created at the end of step 17 and let the complement $G_n^z = G_n \setminus G_n^*$ denote the graph generated by the

Algorithm 7 Disassortative Graph Algorithm

-
- 1: Input: A degree sequence $\mathbf{D}_n$.
 - 2: Rank the nodes by their degrees in ascending order and let $\varrho(k)$ and denote the node with rank k , i.e. $D_{\varrho(n)} \geq D_{\varrho(n-1)} \geq \dots \geq D_{\varrho(2)} \geq D_{\varrho(1)}$.
 - 3: Create two empty lists S_n and T_n .
 - 4: Set $i = n$ and $j = 1$.
 - 5: **while** $D_{\varrho(i)} \geq z_n$ **do**
 - 6: Fill the next $D_{\varrho(i)}$ slots of S_n with stubs labeled $\varrho(i)$.
 - 7: Set $i = i - 1$.
 - 8: **end while**
 - 9: **while** $D_{\varrho(j)} \leq z_n$ **do**
 - 10: Add to T_n , $D_{\varrho(j)}$ copies of stubs labeled: $\varrho(j), \dots, \varrho(j + N_{D_{\varrho(j)}} - 1)$.
 - 11: Set $j = j + N_{D_{\varrho(j)}}$.
 - 12: **end while**
 - 13: Set $t = 1$, $i = S_n[t]$ and $j = T_n[t]$
 - 14: **while not** $D_i = z_n = D_j$ **do**
 - 15: Add edge $i - j$ to G_n .
 - 16: Set $t = t + 1$, $i = S_n[t]$ and $j = T_n[t]$.
 - 17: **end while**
 - 18: Set $\mathbf{D}_n^z$ to be the degree sequence corresponding to the remaining unpaired stubs.
 - 19: Pair the stubs in $\mathbf{D}_n^z$ using the configuration model.
 - 20: Output: G_n .
-

Figure 7.1: Top part of the two lists of stubs.

configuration model in step 19. In addition we will write $G_n = \text{DGA}(\mathbf{D}_n)$ if G_n is generated by the **Disassortative Graph Algorithm** with degree sequence $\mathbf{D}_n$ as input. An illustration of the lists S_n and T_n is displayed in Figure 7.1.

We will now illustrate the **DGA** on the simple degree sequence $\{1, 2, 2, 3\}$, see Figure 7.2. Observe that in this case $z_n = 2$. Figure 7.2a shows the initialization state where we have created the lists S_n and T_n and no stubs have been connected. We start, Figure 7.2b, by connecting the nodes at the top of the lists, 4 and 1. Then we move down the lists, Figure 7.2c, and connect 4 and 2. The next step, Figure 7.2d, is where the specific way the algorithm ordered the stubs in both lists comes into play.

There is one stub left on the node with the largest degree, node 4. The smallest degree among the still available nodes is two. Therefore we want to connect node 4 to a node with degree two which are 2 and 3. However, since there is already an edge between 4 and 2, connecting them again will create multi-edges between these nodes. The ordering of the lists resolves this by making sure we first connected to each different node with the same degree before we can create an edge between two nodes that have already been connected. In this example we therefore connect 4 and 3.

After this step the algorithm reaches a pair of nodes that both have degree $z_n = 2$, Figure 7.2e. This is where we stop and pair the remaining stubs using the configuration model. Since in this specific example only nodes 2 and 3 have a stub left, we connect these, Figure 7.2f.

Although the **DGA** is defined for arbitrary degree sequences, in practice we would like to have $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$ for some $\eta, \varepsilon > 0$ and limit probability mass functions f and f^* . In the following sections we will derive several properties of the **DGA** under this assumption.

7.4 Joint degree distribution of graphs generated by the DGA

Before we turn to analysis of the **DGA** it is useful to look at the empirical joint degree distribution for graphs generated by the algorithm. We will give a complete characterization of both the empirical and limiting joint degree distribution in Proposition 7.3 and Theorem 7.4, respectively.

7.4.1 Partitioned representation of the DGA

In order to analyze the structure of the joint degree distribution we approach the algorithm from a different angle. First observe that if we are only interested in the degrees D_i and D_j for an undirected edge $i - j$, then the specific way in which the stubs are ordered by the algorithm is irrelevant for $h_n^*(k, \ell)$. This means we do not have to consider the label of the nodes to which stubs belong, only their degree. Note that the number of stubs belonging to nodes of degree k equals kN_k . Moreover, due to the symmetry in the transition to directed edges, by replacing $i - j$ with $i \rightarrow j$ and $j \rightarrow i$, the directed structure of the graph generated by **DGA** can be seen as follows.

Figure 7.2: Example of the DGA on a simple degree sequence with $z_n = 2$.

Figure 7.3: Partitioned representation of the joint degree structure of graphs generated by the DGA.

Consider the partition of the set $\{1, 2, \dots, L_n\}$ given by kN_k for $k = 0, 1, \dots$, represented as a line of length L_n partitioned into intervals of size kN_k . Now take a copy of this partitioned line, reverse it and place it below the original one, see Figure 7.3. Both lines can be seen as the lists of all stubs, ordered by the degree of the nodes to which they belong. For the top line the stubs are ordered, from left to right, in increasing order of the degree, while for the bottom line the degrees are in decreasing order. Then the DGA can be seen as creating directed edges $i \rightarrow j$ between the nodes i corresponding to the stubs on the bottom line and nodes j corresponding to stubs in the top line.

From this representation we observe that an edge $i \rightarrow j$ between nodes of degree $D_i = k$ and $D_j = \ell$ exists if and only if the interval corresponding to kN_k in the partitioned bottom line has an intersection with the interval corresponding to ℓN_ℓ in the partitioned upper line. In terms of N_t this holds, if and only if,

$$\sum_{t=k+1}^{\infty} tN_t < \sum_{t=1}^{\ell} tN_t \quad \text{and} \quad \sum_{t=1}^{\ell-1} tN_t < \sum_{t=k}^{\infty} tN_t. \quad (7.6)$$

Moreover, the number of edges that connect nodes of degree k and ℓ is equal to the size of the intersection,

$$\min \left\{ \sum_{t=k}^{\infty} tN_t, \sum_{t=1}^{\ell} tN_t \right\} - \max \left\{ \sum_{t=k+1}^{\infty} tN_t, \sum_{t=1}^{\ell-1} tN_t \right\}. \quad (7.7)$$

This partitioned representation of both the DGA and the joint degree structure, as displayed in Figure 7.3, will be crucial for the analysis of the structure of maximally disassortative graphs.

First let $\psi_n(k, \ell)$ denote the indicator that there exists a directed edge $i \rightarrow j$ with $D_i = k$ and $D_j = \ell$. Then since for any $k \geq 0$,

$$\frac{1}{L_n} \sum_{t=0}^k tN_t = \frac{1}{L_n} \sum_{t=0}^k t \sum_{i=1}^n \mathbb{1}_{\{D_i=t\}} = F_n^*(k),$$

it follows from (7.6) that

$$\psi_n(k, \ell) := \mathbb{1}_{\{1 - F_n^*(k) < F_n^*(\ell)\}} \mathbb{1}_{\{1 - F_n^*(k-1) > F_n^*(\ell-1)\}}. \quad (7.8)$$

Moreover, if we let $\Theta_n(k, \ell)$ denote the average number of edges between nodes of degree k and ℓ , then (7.7) implies that

$$\Theta_n(k, \ell) = \min(1 - F_n^*(k-1), F_n^*(\ell)) - \max(1 - F_n^*(k), F_n^*(\ell-1)). \quad (7.9)$$

Summarizing this we have the following result.

Proposition 7.3. *Let $G_n = \text{DGA}(\mathbf{D}_n)$ and let $\psi_n(k, \ell)$ and $\Theta_n(k, \ell)$ be defined as is, respectively, (7.8) and (7.9). Then*

$$h_n^*(k, \ell) = \psi_n(k, \ell) \Theta_n(k, \ell).$$

7.4.2 Limiting joint degree distribution

Proposition 7.3 gives a complete characterization of the joint degree structure of graphs with degree sequence $\mathbf{D}_n$, generated by the DGA, in terms of the empirical size-biased degree distribution f_n^* . When $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$, this distribution converges in probability to a limiting distribution f^* . Let us now define the functions $\phi(k, \ell)$ and $\Theta(k, \ell)$ by

$$\psi(k, \ell) = \mathbb{1}_{\{1-F^*(k) < F^*(\ell)\}} \mathbb{1}_{\{1-F^*(k-1) > F^*(\ell-1)\}}, \quad (7.10)$$

$$\Theta(k, \ell) = \min(1 - F^*(k-1), F^*(\ell)) - \max(1 - F^*(k), F^*(\ell-1)), \quad (7.11)$$

and, in addition, define the joint probability mass function h^* by

$$h^*(k, \ell) = \psi(k, \ell)\Theta(k, \ell), \quad k, \ell = 1, 2, \dots \quad (7.12)$$

Then h_n^* converges in probability to h^* as $n \rightarrow \infty$, when $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$:

Theorem 7.4. *Let $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$ and $G_n = \text{DGA}(\mathbf{D}_n)$. In addition let $0 < \delta < \varepsilon$, $K > 0$ and define the event*

$$\mathcal{J}_n = \left\{ \sum_{k, \ell=0}^{\infty} |h_n(k, \ell) - h(k, \ell)| \leq K n^{-\delta} \right\}.$$

Then

$$\mathbb{P}(\mathcal{J}_n) \geq 1 - O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$.

The proof of this theorem is provided in the next section.

7.4.3 Proof of Theorem 7.4

We will use two technical lemmas, which deal with the difference between the functions ψ_n and ψ , and Θ_n and Θ .

Lemma 7.5. *Let $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$. Then, for any $k, \ell \geq 0$, $0 < \delta < \varepsilon$ and $K > 0$*

$$\mathbb{P}(|\psi_n(k, \ell) - \psi(k, \ell)|\Theta_n(k, \ell) > K n^{-\delta}, \Omega_n) \leq O(n^{-\varepsilon+\delta}).$$

Lemma 7.6. *Let $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}(f, f^*)$. Then, for any $k, \ell \geq 0$, $K > 0$ and $0 < \delta < \varepsilon$,*

$$\mathbb{P}(|\Theta(k, \ell) - \Theta_n(k, \ell)| > C n^{-\delta}, \Omega_n) \leq O(n^{-\varepsilon+\delta}).$$

The proof of both lemmas is postponed till the end of this section. We will first give the proof of Theorem 7.4:

Proof of Theorem 7.4. Let p be the smallest integer satisfying

$$1 - F^*(p) < f^*(1), \quad (7.13)$$

and define p_n as the smallest integer that satisfies

$$1 - F_n^*(p_n) < f_n^*(1).$$

Then we have that $\mathbb{P}(p_n = p)$ converges to one, since

$$\begin{aligned}
 & \mathbb{P}(p \neq p_n, \Omega_n) \\
 &= \mathbb{P}(1 - F_n^*(p) \geq f_n^*(1), \Omega_n) \\
 &\leq \mathbb{P}((F^*(p) - F_n^*(p)) + (f^*(1) - f_n^*(1)) > f^*(1) - 1 + F^*(p), \Omega_n) \\
 &\leq \mathbb{P}(|F^*(p) - F_n^*(p)| > f^*(1) - 1 + F^*(p), \Omega_n) \\
 &\quad + \mathbb{P}(|f^*(1) - f_n^*(1)| > f^*(1) - 1 + F^*(p), \Omega_n) \\
 &\leq \frac{\mathbb{E}[|F^*(p) - F_n^*(p)| \mathbb{1}_{\Omega_n}]}{f^*(1) - 1 + F^*(p)} + \frac{\mathbb{E}[|f^*(1) - f_n^*(1)| \mathbb{1}_{\Omega_n}]}{f^*(1) - 1 + F^*(p)} \\
 &\leq \frac{2n^{-\varepsilon}}{f^*(1) - 1 + F^*(p)} \leq O(n^{-\varepsilon}),
 \end{aligned}$$

as $n \rightarrow \infty$, where we have used that by definition of p it holds that $f^*(1) - 1 + F^*(p) > 0$. Therefore, if we define the event $P_n = \{p = p_n\}$ and let $\Lambda_n = P_n \cap \Omega_n$, then

$$\mathbb{P}(\Lambda_n) \geq 1 - O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)),$$

so that for Theorem 7.4 it is enough to show that

$$\mathbb{P}(\mathcal{J}_n^c, \Lambda_n) \leq O(n^{-\varepsilon+\delta}), \quad (7.14)$$

as $n \rightarrow \infty$.

Now, observe that p_n is the smallest degree such that nodes i with degree $D_i > p_n$ will be connected to nodes with degree 1, by the DGA, while p is the corresponding degree for the limit distribution. Therefore

$$\psi(k, \ell) = \begin{cases} 1 & \text{for all } k > p \text{ and } \ell = 1, \\ 1 & \text{for all } k = 1 \text{ and } \ell > p, \\ \psi(k, \ell) & \text{for all } k \leq p \text{ and } \ell \leq p, \\ 0 & \text{else,} \end{cases} \quad (7.15)$$

while, on the event P_n , the same relations hold for ψ_n . The idea of the proof is to split the analysis into the three regions

$$(k = 1, \ell > p), \quad (k, \ell \leq p) \quad \text{and} \quad (k > p, \ell = 1).$$

The hard work is in the second region. However, since here all degree are bounded by p , it suffices to analyze individual terms

$$|\psi_n(k, \ell)\Theta_n(k, \ell) - \psi(k, \ell)\Theta(k, \ell)|,$$

instead of the full sum

$$\sum_{k, \ell=0}^p |\psi_n(k, \ell)\Theta_n(k, \ell) - \psi(k, \ell)\Theta(k, \ell)|.$$

Recall that

$$\mathcal{J}_n = \left\{ \sum_{k, \ell=0}^{\infty} |h_n(k, \ell) - h(k, \ell)| \leq Kn^{-\delta} \right\}.$$

and let us bound the probability in (7.14) as

$$\mathbb{P}(\mathcal{J}_n^c, \Lambda_n) \leq \mathbb{P}\left(\sum_{k,\ell=1}^{\infty} |\psi_n(k, \ell) - \psi(k, \ell)| \Theta_n(k, \ell) > \frac{Kn^{-\delta}}{2}, \Lambda_n\right) \quad (7.16)$$

$$+ \mathbb{P}\left(\sum_{k,\ell=1}^{\infty} \psi(k, \ell) |(\Theta(k, \ell) - \Theta_n(k, \ell))| > \frac{Kn^{-\delta}}{2}, \Lambda_n\right). \quad (7.17)$$

We will first deal with (7.16). By (7.15) and conditioned on Λ_n , we have that $|\psi_n(k, \ell) - \psi(k, \ell)| \neq 0$ only when $k, \ell \leq p$. Hence we get, using the union bound,

$$\begin{aligned} & \mathbb{P}\left(\sum_{k,\ell=1}^{\infty} |\psi_n(k, \ell) - \psi(k, \ell)| \Theta_n(k, \ell) > \frac{Kn^{-\delta}}{2}, \Lambda_n\right) \\ &= \mathbb{P}\left(\sum_{k,\ell=1}^p |\psi_n(k, \ell) - \psi(k, \ell)| \Theta_n(k, \ell) > \frac{Kn^{-\delta}}{2}, \Lambda_n\right) \\ &\leq \sum_{k,\ell=1}^p \mathbb{P}\left(|\psi_n(k, \ell) - \psi(k, \ell)| \Theta_n(k, \ell) > \frac{Kn^{-\delta}}{2p^2}, \Lambda_n\right) \\ &\leq O(n^{-\varepsilon+\delta}), \end{aligned}$$

where the last line follows from Lemma 7.5.

Next we consider (7.17). First we use (7.15) to bound the term inside the probability as

$$\sum_{k,\ell=1}^{\infty} \psi(k, \ell) |(\Theta(k, \ell) - \Theta_n(k, \ell))| \leq \sum_{k=1}^p \sum_{\ell=1}^p \psi(k, \ell) |\Theta(k, \ell) - \Theta_n(k, \ell)| \quad (7.18)$$

$$+ \sum_{\ell=p+1}^{\infty} |\Theta(1, \ell) - \Theta_n(1, \ell)| \quad (7.19)$$

$$+ \sum_{k=p+1}^{\infty} |\Theta(k, 1) - \Theta_n(k, 1)| \quad (7.20)$$

We will start by analyzing (7.19). For this we notice that $\Theta_n(1, \ell) - \Theta(1, \ell) = f_n^*(\ell) - f^*(\ell)$, so that

$$\sum_{\ell=p+1}^{\infty} |\Theta(1, \ell) - \Theta_n(1, \ell)| \leq \sum_{\ell=0}^{\infty} |f_n^*(\ell) - f^*(\ell)|.$$

The upper bound for (7.20) is the same. Therefore, again using the union bound,

$$\begin{aligned} & \mathbb{P}\left(\sum_{k,\ell=0}^{\infty} \psi(k, \ell) |(\Theta(k, \ell) - \Theta_n(k, \ell))| > \frac{Kn^{-\delta}}{2}, \Lambda_n\right) \\ &\leq 2\mathbb{P}\left(\sum_{\ell=0}^{\infty} |f_n^*(\ell) - f^*(\ell)| > \frac{Kn^{-\delta}}{6}, \Omega_n\right) \end{aligned}$$

$$\begin{aligned}
 & + \sum_{k, \ell=0}^p \mathbb{P} \left(|\Theta(k, \ell) - \Theta_n(k, \ell)| > \frac{Kn^{-\delta}}{6p^2}, \Omega_n \right) \\
 & \leq O \left(n^{-\varepsilon+\delta} \right).
 \end{aligned}$$

Here we have used Lemma 7.6 to bound the last probability in the second line.

With this final result we have proven (7.14) and hence Theorem 7.4. $\square$

All that is left is to prove the two technical lemmas 7.5 and 7.6. Due to the use of both a minimum and maximum, in the definitions of $\Theta_n(k, \ell)$ and $\Theta(k, \ell)$ and the double cases in $\psi_n(k, \ell)$ and $\psi(k, \ell)$, the proofs consists of many case distinctions, where we have to bound each specific case. In order to improve the readability of the proofs we define, for any $k, \ell \geq 0$,

$$\begin{aligned}
 A_n &= \{1 - F_n^*(k) < F_n^*(\ell)\}, \\
 B_n &= \{1 - F_n^*(k-1) > F_n^*(\ell-1)\}, \\
 I_n &= \{1 - F_n^*(k-1) \leq F_n^*(\ell)\}, \\
 J_n &= \{1 - F_n^*(k) \geq F_n^*(\ell-1)\}.
 \end{aligned}$$

With these definitions we have that $\psi_n(k, \ell) = \mathbb{1}_{A_n} \mathbb{1}_{B_n}$. Moreover since $A_n^c \cap B_n^c = \emptyset$ we have that

$$1 - \psi_n(k, \ell) = \mathbb{1}_{A_n} \mathbb{1}_{B_n^c} + \mathbb{1}_{A_n^c} \mathbb{1}_{B_n}. \quad (7.21)$$

Where the event A_n and B_n determine the value of $\psi_n(k, \ell)$, so do the events I_n and J_n define the expression for $\Theta_n(k, \ell)$, as follows:

$$\Theta_n(k, \ell) = \begin{cases} f_n^*(k) & \text{on the event } I_n \cap J_n, \\ 1 - F^*(k-1) - F^*(\ell-1) & \text{on the event } I_n \cap J_n^c, \\ F^*(k) + F^*(\ell) - 1 & \text{on the event } I_n^c \cap J_n, \\ f_n^*(\ell) & \text{on the event } I_n^c \cap J_n^c. \end{cases} \quad (7.22)$$

Note that by their definitions,

$$0 \leq \psi_n(k, \ell), \psi(k, \ell), \Theta_n(k, \ell), \Theta(k, \ell) \leq 1,$$

for all $k, \ell \geq 0$. In addition we will often use the following result:

Lemma 7.7. *Let $k, \ell \geq 0$ be such that $1 - F^*(k) < F^*(\ell)$. Then*

$$\mathbb{P}(1 - F_n^*(k) \geq F_n^*(\ell), \Omega_n) \leq O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$.

If, on the other hand, $1 - F^(k) > F^*(\ell)$, then*

$$\mathbb{P}(1 - F_n^*(k) \leq F_n^*(\ell), \Omega_n) \leq O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$.

Proof. We will prove the first statement, since the proof for the second is similar. First we write

$$\begin{aligned} & \mathbb{P}(1 - F_n^*(k) \geq F_n^*(\ell), \Omega_n) \\ &= \mathbb{P}((F^*(k) - F_n^*(k)) + 1 - F^*(k) \geq F^*(\ell) + (F_n^*(\ell) - F^*(\ell)), \Omega_n) \\ &= \mathbb{P}((F^*(k) - F_n^*(k)) + (F_n^*(\ell) - F^*(\ell)) \geq F^*(\ell) - 1 + F^*(k), \Omega_n). \end{aligned}$$

Next we use the union bound and Markov's inequality to obtain

$$\begin{aligned} & \mathbb{P}(1 - F_n^*(k) \geq F_n^*(\ell), \Omega_n) \\ & \leq \mathbb{P}\left(|F^*(k) - F_n^*(k)| \geq \frac{F^*(\ell) - 1 + F^*(k)}{2}, \Omega_n\right) \\ & \quad + \mathbb{P}\left(|F^*(\ell) - F_n^*(\ell)| \geq \frac{F^*(\ell) - 1 + F^*(k)}{2}, \Omega_n\right) \\ & \leq \frac{4\mathbb{E}[\sup_{k \geq 0} |F_n^*(k) - F^*(k)| \mathbb{1}_{\Omega_n}]}{F^*(\ell) - 1 + F^*(k)} \\ & \leq \frac{4\mathbb{E}[\sum_{k=0}^{\infty} |f_n^*(k) - f^*(k)| \mathbb{1}_{\Omega_n}]}{F^*(\ell) - 1 + F^*(k)} = O(n^{-\varepsilon}), \end{aligned}$$

as $n \rightarrow \infty$, where we have used $1 - F^*(k) < F^*(\ell)$ for the last equality. $\square$

Proof of Lemma 7.5. Note that the specific expression of $\psi(k, \ell)$ depends on the ordering between

$$1 - F^*(k) \quad \text{and} \quad F^*(\ell),$$

and

$$1 - F^*(k - 1) \quad \text{and} \quad F^*(\ell - 1).$$

Therefore, we need to consider all different cases ($<$, $=$, $>$), where we treat equality as a separate case. This gives a total of nine cases. However, there are several combinations that do not need to be considered. For instance, $1 - F^*(k) > F^*(\ell)$ implies that $1 - F^*(k - 1) > F^*(\ell - 1)$. In the end, we are left with the following cases:

- I) $1 - F^*(k) < F^*(\ell)$ and $1 - F^*(k - 1) < F^*(\ell - 1)$
- II) $1 - F^*(k) < F^*(\ell)$ and $1 - F^*(k - 1) = F^*(\ell - 1)$
- III) $1 - F^*(k) < F^*(\ell)$ and $1 - F^*(k - 1) > F^*(\ell - 1)$
- IV) $1 - F^*(k) = F^*(\ell)$ and $1 - F^*(k - 1) > F^*(\ell - 1)$
- V) $1 - F^*(k) > F^*(\ell)$ and $1 - F^*(k - 1) > F^*(\ell - 1)$

We will start with the first case:

I) $1 - F^*(k) < F^*(\ell)$ and $1 - F^*(k - 1) < F^*(\ell - 1)$

First, note that in this case $\psi(k, \ell) = 0$. Moreover, since $F^*(\ell - 1) > 1 - F^*(k - 1)$, it follows from Lemma 7.7 that

$$\mathbb{P}(B_n) \leq \mathbb{P}(B_n, \Omega_n) + \mathbb{P}(\Omega_n^c) \leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$. Hence, since $\psi_n(k, \ell) = 0$ on the event B_n^c ,

$$\begin{aligned} & \mathbb{P}(|\psi_n(k, \ell) - \psi(k, \ell)| \Theta_n(k, \ell) > Kn^{-\delta}) \\ &= \mathbb{P}(\psi_n(k, \ell) \Theta_n(k, \ell) > Kn^{-\delta}, B_n^c) + \mathbb{P}(B_n) \\ &\leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)), \end{aligned}$$

as $n \rightarrow \infty$.

II) $1 - F^*(k) = F^*(\ell)$ and $1 - F^*(k-1) = F^*(\ell-1)$

In this case we have

$$1 - F^*(k) < 1 - F^*(k-1) = F^*(\ell-1) < F^*(\ell),$$

so that, by Lemma 7.7,

$$\mathbb{P}(I_n^c) \leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)) \quad \text{and} \quad \mathbb{P}(J_n) \leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)).$$

Therefore, using (7.22) and $1 - F^*(k) = F^*(\ell)$, it follows that

$$\begin{aligned} & \mathbb{P}(|\psi_n(k, \ell) - \psi(k, \ell)| \Theta_n(k, \ell) > Kn^{-\delta}) \\ &\leq \mathbb{P}(2\Theta_n(k, \ell) > Kn^{-\delta}) \\ &\leq \mathbb{P}(2\Theta_n(k, \ell) > Kn^{-\delta}, I_n^c, J_n) + \mathbb{P}(I_n) + \mathbb{P}(J_n^c) + \mathbb{P}(I_n, J_n^c) \\ &\leq \mathbb{P}\left(|1 - F_n^*(k-1) - F_n^*(\ell-1)| > \frac{Kn^{-\delta}}{2}\right) + O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)) \\ &\leq \mathbb{P}\left(|F^*(k-1) - F_n^*(k-1)| > \frac{Kn^{-\delta}}{4}\right) \\ &\quad + \mathbb{P}\left(|F^*(\ell-1) - F_n^*(\ell-1)| > \frac{Kn^{-\delta}}{4}\right) + O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)) \\ &\leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)), \end{aligned}$$

where for the fifth line we have used that

$$\begin{aligned} & |1 - F_n^*(k-1) - F_n^*(\ell-1)| \\ &= |1 - F_n^*(k-1) - (1 - F^*(k-1)) + F^*(\ell-1) - F_n^*(\ell-1)| \\ &\leq |F^*(k-1) - F_n^*(k-1)| + |F^*(\ell-1) - F_n^*(\ell-1)|, \end{aligned}$$

since $1 - F^*(k) = F^*(\ell)$.

Case IV) can be dealt with using arguments similar to case II), while case V) is similar to I). Therefore, there is only one case left:

III) $1 - F^*(k) < F^*(\ell)$ and $1 - F^*(k-1) > F^*(\ell-1)$

We first note that, since $1 - F^*(k) < F^*(\ell)$,

$$\mathbb{P}(A_n^c) \leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$, by Lemma 7.7, and similarly, since $1 - F^*(k-1) > F^*(\ell-1)$,

$$\mathbb{P}(B_n^c) \leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)).$$

Since for this case $\psi(k, \ell) = 1$, we have,

$$\begin{aligned}
 & \mathbb{P}(|\psi_n(k, \ell) - \psi(k, \ell)| \Theta_n(k, \ell) > Kn^{-\delta}) \\
 &= \mathbb{P}((1 - \psi_n(k, \ell))\Theta_n(k, \ell) > Kn^{-\delta}) \\
 &\leq \mathbb{P}((1 - \psi_n(k, \ell))\Theta_n(k, \ell) > Kn^{-\delta}, A_n, B_n) \\
 &\quad + \mathbb{P}(A_n^c) + \mathbb{P}(B_n^c) + \mathbb{P}(A_n^c, B_n^c) \\
 &\leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)),
 \end{aligned}$$

as $n \rightarrow \infty$, where we have used (7.21) for the third line. □

Proof of Lemma 7.6. Similar to the proof of Lemma 7.5 we have to consider different cases. Here these are with respect to the different relations between

$$1 - F^*(k - 1) \quad \text{and} \quad F^*(\ell),$$

and

$$1 - F^*(k) \quad \text{and} \quad F^*(\ell - 1),$$

which determine the expression for $\Theta(k, \ell)$. To analyze each case we will also need to distinguish between the different expression of $\Theta_n(k, \ell)$, which are determined by the events I_n and J_n .

We will consider the three cases where $1 - F^*(k) > F^*(\ell - 1)$. The other six cases can be dealt with using similar arguments. First note that in this case, by Lemma 7.7,

$$\mathbb{P}(J_n^c) \leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$.

I) $1 - F^*(k - 1) < F^*(\ell)$ and $1 - F^*(k) > F^*(\ell - 1)$

Similar to $\mathbb{P}(J_n^c)$, it follows from Lemma 7.7 that, as $n \rightarrow \infty$,

$$\mathbb{P}(I_n^c) \leq O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)).$$

Therefore, by conditioning on the different combinations of I_n and J_n , we get

$$\begin{aligned}
 & \mathbb{P}(|\Theta(k, \ell) - \Theta_n(k, \ell)| > Kn^{-\delta}) \\
 &\leq \mathbb{P}(|\Theta(k, \ell) - \Theta_n(k, \ell)| > Kn^{-\delta}, I_n, J_n) + \mathbb{P}(J_n^c) + \mathbb{P}(I_n^c) + \mathbb{P}(J_n^c, I_n^c) \\
 &\leq \mathbb{P}(|f^*(k) - f_n^*(k)| > Kn^{-\delta}, \Omega_n) + O(n^{-\varepsilon} + \mathbb{P}(\Omega_n^c)) \\
 &\leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)),
 \end{aligned}$$

as $n \rightarrow \infty$.

II) $1 - F^*(k - 1) = F^*(\ell)$ and $1 - F^*(k) > F^*(\ell - 1)$

Since $f^*(k) = F^*(k) - F^*(k - 1)$,

$$\begin{aligned}
 |F_n^*(k) + F_n^*(\ell) - 1 - f^*(k)| &= |F_n^*(k) - F^*(k) + F_n^*(\ell) - 1 + F^*(k - 1)| \\
 &\leq |F_n^*(k) - F^*(k)| + |F_n^*(\ell) - F^*(\ell)|,
 \end{aligned}$$

from which it follows that

$$\begin{aligned}
 \mathbb{P}(|\Theta(k, \ell) - \Theta_n(k, \ell)| > Kn^{-\delta}, I_n^c, J_n) \\
 &\leq \mathbb{P}(|F_n^*(k) + F_n^*(\ell) - 1 - f^*(k)| > Kn^{-\delta}) \\
 &\leq \mathbb{P}\left(|F_n^*(k) - F^*(k)| > \frac{Kn^{-\delta}}{2}\right) + \mathbb{P}\left(|F_n^*(\ell) - F^*(\ell)| > \frac{Kn^{-\delta}}{2}\right) \\
 &\leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)),
 \end{aligned}$$

as $n \rightarrow \infty$. Hence, we obtain

$$\begin{aligned}
 \mathbb{P}(|\Theta(k, \ell) - \Theta_n(k, \ell)| > Kn^{-\delta}) \\
 &\leq \mathbb{P}(|f^*(k) - f_n^*(k)| > Kn^{-\delta}) \\
 &\quad + \mathbb{P}(|\Theta(k, \ell) - \Theta_n(k, \ell)| > Kn^{-\delta}, I_n^c, J_n) + 2\mathbb{P}(J_n^c) \\
 &\leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)),
 \end{aligned}$$

as $n \rightarrow \infty$.

III) $1 - F^*(k - 1) > F^*(\ell)$ and $1 - F^*(k) > F^*(\ell - 1)$

First we notice that in this case $\Theta(k, \ell) = F^*(\ell) + F^*(k) - 1$. Next, using Lemma 7.7, we have, as $n \rightarrow \infty$,

$$\mathbb{P}(I_n) \leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)).$$

Therefore it follows that

$$\begin{aligned}
 \mathbb{P}(|\Theta(k, \ell) - \Theta_n(k, \ell)| > Kn^{-\delta}) \\
 &\leq \mathbb{P}(|F^*(\ell) + F^*(k) - 1 - \Theta_n(k, \ell)| > Kn^{-\delta}, I_n^c, J_n) \\
 &\quad + \mathbb{P}(I_n) + 2\mathbb{P}(J_n^c) \\
 &\leq \mathbb{P}\left(|F_n^*(k) - F^*(k)| > \frac{Kn^{-\delta}}{2}\right) + \mathbb{P}\left(|F_n^*(\ell) - F^*(\ell)| > \frac{Kn^{-\delta}}{2}\right) \\
 &\quad + \mathbb{P}(I_n) + 2\mathbb{P}(J_n^c) \\
 &\leq O(n^{-\varepsilon+\delta} + \mathbb{P}(\Omega_n^c)),
 \end{aligned}$$

as $n \rightarrow \infty$. □

7.5 Properties of the Disassortative Graph Algorithm

We will now address several properties of the **Disassortative Graph Algorithm**. The first is concerned with the optimization problem (7.4). Then we will discuss simplicity of graphs generated by the DGA.

7.5.1 Minimizing Spearman's rho

Theorem 7.8. *The Dissassortative Graph Algorithm solves (7.4).*

We will prove this theorem at the end of this section. First, let $G_n = \text{DGA}(\mathbf{D}_n)$ and observe that if the DGA solves (7.4), it also minimizes $\sum_{i \rightarrow j \in G} D_i D_j$. Therefore, with our algorithm we also obtain the minimum value for the s metric, $s_{\min}$ as considered in [2], for graphs G_n with degree sequence $\mathbf{D}_n$. Moreover, the fact that we could use an arbitrary algorithm in step 19 to connect the nodes of degree z_n confirms the observation in [2] that graphs with minimal s metric are not unique with respect to their structure.

Theorem 7.8 is a consequence of the following lemma.

Lemma 7.9. *Consider a sequence $0 \leq a_1 \leq \dots \leq a_m$ and let $\mathcal{P}_m$ denote the set of permutations of $\{1, \dots, m\}$. Then*

$$\min_{\sigma \in \mathcal{P}_m} \sum_k a_k a_{\sigma(k)} = \sum_k a_k a_{m-k+1}$$

and this minimum is achievable for a permutation σ if and only if

$$a_{\sigma(1)} \geq a_{\sigma(2)} \geq \dots \geq a_{\sigma(n)}.$$

Proof.

[$\Rightarrow$] If $a_{\sigma(1)} \geq \dots \geq a_{\sigma(n)}$ then $\sum_k a_k a_{\sigma(k)} = \sum_k a_k a_{m-k+1}$

[$\Leftarrow$] Assume that $\sigma = \arg \min_{\sigma \in S_m} \sum_k a_k a_{\sigma(k)}$ but there exist $a_i < a_j$ such that $a_{\sigma(i)} < a_{\sigma(j)}$. Consider $\sigma^* = \sigma \cdot (ij)$ then $\sum_k a_k a_{\sigma(k)} - \sum_k a_k a_{\sigma^*(k)} = (a_i - a_j)(a_{\sigma(i)} - a_{\sigma(j)}) > 0$ which contradicts the initial assumption. $\square$

Proof of Theorem 7.8. Consider a degree sequence $\mathbf{D}_n$, rank it in ascending order and let $\varrho(k)$ denote the node with rank k among this degree sequence, as defined in the description of the DGA. Now define the sequence $\mathbf{a}_{L_n}$ by

$$a_k = \mathcal{F}_n^*(D_{\varrho(i)}) \quad \text{for all } \sum_{t=1}^{i-1} D_{\varrho(t)} < k \leq \sum_{t=1}^i D_{\varrho(t)}, \quad (7.23)$$

where we use the convention that $\sum_{t=1}^0 D_{\varrho(t)} = 0$. With this definition, the sequence $\mathbf{a}_{L_n}$ looks as follows

$$a_1 \leq \dots \leq a_k \leq \dots = \underbrace{\mathcal{F}_n^*(D_{\varrho(1)}) \leq \dots \leq \mathcal{F}_n^*(D_{\varrho(1)})}_{D_{\varrho(1)}} \leq \underbrace{\mathcal{F}_n^*(D_{\varrho(2)}) \leq \dots \leq \mathcal{F}_n^*(D_{\varrho(2)})}_{D_{\varrho(2)}} \leq \dots$$

Next, we note that for each graph $G \in \mathcal{G}(\mathbf{D}_n)$ there exists a permutation σ_G such that

$$\sum_{i \rightarrow j \in G} \mathcal{F}_n^*(D_i) \mathcal{F}_n^*(D_j) = \sum_{k=1}^{L_n} a_k a_{\sigma_G(k)}.$$

For a bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$, any permutation σ of $\{1, \dots, L_n\}$ can be seen as a pairing of the stubs and hence as σ corresponds to a directed graph with this bi-degree sequence. However, not every such permutation corresponds to a graph which is the bi-directed version of an undirected graph, i.e. for each edge $i \rightarrow j$ there is exactly one edge $j \rightarrow i$. Therefore let $\mathcal{P}(\mathbf{D}_n)$ denote the set

of all permutations of $\{1, \dots, L_n\}$ which do corresponds to an undirected graph, in its directed representation. Then the optimization problem (7.4) is equivalent to the following problem:

$$\min_{\sigma \in \mathcal{P}(\mathbf{D}_n)} \sum_{k=1}^{L_n} a_k a_{\sigma^*(k)}. \quad (7.24)$$

Now, recall the partitioned representation of the DGA that we introduced in Section 7.4, see Figure 7.3. From this description of the algorithm it is not hard to see that if $\mathbf{a}_{L_n}$ is defined by (7.23), then there exists a permutation σ^* with the property that

$$a_{\sigma^*(1)} \geq a_{\sigma^*(2)} \geq \dots \geq a_{\sigma^*(L_n)},$$

such that the DGA pairs the stubs corresponding to a_i and $a_{\sigma^*(i)}$. Therefore, Lemma 7.9 implies that

$$\sum_{k=1}^{L_n} a_k a_{\sigma^*(k)} = \min_{\sigma \in \mathcal{P}_{L_n}} \sum_{k=1}^{L_n} a_k a_{\sigma(k)},$$

where $\mathcal{P}_{L_n}$ denotes the set of all permutations of $\{1, \dots, L_n\}$. Since $\mathcal{P}(\mathbf{D}_n) \subseteq \mathcal{P}_{L_n}$, this implies that

$$\sum_{k=1}^{L_n} a_k a_{\sigma^*(k)} = \min_{\sigma \in \mathcal{P}(\mathbf{D}_n)} \sum_{k=1}^{L_n} a_k a_{\sigma^*(k)},$$

which proves that the DGA solves (7.24) and hence it solves (7.4). $\square$

7.5.2 Simplicity of DGA graphs

As we have already mentioned, the joint degree structure, and hence the optimality of the DGA, depends only on the degrees of nodes that are connected and not on their labels. In the algorithm, however, we use an ordering for filling the lists of stubs S_n and T_n . This is to make sure that the probability that G_n^* is simple, i.e. it has no self-loops and no more than one edge between nodes i and j , converges to one as $n \rightarrow \infty$.

To understand this, consider the first time that the algorithm sees a stub belonging to a node i in the list S_n with degree $D_i > z_n$. Then node i will be connected to the nodes corresponding to the next D_i stubs in T_n . Now consider such a stub, belonging to node j . Then there will be more than one edge $i - j$ if and only if there is more than one stub belonging to node j among these D_i stubs in T_n , which can only happen when $D_i > N_{D_j}$. Since the degree of nodes in T_n is bounded by z_n , we have that N_{D_j} scales as n , while the maximal degree is $o(n)$, since f has finite mean. Therefore, the event $D_i > N_{D_j}$ for $D_i > z_n$ and $D_j \leq z_n$ has vanishing probability. We hence have the following result, the details of the proof can be found at the end of this section:

Proposition 7.10. *Let $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$, $G_n = \text{DGA}(\mathbf{D}_n)$ and denote by $\mathcal{S}_n^*$ the event that G_n^* is simple, then*

$$\mathbb{P}(\mathcal{S}_n^*) \geq 1 - O\left(n^{-\varepsilon} + n^{-1/2} + n^{-\eta/2} + \mathbb{P}(\Omega_n^c)\right),$$

as $n \rightarrow \infty$.

This proposition implies that the simplicity of the graph G_n , generated by the DGA, solely depends on the simplicity of G_n^z , constructed in Step 19. Now consider the degree sequence $\mathbf{D}_n^z$ corresponding to the remaining stubs, obtained in Step 18 of the algorithm, and observe that these degrees are uniformly bounded by z_n . Now take $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$ and let z be the median of F^* , i.e. the unique integer such that

$$F^*(z) \geq \frac{1}{2} \quad \text{and} \quad F^*(z-1) < \frac{1}{2}. \quad (7.25)$$

We can show that

$$\lim_{n \rightarrow \infty} \mathbb{P}(z_n \leq z+1) = 1,$$

see the proof of Proposition 7.10. Therefore, if we define the event $A_n = \{z_n \leq z+1\}$, then conditioned on A_n the degrees in $\mathbf{D}_n^z$ are bounded by $z+1$. Hence, if we connect these stubs using the configuration model, and let $\mathcal{S}_n^z$ denote the event that G_n^z is simple, then it follows, see e.g. [79, Theorem 7.12], that there exist a constant $\mathfrak{s} > 0$, such that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{S}_n^z) = \mathfrak{s}.$$

From this and Proposition 7.10 we obtain the following corollary:

Corollary 7.11. *Let $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}$, $G_n = \text{DGA}(\mathbf{D}_n)$ and denote by $\mathcal{S}_n$ the event that G_n is simple. Then there exists a constant $\mathfrak{s} > 0$ such that*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{S}_n) = \mathfrak{s}.$$

Note that by Lemma 3.1 it follows that if $\mathcal{D}$ is an integer-valued random variable that satisfies, for some $\eta > 0$,

$$\nu := \mathbb{E}[\mathcal{D}] < \infty \quad \text{and} \quad \mathbb{E}[\mathcal{D}^{1+\eta}] < \infty,$$

then the degree sequence $\text{IID}(\mathcal{D}) \in \mathcal{D}_{\eta, \varepsilon}$, for any $0 < \varepsilon < \eta/(8+4\eta)$, while

$$\mathbb{P}(\Omega_n) \geq 1 - O(n^{-\varepsilon}),$$

as $n \rightarrow \infty$. Therefore, if we want to generate maximally disassortative graphs with limit degree density f , we can first generate a degree sequence using the IID algorithm, by sampling from f , and then connect the nodes using the DGA. From Corollary 7.11 it now follows that, in order to generate maximally disassortative simple graphs, we could repeat this procedure until the resulting graph is simple.

Proof of Proposition 7.10. Let $\mathcal{A}(f)$ be the set of all $k \geq 0$ such that $f(k) > 0$. We will prove the statement under the assumption that $\mathcal{A}(f) = \mathbb{N} = \{0, 1, 2, \dots\}$, i.e. $f(k) > 0$ for all $k \geq 0$. The proof for other cases follows similarly, by considering only $k \in \mathcal{A}(f)$ instead of $k \geq 0$.

Let z_n and z be defined as in (7.5) and (7.25), respectively, and define the event

$$A_n = \{z_n \leq z+1\}.$$

Then, by definition of z_n , we have that $F^*(z+1) > 1/2$ and hence

$$\begin{aligned} \mathbb{P}(z_n > z+1, \Omega_n) &\leq \mathbb{P}\left(F_n^*(z+1) < \frac{1}{2}, \Omega_n\right) \\ &\leq \mathbb{P}\left(F^*(z+1) - \frac{1}{2} < |F^*(z+1) - F_n^*(z+1)|, \Omega_n\right) \\ &\leq \frac{\mathbb{E}[|F^*(z+1) - F_n^*(z+1)| \mathbb{1}_{\Omega_n}]}{F^*(z+1) - \frac{1}{2}} \\ &\leq O(n^{-\varepsilon}), \end{aligned}$$

as $n \rightarrow \infty$. Therefore, if we define $\Lambda_n = A_n \cap \Omega_n$, it is enough to show that

$$\mathbb{P}((\mathcal{S}_n^*)^c, \Lambda_n) \leq O(n^{-\varepsilon} + n^{-1/2} + n^{-\eta/2}).$$

In order to analyze this probability, note that by construction there are no self-loops in G_n^* . Moreover, a node i with $D_i < z_n$ can only have more than one edge to a node j when $D_j > N_{D_i}$. Hence, when G_n^* is not simple it means that for some $1 \leq k \leq z_n$ we must have that $0 < N_k < \max_{1 \leq j \leq n} D_j$, hence

$$(\mathcal{S}_n^*)^c \subseteq \bigcup_{k=1}^{z_n} \left\{ 0 < N_k < \max_{1 \leq j \leq n} D_j \right\}.$$

Therefore, if we denote $f_{\min} = \min_{1 \leq k \leq z} f(k) > 0$, it follows from the union bound that

$$\begin{aligned} \mathbb{P}((\mathcal{S}_n^*)^c, \Lambda_n) &\leq \sum_{k=1}^{z+1} \mathbb{P}\left(0 < N_k < \max_{1 \leq j \leq n} D_j, \Lambda_n\right) \\ &= \sum_{k=1}^{z+1} \mathbb{P}\left(0 < f_n(k) < \frac{\max_{1 \leq j \leq n} D_j}{n}, \Lambda_n\right) \\ &\leq \sum_{k=1}^{z+1} \mathbb{P}\left(f(k) - n^{-\varepsilon} < \frac{\max_{1 \leq j \leq n} D_j}{n}, \Lambda_n\right) \\ &\leq \sum_{k=1}^{z+1} \mathbb{P}\left(f_{\min} < \frac{\max_{1 \leq j \leq n} D_j}{n} + n^{-\varepsilon}, \Lambda_n\right) \\ &\leq \frac{(z+1)n^{-1} \mathbb{E}[\max_{1 \leq j \leq n} D_j \mathbb{1}_{\Lambda_n}] + (z+1)n^{-\varepsilon}}{f_{\min}} \\ &\leq \frac{(z+1) \mathbb{E}[\max_{1 \leq j \leq n} D_j \mathbb{1}_{\{D_j > \sqrt{n}\}} \mathbb{1}_{\Lambda_n}]}{nf_{\min}} \\ &\quad + \frac{(z+1)n^{-1/2}}{f_{\min}} + \frac{(z+1)n^{-\varepsilon}}{f_{\min}}. \end{aligned}$$

The last probability is $O(n^{-1/2} + n^{-\varepsilon})$, as $n \rightarrow \infty$. We will now show that the

other probability is $O(n^{-\varepsilon} + n^{-\eta/2})$. For this we note that

$$\frac{\max_{1 \leq j \leq n} D_j \mathbb{1}_{\{D_j > \sqrt{n}\}}}{n} \leq \frac{1}{n} \sum_{i=1}^n D_i \mathbb{1}_{\{D_i > \sqrt{n}\}} = \frac{L_n}{n} (1 - F_n^*(\sqrt{n})),$$

and

$$1 - F^*(\sqrt{n}) = \mathbb{E} \left[D \mathbb{1}_{\{D > \sqrt{n}\}} \right] \leq n^{-\eta/2} \mathbb{E} [D^{1+\eta}].$$

Therefore we obtain that, as $n \rightarrow \infty$,

$$\begin{aligned} \frac{\mathbb{E} [\max_{1 \leq j \leq n} D_j \mathbb{1}_{\Lambda_n}]}{nf_{\min}} &\leq \frac{\mathbb{E} [L_n |F_n^*(\sqrt{n}) - F^*(\sqrt{n})| \mathbb{1}_{\Lambda_n}]}{nf_{\min}} \\ &\quad + \frac{\mathbb{E} [2L_n (1 - F^*(\sqrt{n}) \mathbb{1}_{\Lambda_n})]}{nf_{\min}} \\ &\leq \frac{\mathbb{E} [(\nu n + n^{1-\varepsilon}) \sup_{k \geq 0} |F_n^*(k) - F^*(k)| \mathbb{1}_{\Lambda_n}]}{nf_{\min}} \\ &\quad + \frac{\mathbb{E} [2(\nu n + n^{1-\varepsilon}) (1 - F^*(\sqrt{n}) \mathbb{1}_{\Lambda_n})]}{nf_{\min}} \\ &\leq \frac{(\nu + n^{-\varepsilon})n^{-\varepsilon}}{f_{\min}} + \frac{2(\nu + n^{-\varepsilon})n^{-\eta/2} \mathbb{E} [D^{1+\eta}]}{f_{\min}} \\ &\leq O(n^{-\varepsilon} + n^{-\eta/2}), \end{aligned}$$

which completes the proof. $\square$

7.6 A lower bound for Spearman's rho.

In this section we will use the DGA to obtain a general lower bound for Spearman's rho, on any graph G_n , for which the degree sequence $\mathbf{D}_n$ satisfies Assumption 2.3:

Theorem 7.12. *Let $G_n \in \mathcal{G}_{\eta, \varepsilon}$ and let D_*, D^* be random variables with joint distribution h , defined in (7.12). In addition, define*

$$\kappa = \frac{\min\{\varepsilon, 1/2\} + \delta}{2}.$$

Then, for any $0 < \delta < \min\{\varepsilon, 1/2\}$ and $K > 0$,

$$\mathbb{P}(\rho(G_n) \geq \rho(D_*, D^*) - Kn^{-\delta}) \geq 1 - O(n^{-\varepsilon+\kappa} + n^{-1+\delta} + \mathbb{P}(\Omega_n^c)), \quad (7.26)$$

as $n \rightarrow \infty$.

Moreover, there exists graphs $\tilde{G}_n$ with the same degree sequence as G_n , such that, as $n \rightarrow \infty$,

$$\mathbb{P}(|\rho(\tilde{G}_n) - \rho(D_*, D^*)| > Kn^{-\delta}) \leq O(n^{-\varepsilon+\delta} + n^{-1+\delta} + \mathbb{P}(\Omega_n^c)). \quad (7.27)$$

This result can be understood in terms of the following optimization problem. Given density functions f and f^* , consider, for a fixed n , the following objective function

$$\min_{G_n \in \mathcal{G}_{\eta, \varepsilon}(f, f^*)} \rho(G_n), \quad (7.28)$$

where the minimum is understood to be taken over all graphs G_n with degree sequence satisfying Assumption 2.3 with probability mass functions f and f^* . Then (7.26) in Theorem 7.12 states that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\min_{G_n \in \mathcal{G}_{\eta, \varepsilon}(f, f^*)} \rho(G_n) < \rho(D_*, D^*) \right) = 0,$$

where $\rho(D_*, D^*)$ is given by, see (5.8),

$$\rho(D_*, D^*) = \sum_{k, \ell=0}^{\infty} \mathcal{F}^*(k) \mathcal{F}^*(\ell) \psi(k, \ell) \Theta(k, \ell),$$

with ψ and Θ as defined in (7.10) and (7.11), respectively. Since the theorem also gives the existence of a sequence of graphs $\tilde{G}_n$ that attain this value it follows that

$$\min_{G_n \in \mathcal{G}_{\eta, \varepsilon}(f, f^*)} \rho(G_n) \xrightarrow{\mathbb{P}} \rho(D_*, D^*),$$

as $n \rightarrow \infty$.

It will come as no surprise that the graphs $\tilde{G}_n$ from Theorem 7.12 are generated by the **Disassortative Graph Algorithm** with the degree sequences of the original graphs as input. Therefore we refer to graphs generated by the **DGA** as maximally disassortative graphs.

We remark that although the **DGA** solves the minimization problem of degree-degree correlations in undirected graphs by giving an the asymptotic minimum $\rho(D_*, D^*)$ on Spearman's rho, this minimum is, in general, hard to derive since it depends on the full size-biased limit probability mass function f^* . However, for specific cases it can be computed numerically by computing

$$\sum_{k=0}^K \sum_{\ell=0}^L \mathcal{F}^*(k) \mathcal{F}^*(\ell) \psi(k, \ell) \Theta(k, \ell),$$

for certain upper bounds K and L .

Proof of Theorem 7.12. Consider a graph $G_n \in \mathcal{G}_{\eta, \varepsilon}$, denote its degree sequence by $\mathbf{D}_n$ and let $\tilde{G}_n = \text{DGA}(\mathbf{D}_n)$. Then, since $\delta < \kappa < \varepsilon$, Theorem 7.4 implies that $\tilde{G}_n \in \mathcal{G}_{\eta, \kappa}(D_*, D^*)$ and hence it follows from Theorem 7.2 that

$$\mathbb{P} \left(\left| \rho(\tilde{G}_n) - \rho(D_*, D^*) \right| > K n^{-\delta} \right) \leq O \left(n^{-1+2\delta} + n^{-\varepsilon+\kappa} + \mathbb{P}(\Omega_n^c) \right),$$

which proves (7.27).

For the first statement, note that $\kappa < \varepsilon$ implies that $G_n \in \mathcal{G}_{\eta, \kappa}$ so that for any $K > 0$, by Proposition 7.1,

$$\mathbb{P} \left(\left| \rho(G_n) - \rho(D_*, D^*) \right| > \frac{K n^{-\delta}}{2} \right) \leq O \left(n^{-1+2\kappa} + n^{-\varepsilon+\kappa} + \mathbb{P}(\Omega_n^c) \right), \quad (7.29)$$

as $n \rightarrow \infty$. In addition, it follows from Theorem 7.8 that

$$\sum_{i \rightarrow j \in G_n} \mathcal{F}_n^*(D_i) \mathcal{F}_n^*(D_j) \geq \sum_{i \rightarrow j \in \tilde{G}_n} \mathcal{F}_n^*(D_i) \mathcal{F}_n^*(D_j)$$

so that

$$\tilde{\rho}(G_n) \geq \tilde{\rho}(\tilde{G}_n),$$

with $\tilde{\rho}$ as in (7.3). Therefore we have, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{P} \left(\tilde{\rho}(G_n) < \rho(D_*, D^*) - \frac{Kn^{-\delta}}{2} \right) &\leq \mathbb{P} \left(\tilde{\rho}(\tilde{G}_n) < \rho(D_*, D^*) - \frac{Kn^{-\delta}}{2} \right) \\ &\leq \mathbb{P} \left(\left| \tilde{\rho}(\tilde{G}_n) - \rho(D_*, D^*) \right| > \frac{Kn^{-\delta}}{2} \right) \\ &\leq O \left(n^{-1+2\delta} + n^{-\varepsilon-\kappa} + \mathbb{P}(\Omega_n^c) \right), \end{aligned}$$

where we have combined Proposition 7.1 and Theorem 7.2 for the last line.

Combining the above with (7.29) yields, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{P} \left(\rho(G_n) < \rho(D_*, D^*) - Kn^{-\delta} \right) \\ &\leq \mathbb{P} \left(\tilde{\rho}(G_n) < \rho(D_*, D^*) - \frac{Kn^{-\delta}}{2} \right) + \mathbb{P} \left(|\tilde{\rho}(G_n) - \rho(G_n)| > \frac{Kn^{-\delta}}{2} \right) \\ &\leq O \left(n^{-1+2\delta} + n^{-\varepsilon-\kappa} + \mathbb{P}(\Omega_n^c) \right), \end{aligned}$$

which implies (7.26). $\square$

7.7 Spearman's rho and the tail of the degree distribution

We will now investigate the influence of the degree distribution on the value of Spearman's rho, on maximally disassortative graphs, i.e. graphs generated by the DGA. We will show that the tail of the distribution does not influence this value. This is achieved by transforming a given degree distribution, such that the asymptotic behavior of the tail of this distribution is preserved, while we increase the probability mass of the corresponding size-biased degree distribution at one.

Let us start by considering a degree distributions f , for which the size-biased distributions f^* satisfies $f^*(1) \geq 1/2$, i.e. f^* has median 1. Observe that if h is defined as in (7.12) then we have in this setting that

$$h(k, \ell) = \begin{cases} f^*(k) & \text{if } k > 1 \text{ and } \ell = 1, \\ 2f^*(1) - 1 & \text{if } k = 1 \text{ and } \ell = 1, \\ f^*(\ell) & \text{if } k = 1 \text{ and } \ell > 1, \\ 0 & \text{else.} \end{cases}$$

Hence, if D_*, D^* have joint probability mass function h then

$$\mathbb{E} [\mathcal{F}^*(D_*) \mathcal{F}^*(D^*)] = \sum_{k, \ell=1}^{\infty} \mathcal{F}^*(k) \mathcal{F}^*(\ell) h(k, \ell)$$

$$\begin{aligned}
&= f^*(1)^2(2f^*(1) - 1) + 2f^*(1) \sum_{\ell=2}^{\infty} \mathcal{F}^*(\ell) f^*(\ell) \\
&\geq f^*(1)^2(2f^*(1) - 1) + 4f^*(1)^2 \sum_{\ell=2}^{\infty} f^*(\ell) \\
&= f^*(1)^2(2f^*(1) - 1) + 4f^*(1)^2(1 - f^*(1)) \\
&= 3f^*(1)^2 - 2f^*(1)^3,
\end{aligned}$$

where we have used, see (7.2), that $\mathcal{F}^*(\ell) \geq 2f^*(1)$ for all $\ell > 1$. From this it follows that whenever $f^*(1) \geq 1/2$ and D_*, D^* have joint probability mass function h ,

$$3\mathbb{E}[\mathcal{F}^*(D_*)\mathcal{F}^*(D^*)] - 3 \geq 9f^*(1)^2 - 6f^*(1)^3 - 3. \quad (7.30)$$

Together with Theorem 7.12 this implies the following result:

Proposition 7.13. *Let f and f^* be such that $f^*(1) \geq \frac{1}{2}$ and $G_n \in \mathcal{G}_{n,\varepsilon}(f, f^*)$. Then, for any $0 < \delta < \min(\varepsilon, 1/2)$ and $K > 0$,*

$$\mathbb{P}(\rho(G_n) \geq 9f^*(1)^2 - 6f^*(1)^3 - 3 - Kn^{-\delta}) \geq 1 - O(n^{-1+2\delta} + n^{-\varepsilon+\kappa} + \mathbb{P}(\Omega_n^c)),$$

as $n \rightarrow \infty$, where $\kappa = (\min\{\varepsilon, 1/2\} + \delta)/2$.

Since the function on the right side of (7.30) is strictly monotonically increasing and is 0 when $f^*(1) = 1$, it follows that the limit of Spearman's rho on maximally disassortative graphs can be bounded from below by a value that is arbitrary close to 0, if $f^*(1)$ is large enough.

Given $0 < \xi < 1$, we will now describe a construction, see Algorithm 8, that transforms any given distribution f with support on the positive integers, into a distribution f_ξ , with support on the positive integers, such that $f_\xi(1) = \xi$ and

$$\lim_{k \rightarrow \infty} \frac{1 - F_\xi(k)}{1 - F(k)} = 1, \quad (7.31)$$

where F and F_ξ are the cumulative distribution functions of f and f_ξ , respectively.

To see that the function f_ξ , created by Algorithm 8, indeed defines a probability mass function we compute

$$\begin{aligned}
\sum_{t=1}^{\infty} f_\xi(t) &= \xi + \sum_{t=2}^{K_\xi+1} f_\xi(t) + \sum_{t=K_\xi+2}^{\infty} f(t) \\
&= \xi + x + f(K_\xi + 1) + 1 - F(K_\xi + 1) = 1.
\end{aligned}$$

Moreover, since $f_\xi(k) = f(k)$ for all $k > K_\xi$ it follows that F_ξ satisfies (7.31). We will refer to f_ξ as the ξ -transform of f .

With this transformation we can now transform a given distribution f , to get a distribution f_ρ whose size-biased distribution f_ρ^* satisfies

$$9f_\rho^*(1)^2 - 6f_\rho^*(1)^3 - 3 > \rho,$$

Algorithm 8 ξ -transform of a probability mass function f

```

1: Given: a probability mass function  $f$  and  $0 < \xi < 1$ 
2: Let  $K_\xi$  be the smallest integer such that  $F(K_\xi) > \xi$ .
3: Set  $x = F(K_\xi) - \xi > 0$  and  $f_\xi(1) = \xi$ .
4: if  $K_\xi = 1$  then
5:   Set  $f_\xi(2) = f(2) + x$ 
6: else
7:   for  $2 \leq k \leq K_\xi$  do
8:     Set  $f_\xi(k) = x/(K_\xi - 1)$ 
9:   end for
10:  Set  $f_\xi(K_\xi + 1) = f(K_\xi + 1)$ 
11: end if
12: for  $k > K_\xi$  do
13:   Set  $f_\xi(k) = f(k)$ 
14: end for
15: Output: Probability mass function  $f_\xi$ 

```

without affecting the asymptotic behavior of the tail of the original distribution f . It then follows from Proposition 7.13 that for any sequence of graphs $G_n \in \mathcal{G}_{\eta, \varepsilon}(f_\rho, f_\rho^*)$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\rho(G_n) > \rho) = 1,$$

This gives the following result:

Theorem 7.14. *Let f be any probability density function with support on the non-negative integers with mean ν and*

$$\sum_{k=0}^{\infty} k^{1+\eta} f(k) < \infty,$$

for some $\eta > 0$. Then, for any $-1 < \rho < 0$, there exists a probability mass function f_ρ on the non-negative integers with a mean ν_ρ which satisfies

$$\lim_{k \rightarrow \infty} \frac{1 - F_\rho(k)}{1 - F(k)} = \frac{\nu_\rho}{\nu}.$$

Moreover, for any sequence of graphs $G_n \in \mathcal{G}_{\eta, \varepsilon}(f_\rho, f_\rho^)$, where $f_\rho^*(k) = k f_\rho(k) / \nu_\rho$,*

$$\mathbb{P}(\rho(G_n) > \rho) \geq 1 - O\left(n^{-1+\kappa} + n^{-\varepsilon+3\kappa/4} + \mathbb{P}(\Omega_n^c)\right),$$

as $n \rightarrow \infty$, where $\kappa = \min(\varepsilon, 1/2)$.

Proof. Let ξ be such that $9(\xi/2)^2 - 6(\xi/2)^3 - 3 = \rho + \epsilon$, for some $\epsilon > 0$, and denote by f^* the size-biased distribution of f . Now take f_ξ^* to be the ξ -transform of f^* and set

$$\nu_\rho = \left(\nu(1 - F(K_\xi)) + \sum_{t=1}^{K_\xi} \frac{f_\xi^*(t)}{t} \right)^{-1},$$

where K_ξ was defined as the smallest integer such that $F^*(K_\xi) > \xi$. Now we define the function f_ρ by

$$f_\rho(0) = \frac{\nu_\rho f_\xi^*(1)}{2} = f_\rho(1) = \frac{\nu_\rho f_\xi^*(1)}{2} \quad \text{and} \quad f_\rho(t) = \frac{\nu_\rho f_\xi^*(t)}{t} \quad \text{for } t \geq 2.$$

Then, since by construction $f_\rho^*(t) = f^*(t)$ for all $t > K_\xi$, it follows that

$$\begin{aligned} \sum_{t=0}^{\infty} f_\rho(t) &= \sum_{t=1}^{\infty} \frac{\nu_\rho f_\rho^*(t)}{t} \\ &= \nu_\rho \left(\sum_{t=1}^{K_\xi} \frac{f_\rho^*(t)}{t} + \sum_{t=K_\xi+1}^{\infty} \frac{f^*(t)}{t} \right) \\ &= \nu_\rho \left(\sum_{t=1}^{K_\xi} \frac{f_\rho^*(t)}{t} + \nu(1 - F(K_\xi)) \right) = 1, \end{aligned}$$

so that f_ρ defines a probability mass function. Moreover, since for all $k > K_\xi$

$$1 - F_\rho(k) = \sum_{t=k+1}^{\infty} f_\rho(t) = \nu_\rho \sum_{t=k+1}^{\infty} \frac{f_\xi^*(t)}{t} = \nu_\rho \sum_{t=k+1}^{\infty} \frac{f^*(t)}{t} = \frac{\nu_\rho}{\nu} \sum_{t=k+1}^{\infty} f(t),$$

it follows that $\sum_{k=0}^{\infty} t^{1+\eta} f_\rho(t) < \infty$ and

$$\lim_{k \rightarrow \infty} \frac{1 - F_\rho(k)}{1 - F(k)} = \frac{\nu_\rho}{\nu}.$$

Now let $\mathcal{D}$ have probability density f_ρ , and hence size-biased density $f_\rho^*(t) = t f_\rho(t) / \nu_\rho$, and let $\mathbf{D}_n = \text{IID}(\mathcal{D})$. Then, by Theorem 3.1, $\mathbf{D}_n \in \mathcal{D}_{\eta, \varepsilon}(f_\rho, f_\rho^*)$ and since by construction of f_ρ we have that $f_\rho^*(1) = \xi/2$, it follows that

$$9f_\rho^*(1)^2 - 6f_\rho^*(1)^3 - 3 = \rho + \epsilon.$$

Hence, if G_n is a graph with degree sequence $\mathbf{D}_n$, we have, by taking $\delta = \min(\varepsilon, 1/2)/2$ in Proposition 7.13, that as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{P}(\rho(G_n) > \rho) &\geq \lim_{n \rightarrow \infty} \mathbb{P}(\rho(G_n) > \rho + \epsilon - n^{-\delta}) \\ &= \mathbb{P}(\rho(G_n) > 9f_\rho^*(1)^2 - 6f_\rho^*(1)^3 - 3 - n^{-\delta}) \\ &\geq 1 - O\left(n^{-1+\kappa} + n^{-\varepsilon+3\kappa/4} + \mathbb{P}(\Omega_n^c)\right), \end{aligned}$$

which finishes the proof. $\square$

The construction that we used in Theorem 7.14, to create the adversary degree distribution f_ρ has one downside. In order to construct degree distributions such that $\rho(G_n)$ is arbitrary close to zero, the value of $f^*(1)$ should be arbitrary close to 1. Therefore, these distributions might not resemble real-world situations. The reason for this downside is that the construction is based on the very crude lower bound (7.30) on Spearman's rho, for which we had to assume $f^*(1) \geq 1/2$.

The main message of Theorem 7.14 is that it is not the tail of the degree distribution that is crucial for the minimal value of $\rho(G_n)$ and we will use this result to show that when the degrees follow a regularly-varying distribution, then Spearman's rho can be arbitrary small, independent of the exponent γ .

Let F be regularly varying with exponent $\gamma > 1$ and slowly-varying function $\mathcal{L}$. Pick any $-1 < \rho < 0$ and let F_ρ be the transformed distribution, given by Theorem 7.14. We will show that F_ρ is again regularly varying with exponent γ . Note that for this it is enough to show that $(1 - F_\rho(x))x^\gamma$ is slowly varying. To this end fix $t > 0$ and write

$$\begin{aligned} \frac{1 - F_\rho(tk)}{t^{-\gamma}(1 - F_\rho(k))} &= \left(\frac{1 - F_\rho(tk)}{1 - F(tk)} \right) \left(\frac{1 - F(k)}{1 - F_\rho(k)} \right) \left(\frac{1 - F(tk)}{t^{-\gamma}(1 - F(k))} \right) \\ &= \left(\frac{1 - F_\rho(tk)}{1 - F(tk)} \right) \left(\frac{1 - F(k)}{1 - F_\rho(k)} \right) \frac{\mathcal{L}(tk)}{\mathcal{L}(k)}. \end{aligned}$$

The product of the first two terms converge to 1, as $k \rightarrow \infty$, by Theorem 7.14, while this holds for the last term since $\mathcal{L}$ is slowly varying. Summarizing, we have

$$\lim_{k \rightarrow \infty} \frac{1 - F_\rho(tk)}{t^{-\gamma}(1 - F_\rho(k))} = 1,$$

which proves that $(1 - F_\rho(x))x^\gamma$ is slowly varying and hence F_ρ is regularly varying with exponent γ . Therefore we obtain the following corollary to Theorem 7.14:

Corollary 7.15. *For any $-1 < \rho < 0$ and $\gamma > 1$, there exist distributions f and f^* , where F is regularly varying with exponent γ , such that for any sequence of graphs $G_n \in \mathcal{G}_{\eta, \varepsilon}(f, f^*)$*

$$\mathbb{P}(\rho(G_n) > \rho) \geq 1 - O\left(n^{-1+\kappa} + n^{-\varepsilon+3\kappa/4} + \mathbb{P}(\Omega_n^c)\right),$$

as $n \rightarrow \infty$, where $\kappa = \min(\varepsilon, 1/2)$.

This corollary tells us that knowing only the exponent of a regularly varying degree distribution does not give any guarantees on the minimum value of Spearman's rho on graphs with this degree distribution.

7.8 Spearman's rho on maximal disassortative graphs.

We will now use simulations to illustrate the behavior of Spearman's rho for two types of degree distributions, regularly varying and Poisson. Each of these types has a parameter that can serve as a proxy for the way in which the mass of the probability density functions is distributed over their support. For the regularly varying distributions this is the exponent γ , while for the Poisson distribution it is the mean λ . We will refer to these as the parameters of the distribution.

For the numerical experiments we have generated degree sequences $\mathbf{D}_n$ by sampling from the given distribution, using the IID algorithm, for different sizes n and values for the parameters. We then generated graphs G_n using the DGA. For each combination of size and parameter, we have generated 10^3 graphs in this

manner and have computed $\rho(G_n)$ on each of them. This gives us 10^3 samples of Spearman's rho on maximal disassortative graphs with the given size and degree distribution.

To analyze the speed of convergence of $\rho(G_n)$ we have computed for each combination of size and parameter

$$X_n := |\rho(G_n) - \mathbb{E}'[\rho(G_n)]|,$$

where $\mathbb{E}'$ denotes the empirical mean, based on the 10^3 realizations per such combination. We then plotted the empirical inverse cumulative distribution of X_n for different sizes $n = 10^4, 10^5, 10^6$ and 10^7 . The results are shown in Figures 7.4 and 7.5.

In addition, to investigate the limit of Spearman's rho in maximally disassortative graphs, we have computed $\mathbb{E}'[\rho(G_n)]$, with $n = 10^7$, for several values of the parameter of the distribution. We then plotted these values with respect to the parameter in Figure 7.6.

We will now describe the specific distributions we have used for the simulations and discuss the results.

7.8.1 Regularly-varying degree distribution

Let X have a Pareto distribution with scale 1 and shape $\gamma > 1$, i.e.

$$f_X(t) = \begin{cases} \gamma t^{-1-\gamma} & \text{if } t \geq 1 \\ 0 & \text{else,} \end{cases} \quad 1 - F_X(t) = \begin{cases} t^{-\gamma} & \text{if } t \geq 1 \\ 1 & \text{else,} \end{cases}$$

and define $\mathcal{D} = \lfloor X \rfloor$. Then we have that $1 - F(k) = 1 - F_X(k+1)$, so that F is regularly varying with exponent $\gamma > 1$, while

$$f(k) = F(k) - F(k-1) = k^{-\gamma} - (k+1)^{-\gamma}. \quad (7.32)$$

Standard calculations yield that $\sum_{k=0}^{\infty} kf(k) = \zeta(\gamma)$, where ζ is the Riemann zeta function. Therefore

$$f^*(k) = \frac{kf(k)}{\zeta(\gamma)},$$

so that $f^*(1) = (1 - 2^{-\gamma})/\zeta(\gamma)$ which is increasing in γ . Moreover $9f^*(1)^2 - 6f^*(1)^3 - 3 > -1$ for all $\gamma \geq 2.5$, which places (7.32) in the class of adversary distributions that we have considered in the previous section.

From Figure 7.4 we see that $\rho(G_n)$ is already strongly concentrated around its mean when $n = 10^5$. Even when the degree distribution has infinite variance ($\gamma = 1.5$) we have that $X_n \leq 0.025$, with high probability, for graphs of size $n = 10^5$. This shows, complementary to Theorem 7.12, that the DGA performs very well in practice with respect to the convergence of Spearman's rho to the minimal achievable value $\rho(D_*, D^*)$.

Interestingly, the simulations suggest that the concentration of $\rho(G_n)$ around its mean for graphs of small size becomes tighter when γ decreases. Compare, for instance, the plots for $n = 10^4$ in the Figure 7.4a - 7.4c.

In Figure 7.6a we have plotted the empirical average of $\rho(G_n)$ against the parameter γ of the degree mass function (7.32). Observe that in contrary to

Figure 7.4: Plot of the inverse cdf of X_n for graphs of different sizes and degree distribution (7.32), generated by the DGA, for four different choices of γ .

Figure 7.5: Plot of the inverse cdf of X_n for graphs of different sizes and Poisson degree distribution, generated by the DGA, for four different choices of λ .

Figure 7.6: Plot of the empirical average of $\tilde{\rho}(G_n)$ for graph of size 10^7 and degree distributions (7.32) and Poisson, generated by the DGA, for different values of, respectively γ and λ .

the lower bound related to $f^*(1)$, we clearly see that Spearman's rho is strongly increasing as a function of γ and $\rho(G_n) > -0.8$ for $\gamma > 2$. Therefore it follows that the Spearman's rho rank-correlation measure on any graph with degree distribution (7.32) and $\gamma > 2$ will not have a value smaller than -0.8 . Moreover, when $\gamma \geq 2.5$ we see that $\mathbb{E}'[\rho(G_n)] > -0.5$. Since this is a lower bound for Spearman's rho on any graph with degree probability mass function (7.32), a consequence could be that even if such graphs have a very disassortative joint degree structure they could potentially be classified differently.

7.8.2 Poisson degree distribution

Let $\mathcal{D}$ be a Poisson random variable with mean λ and denote its probability mass function by f . Then it follows that $f^*(k) = f(k-1)$. Hence $f^*(1) = e^{-\lambda}$ is a decreasing function of λ and $9f^*(1)^2 - 6f^*(1)^3 - 3 > -1$ for at least all $\lambda \leq 0.4$. This is opposite to the degree distribution (7.32), where $f^*(1)$ was an increasing function of the parameter γ . This is reflected in Figure 7.6b, where we see that $\mathbb{E}'[\rho(G_n)]$ decreases with λ . Here we again see that the shape of the degree distribution strongly influences the value of $\rho(G_n)$ for maximally disassortative graphs, and hence the minimal value that ρ can attain for any graph with this degree distribution. Note that, in contrast to the case with the regularly varying distribution, $\rho(G_n)$ is not monotonic with respect to λ . This could be due to the fact that the Poisson density is non-monotonic, while the density (7.32) is monotonically decreasing.

In addition we also observe that, similar to the previous setting, the DGA performs very well with respect to the convergence of $\rho(G_n)$. Already for very reasonable sizes, $n \geq 10^5$, the deviations around the mean are, with high probability, smaller than 0.02 for all three values of λ .

7.8.3 Important observations and insights

The main observation from the simulations that we did is that the distribution of the mass of the degree probability mass function is of crucial importance for the minimal value that Spearman's rho can attain. Moreover, it seems that already for very reasonable degree distributions this minimum is much larger than -1 . Therefore, one should be careful when classifying a network as not being very disassortative when a small negative value of Spearman's rho is computed.

The simulations suggest something even stronger. For this, consider the probability mass function (7.32) and observe that if we increase γ then the atoms at the end of this probability mass function loose mass, while those at the beginning gain mass. In this way we can use the parameter γ to 'shift' mass between the head and tail of the distribution. The mean, λ , of the Poisson can be used in a similar way, although in this case we need to decrease λ in order to move mass towards the head. Similar to the case with the regularly-varying degree distribution we see that, as the mass of the probability mass function f is moved towards the tail (increasing λ), the value of $\rho(G_n)$ in maximally disassortative graphs decreases and seems to approach -1 . This leads us to the following conjecture.

Conjecture 7.1. *Let $G_n^{(\gamma)}$ and $G_n^{(\lambda)}$ denote graphs generated by the DGA with, respectively, degree distribution (7.32) with parameter γ and Poisson degree dis-*

tribution with parameter λ . Then

$$\lim_{\gamma \rightarrow 1} \lim_{n \rightarrow \infty} \mathbb{E} \left[\rho(G_n^{(\gamma)}) \right] = -1 \quad \text{and} \quad \lim_{\gamma \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E} \left[\rho(G_n^{(\gamma)}) \right] = 0,$$

and

$$\lim_{\lambda \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E} \left[\rho(G_n^{(\lambda)}) \right] = -1 \quad \text{and} \quad \lim_{\lambda \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{E} \left[\rho(G_n^{(\lambda)}) \right] = 0.$$

Proving this conjecture would be a next important step in understanding the impact of the shape of the degree distribution on the value of Spearman's rho and, potentially, other rank-correlation measures.

Part III

Structural properties of the directed configuration model

Chapter 8

Scaling of the number of removed edges in the erased configuration model

Based on [86]:

van der Hoorn, P. and Litvak, N.

Upper bounds for number of removed edges in the Erased Configuration Model

Algorithms and Models for the Web Graph, Springer, 2015, 54-65

8.1 Introduction

The erased configuration model, see Chapter 6, is a very efficient model for generating simple (un)directed graphs for which the degree distributions converge to given limit distribution. This convergence result is due to the fact that the average number of edges that are removed converges to zero, as the size of the graph goes to infinity. However, for any finite instance, there will be some fraction of erased edges which could potentially effect other structural properties of the graph.

In a recent study [72], the authors found that the number of removed edges did not impact the degree sequence in any significant way. However, several other measures on the graph, for instance average neighbor degree, did seem to be altered by the removal of edges. This emphasizes the fact that asymptotic results for the average number of removed edges, by themselves, are not sufficient. We need to understand how many edges are removed, given the size of the graph and the distribution from which the degree are sampled. However until now, no such results are known.

In this chapter we address the problem of determining the scaling of the number of edges that are removed by the erased configuration model, in terms of the graph size. We will restrict our analysis to degree sequences that are generated by sampling from regularly-varying random variables, since this is often the setting used for the analysis of real-world networks with scale-free degree distributions. We will prove several scaling results for the average number of removed edges, in terms of the exponents of the distribution, for both the undirected and directed model.

First we establish an expression for the expectation of the average number of removed edges, conditioned on the degree sequence, in the undirected configuration model. We then use this expression to derive a collection of upper bounds for this term. Here each bound genuinely arises from different meaningful terms. Some follow from well-known results in the current literature, while other scaling terms are obtained by using the Stable Law CLT and the Tauberian theorem for regularly-varying random variables. The analysis and ideas will be extended to prove similar upper bounds in the directed erased configuration model. We treat the undirected model first since the analysis and ideas are easier to follow and understand in the undirected setting. We then improve our analysis of the expression for the average number of removed to obtain a tighter scaling in terms of the graph size.

The motivation for deriving the first set of upper bounds is twofold. Firstly, these bounds hint to interesting phase transitions in the scaling of the number of erased edges, in terms of the exponents of the degree distributions, for both the undirected and directed model. Here the different scaling terms come from different sources which determine the average number of removed edges. In the next chapter, we will encounter a phase transition, for Spearman's rho in the erased configuration model, that exhibits the same change in scaling as suggested for the erased edges by these results. Secondly, when trying to prove upper bounds in general, minor changes in the analysis can make the difference between finding one bound and finding a better bound. This is the case here and we therefore included the proof for both to be able to highlight where, and how, the

improvement was achieved.

We structured this chapter as follows. We start with establishing upper bounds on the scaling of the average number of erased edges in the undirected erased configuration model in Section 8.2 and extend these results to the directed model in Section 8.3. Finally we improve our bounds in Section 8.4.

8.2 Upper bounds for the number of removed edges in the undirected model

Let E_n denote the total number of edges removed by the erased configuration model. Then

$$E_n = \sum_{i=1}^n E_{ii} + \sum_{1 \leq i < j \leq n} E_{ij},$$

where E_{ij} denote the number of edges $i - j$ that have been removed. Note that, since we remove all self-loops, $E_{ii} = G_{ii}$, while the number of erased edges between two distinct nodes i and j equals the total number of edges between them (G_{ij}) minus one if there is more than one edge. Hence

$$E_n = \sum_{i=1}^n G_{ii} + \sum_{1 \leq i < j \leq n} (G_{ij} - \mathbb{1}_{\{G_{ij} > 0\}}), \quad (8.1)$$

or written differently

$$E_n = S_n + M_n, \quad (8.2)$$

where

$$S_n = \sum_{i=1}^n G_{ii} \quad \text{and} \quad M_n = \sum_{1 \leq i < j \leq n} (G_{ij} - \mathbb{1}_{\{G_{ij} > 0\}})$$

denote the total numbers of self-loops and multiple edges that are removed, respectively. We will refer to the latter as just the number of multiple edges.

In this section we will derive upper bounds for the average number of removed edges. More specifically we study $\mathbb{E}_n[E_n]/L_n$. Observe that this quantity is determined by the wiring procedure in the general configuration model. Therefore our results will be stated for graphs G_n generated by CM, i.e. before the removal of the edges.

For the scaling results we recall the $O_{\mathbb{P}}$ notation from Definition 2.4:

$$X_n = O_{\mathbb{P}}(n^{\kappa}) \iff \text{for all } \delta > 0 \quad n^{-\kappa-\delta} X_n \xrightarrow{\mathbb{P}} 0, \quad \text{as } n \rightarrow \infty.$$

Remark 8.1. Throughout this section $\mathcal{D}$ will denote an integer-valued regularly-varying random variable with exponent $\gamma > 1$. We recall that in the undirected case we consider degree sequences $\mathbf{D}_n = \text{IID}(\mathcal{D})$ as i.i.d. samples from $\mathcal{D}$.

The first result for the undirected erased configuration model are the following three upper bounds:

Theorem 8.1. *Let $G_n = \text{CM}(\mathcal{D})$. Then*

$$\frac{\mathbb{E}_n[E_n]}{L_n} = \begin{cases} O_{\mathbb{P}}\left(n^{\frac{1}{\gamma}-1}\right) & \text{if } 1 < \gamma \leq \frac{3}{2} \\ O_{\mathbb{P}}\left(n^{\frac{4}{\gamma}-3}\right) & \text{if } \frac{3}{2} < \gamma < 2 \\ O_{\mathbb{P}}\left(n^{-1}\right) & \text{if } \gamma > 2 \end{cases} \quad (8.3)$$

Figure 8.1 shows the exponent in the scaling terms in (8.3), as a function of the exponent γ .

Figure 8.1: The scaling exponent of the average number of erased edges, as a function of γ .

Theorem 8.1 gives several insights into the behavior and usefulness of the erased configuration model. First, consider the case when the degrees have finite variance ($\gamma > 2$). Equation (8.3) tells us that in that case only a number of edges that is almost finite in n will be removed. For large n , this alters the degree sequence in a negligible way. Therefore, since the erased configuration model needs to perform the random wiring step only once to generate a simple graph, it is a better model for generating simple graphs with limiting degree distributions than the repeated model which requires multiple attempts before a simple graph is produced. The latter will be a problem, especially as γ approaches 2, where the probability of generating a simple graph, albeit being strictly larger than zero, becomes very small. We already encountered this effect when running numerical experiments for the configuration model in Chapter 6 where we used $\gamma = 3.5$, since for 2.5 the repeated configuration model took more than 100 attempts to generate a simple graph.

In addition, Theorem 8.1 gives a handle on the scale of the number of removed edges, in terms of the graph size in the infinite-variance regime $1 < \gamma < 2$. This scale can be useful in determining the impact of the removal process on structural properties of graphs generated by the erased configuration model.

It will be seen from the proof in the next section that the upper bounds for $\gamma > 3/2$ in Theorem 8.1 follow readily from the literature. Our contribution is in the upper bound for $1 < \gamma < 3/2$, which is a value that corresponds to the degree distribution of many real-world networks. The proof, see Section 8.2.2,

uses the Central Limit Theorem and Tauberian Theorem for regularly-varying random variables, Theorem 2.3 and 2.2, respectively.

8.2.1 The upper bounds $n^{\frac{4}{\gamma}-3}$ and n^{-1}

From [79] we obtain the following upper bounds for the number of self-loops and multiple edges:

Proposition 8.2 ([79] Proposition 7.11). *Let $G_n = \mathcal{CM}(\mathcal{D})$. Then*

$$\mathbb{E}_n[S_n] \leq \frac{1}{2} \sum_{i=1}^n \frac{D_i^2}{L_n - 1} \quad \text{and} \quad \mathbb{E}_n[M_n] \leq 2 \left(\sum_{i=1}^n \frac{D_i(D_i - 1)}{L_n} \right)^2.$$

Now observe that it follows from (8.2) and Proposition 8.2 that

$$\frac{\mathbb{E}_n[E_n]}{L_n} \leq \frac{1}{2L_n(L_n - 1)} \sum_{i=1}^n D_i^2 + \frac{2}{L_n^3} \left(\sum_{i=1}^n D_i^2 \right)^2.$$

Since

$$\frac{4}{\gamma} - 3 > \frac{2}{\gamma} - 2,$$

for $1 < \gamma < 2$, we obtain the following result as a direct consequence of Lemma 2.5.

Lemma 8.3. *Let $G_n = \mathcal{CM}(\mathcal{D})$, then*

$$\frac{\mathbb{E}[E_n]}{L_n} = \begin{cases} O_{\mathbb{P}}\left(n^{\frac{4}{\gamma}-3}\right) & \text{if } 1 < \gamma < 2, \\ O_{\mathbb{P}}\left(n^{-1}\right) & \text{if } \gamma > 2. \end{cases} \quad (8.4)$$

Lemma 8.3 provides the last two upper bounds from Theorem 8.1. However, the bound $n^{4/\gamma-3}$ is not tight over the whole range $1 < \gamma \leq 2$ since for $1 < \gamma < 4/3$ we have $4/\gamma - 3 > 0$, and hence, in this case diverges as $n \rightarrow \infty$. This is in disagreement with the result from Chapter 6, which states that the average number of erased edges converges to zero in probability. Therefore, there must exist another upper bound on the average number of erased edges, which goes to zero as $n \rightarrow \infty$ for all $\gamma > 1$. This new bound does not follow readily from the literature. Below we establish such upper bound and explain the essential new ingredients needed for its proof.

8.2.2 The upper bound $n^{\frac{1}{\gamma}-1}$

The main ingredient for establishing this bound is a general upper bound for $\mathbb{E}_n[E_n]/L_n$ in terms of the degrees. Define the function

$$\psi(x) = x - 1 + e^{-x}. \quad (8.5)$$

Then we have the following result:

Proposition 8.4. *Let $\mathbf{D}_n$ be a degree sequence and $G_n = \mathbf{CM}(\mathbf{D}_n)$. Then*

$$\frac{\mathbb{E}_n[E_n]}{L_n} \leq \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \psi\left(\frac{D_i D_j}{L_n}\right) + O_{\mathbb{P}}\left(n^{\frac{2}{\gamma}-2}\right). \quad (8.6)$$

Proof. We first observe that since $1 < \gamma < 2$ we have, using Proposition 8.2 and Lemma 2.5, that

$$\frac{\mathbb{E}_n[S_n]}{L_n} = O_{\mathbb{P}}\left(n^{\frac{2}{\gamma}-2}\right). \quad (8.7)$$

Next, we rewrite the expression for $\mathbb{E}[M_n]/L_n$ as follows,

$$\begin{aligned} \frac{\mathbb{E}[M_n]}{L_n} &= \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \mathbb{E}_n[G_{ij} - \mathbb{1}_{\{G_{ij} > 0\}}] \\ &= \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \mathbb{E}_n[G_{ij}] - \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \mathbb{E}_n[1 - \mathbb{1}_{\{G_{ij} = 0\}}] \\ &= \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \left(\frac{D_i D_j}{L_n} - 1 + \mathbb{P}_n(G_{ij} = 0) \right). \end{aligned} \quad (8.8)$$

We can get an upper bound for $\mathbb{P}_n(E_{ij} = 0)$ from the analysis performed in [80, Section 4]. Since the probability of no edges between i and j equals the probability that none of the D_i stubs connects to one of the D_j stubs, it follows from equation [80, (4.9)] that

$$\mathbb{P}_n(E_{ij} = 0) \leq \prod_{k=0}^{D_i-1} \left(1 - \frac{D_j}{L_n - 2D_i - 1} \right) + \frac{D_i^2 D_j}{(L_n - 2D_i)^2}. \quad (8.9)$$

For the second term we note for any $\kappa > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\max_{1 \leq i \leq n} D_i > n^{\frac{1}{\gamma} + \kappa}\right) \leq \lim_{n \rightarrow \infty} n \mathbb{P}\left(D > n^{\frac{1}{\gamma} + \kappa}\right) = 0.$$

Now fix $\kappa = (\gamma - 1)/2\gamma > 0$, so that $(1/\gamma) + \kappa = 1 - \kappa$. Then, for any $\delta, K > 0$,

$$\begin{aligned} &\mathbb{P}\left(\frac{1}{L_n} \sum_{1 \leq i < j \leq n} \frac{D_i^2 D_j}{(L_n - 2D_i)^2} > K n^{\frac{2}{\gamma}-2+\delta}\right) \\ &\leq \mathbb{P}\left(\frac{1}{L_n} \sum_{1 \leq i < j \leq n} \frac{D_i^2 D_j}{(L_n - 2D_i)^2} > K n^{\frac{2}{\gamma}-2+\delta}, \max_{1 \leq i \leq n} D_i \leq n^{\frac{1}{\gamma} + \kappa}\right) \\ &\quad + \mathbb{P}\left(\max_{1 \leq i \leq n} D_i > n^{\frac{1}{\gamma} + \kappa}\right) \\ &\leq \mathbb{P}\left(\frac{1}{(L_n - 2n^{1-\kappa})^2} \sum_{i=1}^n D_i^2 > K n^{\frac{2}{\gamma}-2+\delta}\right) + \mathbb{P}\left(\max_{1 \leq i \leq n} D_i > n^{\frac{1}{\gamma} + \kappa}\right). \end{aligned}$$

Since the first term in the last line converges to zero, as $n \rightarrow \infty$, by Proposition 2.4, it follows that

$$\frac{1}{L_n} \sum_{1 \leq i < j \leq n} \frac{D_i^2 D_j}{(L_n - 2D_i)^2} = O_{\mathbb{P}}\left(n^{\frac{2}{\gamma}-2}\right).$$

Inserting this result into (8.9) and using that

$$\prod_{k=0}^{D_i-1} \left(1 - \frac{D_j}{L_n - 2D_i - 1}\right) \leq \exp \left\{ -\frac{D_i D_j}{L_n} \right\},$$

we obtain

$$\frac{1}{L_n} \sum_{1 \leq i < j \leq n} \mathbb{P}_n(E_{ij} = 0) \leq \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \exp \left\{ -\frac{D_i D_j}{L_n} \right\} + O_{\mathbb{P}} \left(n^{\frac{2}{\gamma}-2} \right). \quad (8.10)$$

Plugging (8.10) back into (8.8) and using (8.7) gives

$$\begin{aligned} \frac{\mathbb{E}_n[E_n]}{L_n} &= \frac{\mathbb{E}_n[S_n]}{L_n} + \frac{\mathbb{E}_n[M_n]}{L_n} \\ &= \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \left(\frac{D_i D_j}{L_n} - 1 + \mathbb{P}_n(G_{ij} = 0) \right) + O_{\mathbb{P}} \left(n^{\frac{2}{\gamma}-2} \right) \\ &\leq \frac{1}{L_n} \sum_{1 \leq i < j \leq n} \left(\frac{D_i D_j}{L_n} - 1 + \frac{1}{L_n} \exp \left\{ -\frac{D_i^+ D_j^-}{L_n} \right\} \right) + O_{\mathbb{P}} \left(n^{\frac{2}{\gamma}-2} \right), \end{aligned}$$

which proves (8.6). $\square$

We will use (8.6) to obtain the scaling $n^{\frac{1}{\gamma}-1}$. The idea is to compare the first term on the right hand side of (8.6) to

$$n \mathbb{E} \left[\frac{\mathcal{D}_1 \mathcal{D}_2}{\nu n} - 1 + \exp \left\{ -\frac{\mathcal{D}_1 \mathcal{D}_2}{\nu n} \right\} \right], \quad (8.11)$$

where $\mathcal{D}_1$ and $\mathcal{D}_2$ are independent regularly-varying with exponent γ , so that their product is again regularly-varying with exponent γ . It then follows from the Tauberian theorem (Theorem 2.2) that the expectation scales as $n^{-\gamma}$ and hence (8.11) scales as $n^{1-\gamma}$, which, for $\gamma > 1$, is of smaller order than $n^{\frac{1}{\gamma}-1}$. The latter scaling will come from the term $|L_n - \nu n|/L_n$, see Lemma 2.5, which we encounter when comparing (8.6) to (8.11).

Proposition 8.5. *Let $G_n = \text{CM}(\mathcal{D})$, where $\mathcal{D}$ is regularly varying with exponent $1 < \gamma < 2$. Then*

$$\frac{\mathbb{E}_n[E_n]}{L_n} = O_{\mathbb{P}} \left(n^{\frac{1}{\gamma}-1} \right). \quad (8.12)$$

Proof. We start with equation (8.6). Since $\gamma > 1$, the last term $O_{\mathbb{P}}(n^{(2/\gamma)-2})$ is also $O_{\mathbb{P}}(n^{(1/\gamma)-1})$. We will rewrite the first term on the right hand side of (8.6) as follows, by extracting a factor $n(n-1)/2L_n$,

$$\frac{n(n-1)}{2L_n} \left(\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi \left(\frac{D_i D_j}{L_n} \right) \right). \quad (8.13)$$

Since the term outside the brackets is $O_{\mathbb{P}}(n)$ it suffices to show that

$$\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi \left(\frac{D_i D_j}{L_n} \right) = O_{\mathbb{P}} \left(n^{\frac{1}{\gamma}-2} \right).$$

For this observe that

$$|\psi(x) - \psi(y)| = |x - y + e^{-x} - e^{-y}| \leq 2|x - y|, \quad (8.14)$$

hence

$$\begin{aligned} \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi\left(\frac{D_i D_j}{L_n}\right) &\leq \left| \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi\left(\frac{D_i D_j}{L_n}\right) - \psi\left(\frac{D_i D_j}{\nu n}\right) \right| \\ &\quad + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi\left(\frac{D_i D_j}{\nu n}\right) \\ &\leq \frac{4}{n(n-1)} \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| \sum_{1 \leq i < j \leq n} D_i D_j \quad (8.15) \end{aligned}$$

$$+ \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi\left(\frac{D_i D_j}{\nu n}\right), \quad (8.16)$$

where we used (8.14) to get (8.15).

Since

$$\frac{4}{n(n-1)} \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| \sum_{1 \leq i < j \leq n} D_i D_j \leq \frac{|L_n - \nu n|}{\nu n^3 L_n} \sum_{i,j=1}^n D_i D_j,$$

it follows from Lemma 2.5 that (8.15) is $O_{\mathbb{P}}(n^{(1/\gamma)-2})$. Next, observe that the function $\psi(x)$ is positive so that for any $\delta, K > 0$ we have, by Markov's inequality,

$$\begin{aligned} \mathbb{P}\left(\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi\left(\frac{D_i D_j}{\nu n}\right) > K n^{-\gamma+\delta}\right) \\ \leq \frac{n^{\gamma-\delta}}{K} \mathbb{E}\left[\psi\left(\frac{D_1 D_2}{\nu n}\right)\right], \quad (8.17) \end{aligned}$$

where D_1 and D_2 are two independent regularly-varying random variables with exponent $1 < \gamma < 2$. Consequently, see for instance the Corollary after Theorem 3 in [41], the product $D_1 D_2$ is again regularly varying with the same exponent. Now we use Theorem 2.2 ii) to find that (8.17) converges to zero, as $n \rightarrow \infty$, and hence (8.16) is $O_{\mathbb{P}}(n^{-\gamma})$. We have thus shown that

$$\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \psi\left(\frac{D_i D_j}{L_n}\right) \leq O_{\mathbb{P}}\left(n^{\frac{1}{\gamma}-2} + n^{-\gamma}\right) = O_{\mathbb{P}}\left(n^{(1/\gamma)-2}\right)$$

since $-\gamma < (1/\gamma) - 2$ for all $\gamma > 1$, from which the result follows. $\square$

8.3 Upper bounds for the number of removed edges in the directed model

We will now turn towards analyzing the scaling of the average number of removed edges in the directed erased configuration model. The strategy for obtaining the

results is the same as for the undirected version. We will first use known bounds on the number of self-loops and multiple edges to derive several upper bounds for all cases $\gamma_+, \gamma_- > 1$. Then we will derive a more accurate expression for the number of removed edges and use Theorem 2.2, to obtain a different upper bound. The scaling of $\mathbb{E}_n[E_n]/L_n$ is then determined by the taking, from all available upper bounds, the bound that scales the fastest. However, since the bounds will be dependent on both γ_+ and γ_- , the result will look less clean than it did for the undirected case.

Remark 8.2. Similar to the undirected case, throughout this section $\mathcal{D}^+$ and $\mathcal{D}^-$ will be two independent, integer-valued, regularly-varying random variables with exponent $\gamma_+, \gamma_- > 1$ and expectation $\mathbb{E}[\mathcal{D}^+] = \nu = \mathbb{E}[\mathcal{D}^-]$. In addition we use the following shorthand notations

$$\gamma_* = (\gamma_+ \wedge \gamma_-) \quad \text{and} \quad \gamma^* = (\gamma_+ \vee \gamma_-),$$

for, respectively, the minimum and maximum of the exponents γ_+ and γ_- .

We now have the following upper bounds for the average number of removed edges in the directed erased configuration model.

Theorem 8.6. *Let $G_n = \mathcal{CM}(\mathcal{D}^+, \mathcal{D}^-)$, with $\mathcal{D}^+$ and $\mathcal{D}^-$ independent. Then*

$$\frac{\mathbb{E}_n[E_n]}{L_n} = \begin{cases} O_{\mathbb{P}}\left(n^{\frac{1}{\gamma^*}-1}\right) & \text{if } \gamma_* \leq \frac{\gamma^*}{2(\gamma^*-1)} \\ O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_+}+\frac{2}{\gamma_-}-3}\right) & \text{if } \frac{\gamma^*}{2(\gamma^*-1)} < \gamma_* \text{ and } \gamma^* < 2 \\ O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_+}-2}\right) & \text{if } \gamma_+ < 2 \text{ and } \gamma_- > 2 \\ O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_-}-2}\right) & \text{if } \gamma_- < 2 \text{ and } \gamma_+ > 2 \\ O_{\mathbb{P}}(n^{-1}) & \text{if } \gamma_* > 2 \end{cases} \quad (8.18)$$

We see from Theorem 8.6 that, similarly to the undirected case, the average total number of edges removed by the erased configuration model scales as n^{-1} when both the out- and in-degree distribution have finite variance. Therefore, even in this setting the erased model should be preferred over the repeated configuration model for generating simple graphs.

8.3.1 Transition to i.i.d. degrees

The techniques for the proof of Theorem 8.6 will, in the end, be similar to those for the undirected version, Theorem 8.1. However, we remark that for those proofs we heavily used Lemma 2.5 for which we made use of the fact that the degree sequence $\mathbf{D}_n$ could be seen as an i.i.d. sample from the random variable $\mathcal{D}$. The latter is no longer true in the directed setting. However, we can still show that the scaling of terms with out- and in-degrees are similar to those with the corresponding i.i.d. samples from, respectively, $\mathcal{D}^+$ and $\mathcal{D}^-$.

Recall that when $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$ then we have that

$$D_i^+ = \mathcal{D}_i^+ + \xi_i^+, \quad D_j^- = \mathcal{D}_j^- + \xi_j^-,$$

where $\mathcal{D}_i^+$ and $\mathcal{D}_j^-$ are i.i.d. samples from, respectively, $\mathcal{D}^+$ and $\mathcal{D}^-$ and $\xi_i^\pm$ were indicators. In addition, recall that

$$\Delta_n = \sum_{i=1}^n \mathcal{D}_i^+ - \mathcal{D}_i^-,$$

controls, in some sense, the difference between the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ and the joint sequence of i.i.d. samples $(\mathcal{D}_n^+, \mathcal{D}_n^-)$. Moreover, the indicators $\xi_i^\pm$ satisfy

$$\sum_{i=1}^n (\xi_i^+ + \xi_i^-) = |\Delta_n| \quad \text{and} \quad \sum_{i=1}^n \sum_{j=1}^n \xi_i^+ \xi_j^- = 0.$$

The next lemma summarizes the scaling results we will need to prove Theorem 8.6.

Lemma 8.7. *Let $(\mathbf{D}_n^+, \mathbf{D}_n^-) = \text{IID}(\mathcal{D}^+, \mathcal{D}^-)$, with $\mathcal{D}^+$ and $\mathcal{D}^-$ independent. Then, for any integer $p \geq 0$,*

- i) $|\Delta_n| = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*}}\right),$
- ii) $|L_n - \nu n| = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*}}\right),$
- iii) $\left(\frac{1}{L_n}\right)^p = O_{\mathbb{P}}\left(n^{-p}\right),$
- iv) $\sum_{i=1}^n (D_i^\pm)^2 = O_{\mathbb{P}}\left(n^{(\frac{2}{\gamma_\pm} \vee 1)}\right)$ and
- v) $\frac{1}{n^2} \sum_{i,j=1}^n \left| \frac{D_i^+ D_j^-}{L_n} - \frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n} \right| = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*} - 2}\right).$

Comparing Lemma 8.7 ii), iii) and iv) to, respectively, Lemma 2.5 i), ii) and iii), we see that, as far as the scaling is concerned, we can treat terms consisting of sums of powers of out- and in-degrees, as if they were sums of powers of i.i.d. samples from $\mathcal{D}^+$ and $\mathcal{D}^-$. This is an important insight which will make it easier to understand scaling results, and follow proofs, that involve such terms.

Proof of Lemma 8.7.

i) Note that since $\mathbb{E}[\mathcal{D}^+] = \nu = \mathbb{E}[\mathcal{D}^-]$, it follows from Lemma 2.5 i) that

$$|\Delta_n| \leq \left| \sum_{i=1}^n \mathcal{D}_i^+ - \nu n \right| + \left| \sum_{i=1}^n \mathcal{D}_i^- - \nu n \right| = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_+}} + n^{\frac{1}{\gamma_-}}\right) = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*}}\right).$$

ii) Using i) and Lemma 2.5 i), we obtain

$$|L_n - \nu n| \leq \left| \sum_{i=1}^n \mathcal{D}_i^+ - \nu n \right| + |\Delta_n| = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_+}} + n^{\frac{1}{\gamma_-}}\right) = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*}}\right).$$

iii) Since $L_n \geq \sum_{i=1}^n \mathcal{D}_i^+$, the result follows from Lemma 2.5 ii).

iv) Since the proofs for the out- and in-degree are similar we will only consider the former:

$$\begin{aligned} \sum_{i=1}^n (D_i^+)^2 &= \sum_{i=1}^n (\mathcal{D}_i^+ + 2\mathcal{D}_i^+ \xi_i^+ + \xi_i^+) \\ &\leq \sum_{i=1}^n (\mathcal{D}_i^+)^2 + 2 \sum_{i=1}^n \mathcal{D}_i^+ + |\Delta_n| \\ &= O_{\mathbb{P}} \left(n^{\frac{2}{\gamma_+}} + n + n^{\frac{1}{\gamma_+}} \right) \\ &= O_{\mathbb{P}} \left(n^{(\frac{2}{\gamma_+} \vee 1)} \right) \end{aligned}$$

v) First we bound this term as follows:

$$\begin{aligned} &\frac{1}{n^2} \sum_{i,j=1}^n \left| \frac{D_i^+ D_j^-}{L_n} - \frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n} \right| \\ &\leq \frac{1}{n^2} \sum_{i,j=1}^n \left| \frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{L_n} - \frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n} \right| + \frac{1}{n^2} \sum_{i,j=1}^n \left| \frac{\mathcal{D}_i^+ x_{ij}^- + \xi_i^+ \mathcal{D}_j^-}{L_n} \right| \\ &\leq \frac{1}{n^2} \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| \sum_{i,j=1}^n \mathcal{D}_i^+ \mathcal{D}_j^- + \frac{|\Delta_n|}{n^2} \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| \sum_{k=1}^n \mathcal{D}_k^+ + \mathcal{D}_k^- \quad (8.19) \end{aligned}$$

Note that from ii) and iii) it follows that,

$$\left| \frac{1}{L_n} - \frac{1}{\nu n} \right| = \frac{|L_n - \nu n|}{L_n \nu n} = O_{\mathbb{P}} \left(n^{\frac{1}{\gamma_*} - 2} \right).$$

Furthermore, by Lemma 2.5 i)

$$\frac{1}{n^2} \sum_{i,j=1}^n \mathcal{D}_i^+ \mathcal{D}_j^- = O_{\mathbb{P}}(1) \quad \text{and} \quad \frac{1}{n^2} \sum_{k=1}^n (\mathcal{D}_k^+ + \mathcal{D}_k^-) = O_{\mathbb{P}}(n^{-1}).$$

Combining this, and using i), we have the following scaling results for the two terms in (8.19)

$$\begin{aligned} &\frac{1}{n^2} \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| \sum_{i,j=1}^n \mathcal{D}_i^+ \mathcal{D}_j^- = O_{\mathbb{P}} \left(n^{\frac{1}{\gamma_*} - 2} \right) \quad \text{and} \\ &\frac{|\Delta_n|}{n^2} \left| \frac{1}{L_n} - \frac{1}{\nu n} \right| \sum_{k=1}^n (\mathcal{D}_k^+ + \mathcal{D}_k^-) = O_{\mathbb{P}} \left(n^{\frac{2}{\gamma_*} - 3} \right). \end{aligned}$$

Since for all $\gamma_+, \gamma_- > 1$ it holds that

$$\frac{2}{\gamma_*} - 3 < \frac{1}{\gamma_*} - 2,$$

the last term is $O_{\mathbb{P}} \left(n^{\frac{1}{\gamma_*} - 2} \right)$, from which the result follows. $\square$

8.3.2 The first set of upper bounds

Similar to the undirected case we write

$$S_n = \sum_{i=1}^n G_{ii} \quad \text{and} \quad M_n = \sum_{1 \leq i \neq j \leq n} (G_{ij} - \mathbb{1}_{\{G_{ij} > 0\}}),$$

for, respectively, the number of self-loops and multiple edges. We start by deriving upper bounds for the scaling of both these quantities.

Proposition 8.8. *Let $G_n = \mathbf{CM}(\mathcal{D}^+, \mathcal{D}^-)$, with $\mathcal{D}^+$ and $\mathcal{D}^-$ independent. Then*

$$\frac{\mathbb{E}_n[S_n]}{L_n} = O_{\mathbb{P}}(n^{-1}),$$

and

$$\frac{\mathbb{E}_n[M_n]}{L_n} = \begin{cases} O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3}\right) & \text{if } \gamma_* < 2, \\ O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_+} - 2}\right) & \text{if } \gamma_+ < 2, \text{ and } \gamma_- > 2 \\ O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_-} - 2}\right) & \text{if } \gamma_- < 2, \text{ and } \gamma_+ > 2 \\ O_{\mathbb{P}}(n^{-1}) & \text{if } \gamma_* > 2. \end{cases}$$

Proof. We start with the self-loops. First note that by Lemma 8.7,

$$\begin{aligned} \frac{\mathbb{E}_n[S_n]}{L_n} &= \frac{1}{L_n} \mathbb{E}_n \left[\sum_{i=1}^n G_{ii} \right] \\ &= \sum_{i=1}^n \frac{D_i^+ D_i^-}{L_n^2} \\ &= \frac{1}{L_n^2} \sum_{i=1}^n \mathcal{D}_i^+ \mathcal{D}_i^- + \frac{1}{L_n^2} \sum_{i=1}^n (\mathcal{D}_i^+ + \mathcal{D}_i^-) + \frac{|\Delta_n|}{L_n^2} \\ &= \frac{1}{L_n^2} \sum_{i=1}^n \mathcal{D}_i^+ \mathcal{D}_i^- + O_{\mathbb{P}}\left(n^{-1} + n^{\frac{1}{\gamma_*} - 2}\right). \end{aligned}$$

Since $\mathcal{D}^+$ and $\mathcal{D}^-$ are independent regularly-varying with exponent γ_+ and γ_- , the product $\mathcal{D}^+ \mathcal{D}^-$ is again regularly varying with exponent $\gamma_* > 1$. Therefore, it follows from Lemma 2.5 i) and Lemma 8.7 ii) that

$$\frac{1}{L_n^2} \sum_{i=1}^n \mathcal{D}_i^+ \mathcal{D}_i^- = O_{\mathbb{P}}(n^{-1}),$$

from which the result follows, since $\frac{1}{\gamma_*} - 2 < -1$.

Next, we first observe that if we bound the number of multiple edges between two distinct nodes i and j by the number of pairs of edges between them, then we can follow the approach in the proof of [79, Proposition 7.11] to obtain

$$\mathbb{E}_n[M_n] \leq \sum_{1 \leq i \neq j \leq n} \frac{D_i^+(D_i^+ - 1)D_j^-(D_j^- - 1)}{L_n(L_n - 1)} \leq \sum_{i,j=1}^n \frac{(D_i^+)^2(D_j^-)^2}{L_n^2}. \quad (8.20)$$

Here for the last line we used that there can only be multiple edges when $L_n \geq 2$ and in this case $2L_n(L_n - 1) \geq L_n^2$. With this upper bound and Lemma 8.7 ii) and iv) we get

$$\begin{aligned} \frac{\mathbb{E}_n[M_n]}{L_n} &\leq \frac{2}{L_n^3} \left(\sum_{i=1}^n (D_i^+)^2 \right) \left(\sum_{j=1}^n (D_j^-)^2 \right) \\ &= O_{\mathbb{P}} \left(n^{\left(\frac{2}{\gamma_+} \vee 1\right) + \left(\frac{2}{\gamma_-} \vee 1\right) - 3} \right). \end{aligned}$$

From this the result follows since

$$\left(\frac{2}{\gamma_+} \vee 1 \right) + \left(\frac{2}{\gamma_-} \vee 1 \right) - 3 = \begin{cases} \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3 & \text{if } \gamma^* < 2, \\ \frac{2}{\gamma_+} - 2 & \text{if } 1 < \gamma_+ < 2 \text{ and } \gamma_- > 2, \\ \frac{2}{\gamma_-} - 2 & \text{if } \gamma_+ > 2 \text{ and } 1 < \gamma_- < 2, \\ -1 & \text{if } \gamma_* > 2. \end{cases}$$

□

Note that since $E_n = S_n + M_n$, Proposition 8.8 implies that

$$\frac{\mathbb{E}_n[E_n]}{L_n} = \begin{cases} O_{\mathbb{P}} \left(n^{\frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3} \right) & \text{if } \gamma^* < 2, \\ O_{\mathbb{P}} \left(n^{\frac{2}{\gamma_+} - 2} \right) & \text{if } \gamma_+ < 2, \text{ and } \gamma_- > 2 \\ O_{\mathbb{P}} \left(n^{\frac{2}{\gamma_-} - 2} \right) & \text{if } \gamma_- < 2, \text{ and } \gamma_+ > 2 \\ O_{\mathbb{P}} (n^{-1}) & \text{if } \gamma_* > 2. \end{cases} \quad (8.21)$$

Again, as was the case for the undirected model, the bound for $\gamma^* < 2$ is not tight, since

$$\frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3 > 0 \quad \text{whenever} \quad \gamma_+ < \frac{2\gamma_-}{3\gamma_- - 2}.$$

The last inequality can be satisfied for $\gamma_+ > 1$, when $\gamma_- < 2$ and vice versa. Therefore, in the regime $\gamma^* < 2$ there is a region where the upper bound diverges, while $\mathbb{E}_n[E_n]/L_n \xrightarrow{\mathbb{P}} 0$, as $n \rightarrow \infty$. We will now establish a tighter bound, using a similar strategy as for the undirected case.

8.3.3 An upper bound that scales as $n^{\frac{1}{\gamma_*} - 1}$

For the directed case we have the following upper bound for the average number of erased edges.

Proposition 8.9. *Let $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ be a bi-degree sequence and $G_n = \mathbf{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$. Then,*

$$\frac{\mathbb{E}_n[E_n]}{L_n} \leq \frac{1}{L_n} \sum_{i,j=1}^n \psi \left(\frac{D_i^+ D_j^-}{L_n} \right) + O_{\mathbb{P}}(n^{-1}), \quad (8.22)$$

where $\psi(x)$ is defined as in (8.5).

Proof. The proof follows a similar strategy as for the undirected case. We will first derive an expression for $\mathbb{P}_n(G_{ij} = 0)$ in the directed configuration model. For this we will adopt the analysis in [80] to the directed setting.

Let $p(A, B, L)$ denote the probability that, in the directed model, none of the outbound stubs from a set of size A connects to one of the inbound stubs from a set of size B , given that there are L outbound/inbound stubs left in total. Since we can only select A inbound stubs outside of the set of size B if $A + B \leq L$ and the probability of selecting the first such stub is $1 - B/L$, we get

$$p(A, B, L) = \mathbb{1}_{\{A+B \leq L\}} \left(1 - \frac{B}{L}\right) p(A-1, B, L-1).$$

Continuing this recursion yields,

$$p(A, B, L) = \mathbb{1}_{\{A+B \leq L\}} \prod_{k=0}^{A-1} \left(1 - \frac{B}{L-k}\right), \quad (8.23)$$

from which it follows that

$$\mathbb{P}(G_{ij} = 0) = \mathbb{1}_{\{D_i^+ + D_j^- \leq L_n\}} \prod_{k=0}^{D_i^+ - 1} \left(1 - \frac{D_j^-}{L_n - k}\right). \quad (8.24)$$

Since, on the event $D_i^+ + D_j^- \leq L_n$, the product term can be bounded from above by $\exp\{-D_i^+ D_j^- / L_n\}$, we have that

$$\begin{aligned} \frac{\mathbb{E}[M_n]}{L_n} &= \frac{1}{L_n} \sum_{1 \leq i \neq j \leq n} \mathbb{E}_n[G_{ij} - \mathbb{1}_{\{G_{ij} > 0\}}] \\ &\leq \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{E}_n[G_{ij} - \mathbb{1}_{\{G_{ij} > 0\}}] \\ &= \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{E}_n[G_{ij}] - \frac{n^2}{L_n} + \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{P}_n(G_{ij} = 0) \\ &= \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{E}_n[G_{ij}] - \frac{n^2}{L_n} \\ &\quad + \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{1}_{\{D_i^+ + D_j^- \leq L_n\}} \prod_{k=0}^{D_i^+ - 1} \left(1 - \frac{D_j^-}{L_n - k}\right) \\ &\leq \frac{1}{L_n} \sum_{i,j=1}^n \mathbb{E}_n[G_{ij}] - \frac{n^2}{L_n} + \frac{1}{L_n} \sum_{i,j=1}^n \exp\left\{-\frac{D_i^+ D_j^-}{L_n}\right\} \\ &= \frac{1}{L_n} \sum_{i,j=1}^n \psi\left(\frac{D_i^+ D_j^-}{L_n}\right). \end{aligned}$$

The result now follows since $E_n = S_n + M_n$ and $\mathbb{E}_n[S_n]/L_n = O_{\mathbb{P}}(n^{-1})$ by Proposition 8.8. $\square$

Recall from Section 8.2.2 that Proposition 8.4 was one of the main ingredients for the proof of Proposition 8.5. In a similar fashion we use Proposition 8.9 to obtain the following scaling bound.

Proposition 8.10. *Let $G_n = \mathcal{CM}(\mathcal{D}^+, \mathcal{D}^-)$, with $\mathcal{D}^+$ and $\mathcal{D}^-$ independent and $1 < \gamma_+, \gamma_- < 2$. Then,*

$$\frac{\mathbb{E}_n[E_n]}{|L_n|} = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*}-1}\right).$$

Note that this result is almost identical to Proposition 8.5, with the exception that here we have γ_* instead of γ . However, if we consider the bi-directed version of the undirected graph where each edge $i = j$ is replaced with $i \rightarrow j$ and $j \rightarrow i$, then the out- and in-degrees can be seen as i.i.d. samples from the same random variable $\mathcal{D}$. Hence, $\gamma_* = \gamma$, which relates Proposition 8.10 to Proposition 8.5.

The idea for the proof follows the same arguments as those for the proof of Proposition 8.4. We will compare the expectation of (8.22) to

$$n\mathbb{E}\left[\psi\left(\frac{\mathcal{D}^+\mathcal{D}^-}{\nu n}\right)\right],$$

which scales as $n^{1-\gamma_*}$, by Theorem 2.2, while the error term will scale as $n^{(1/\gamma_*)-1}$.

Proof of Proposition 8.10. First we rewrite the expression on the right hand side of (8.22) as

$$\frac{n^2}{L_n} \left(\frac{1}{n^2} \sum_{i,j=1}^n \psi\left(\frac{D_i^+ D_j^-}{L_n}\right) \right) + O_{\mathbb{P}}(n^{-1}).$$

Since $1 - \gamma_* > 1$ and $n^2/L_n = O_{\mathbb{P}}(n)$, it is enough to show that

$$\frac{1}{n^2} \sum_{i,j=1}^n \psi\left(\frac{D_i^+ D_j^-}{L_n}\right) = O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*}-2}\right). \quad (8.25)$$

For this we will bound this term as follows, using (8.14),

$$\begin{aligned} \frac{1}{n^2} \sum_{i,j=1}^n \psi\left(\frac{D_i^+ D_j^-}{L_n}\right) &\leq \frac{1}{n^2} \sum_{i,j=1}^n \left| \psi\left(\frac{D_i^+ D_j^-}{L_n}\right) - \psi\left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n}\right) \right| \\ &\quad + \frac{1}{n^2} \sum_{i,j=1}^n \psi\left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n}\right) \\ &\leq \frac{2}{n^2} \sum_{i,j=1}^n \left| \frac{D_i^+ D_j^-}{L_n} - \frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n} \right| \quad (8.26) \\ &\quad + \frac{1}{n^2} \sum_{i,j=1}^n \psi\left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n}\right) \quad (8.27) \end{aligned}$$

It follow from Lemma 8.7 v) that (8.26) is $O_{\mathbb{P}}\left(n^{\frac{1}{\gamma_*}-2}\right)$. For (8.27) we first note that $\mathcal{D}_i^+$ and $\mathcal{D}_j^-$ are independent for all $1 \leq i, j \leq n$. Therefore, since $\psi(x)$ is

positive, we have that for all $\delta, K > 0$

$$\begin{aligned} \mathbb{P} \left(\left| \frac{1}{n^2} \sum_{i,j=1}^n \psi \left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n} \right) \right| > K n^{-\gamma_* + \delta} \right) \\ \leq \frac{n^{\gamma_* - \delta}}{K} \mathbb{E} \left[\psi \left(\frac{\mathcal{D}^+ \mathcal{D}^-}{\nu n} \right) \right]. \end{aligned}$$

From Theorem 2.2 and the fact that $\mathcal{D}^+ \mathcal{D}^-$ is again regularly varying with exponent $\gamma_* > 1$ we have that for any $\delta > 0$,

$$\lim_{n \rightarrow \infty} \frac{n^{\gamma_* - \delta}}{K} \mathbb{E} \left[\frac{1}{n^2} \sum_{i,j=1}^n \psi \left(\frac{\mathcal{D}^+ \mathcal{D}^-}{\nu n} \right) \right] = 0,$$

from which we deduce that (8.27) is $O_{\mathbb{P}}(n^{-\gamma})$. The result now follows by observing that, for all $\gamma_* > 1$, it holds that $1 - \gamma_* < \frac{1}{\gamma_*} - 1$, so that

$$\frac{n^{\gamma_* - \delta}}{K} \mathbb{E} \left[\psi \left(\frac{\mathcal{D}^+ \mathcal{D}^-}{\nu n} \right) \right] = O_{\mathbb{P}} \left(n^{\frac{1}{\gamma_*} - 1} + n^{1 - \gamma_*} \right) = O_{\mathbb{P}} \left(n^{\frac{1}{\gamma_*} - 1} \right),$$

as $n \rightarrow \infty$, which completes the proof. $\square$

We will end this section by proving Theorem 8.6.

Proof of Theorem 8.6. Consider the case where $\gamma^* < 2$. Then, by (8.21) and Proposition 8.10, it follows that the exponent in the scaling of $\mathbb{E}_n[E_n]/L_n$ is determined by the smallest out of

$$\frac{1}{\gamma_*} - 1 \quad \text{and} \quad \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3.$$

Since

$$\frac{1}{\gamma_*} - 1 \leq \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3 \iff \gamma_* \leq \frac{\gamma^*}{2(\gamma^* - 1)},$$

we obtain the first two cases in (8.18). The other three cases follow from (8.21) and the fact that

$$\frac{2}{\gamma_+} - 2 \leq \frac{1}{\gamma_*} - 1,$$

for all $1 < \gamma_+ < 2$ and $\gamma_- > 2$, and similar for $\frac{2}{\gamma_-} - 2$. $\square$

8.4 A better scaling for average number of removed edges

Theorem 8.1 and Theorem 8.6 give some bounds on the scaling of the average number of erased edges in the undirected and directed erased configuration model, respectively. This is an important contribution to the literature, since no results

on the scaling of the number of erased edges have been known. However, an even more striking implication of both theorems is the change of scaling that seems to occur within the regions $1 < \gamma < 2$ and $1 < \gamma_+, \gamma_- < 2$ for, respectively, the undirected and directed model. This change is due to the interaction between the contributions to the scaling from different sources: finite variance, Stable Law CLT and the Tauberian Theorem. We will see in the next chapter that such a phase-transition occurs for rank-correlations in the erased configuration model. This shows that in this case, the above sources all play non-trivial roles in determining the scaling.

For the average number of removed edges however, it turns out that we can obtain a tighter upper bound. We will first consider the directed case.

Theorem 8.11. *Let $G_n = \mathcal{CM}(\mathcal{D}^+, \mathcal{D}^-)$, with $\mathcal{D}^+$ and $\mathcal{D}^-$ independent and $1 < \gamma_+, \gamma_- < 2$. Then*

$$\mathbb{E}_n[E_n] = O_{\mathbb{P}}(n^{2-\gamma_*}).$$

Since by Lemma 8.7 iii) $L_n^{-1} = O_{\mathbb{P}}(n^{-1})$, the above result is equivalent to

$$\frac{\mathbb{E}_n[E_n]}{L_n} = O_{\mathbb{P}}(n^{1-\gamma_*}). \quad (8.28)$$

Now observe that for all $1 < \gamma_+, \gamma_- < 2$

$$1 - \gamma_* \leq \frac{1}{\gamma} - 2 \quad \text{and} \quad 1 - \gamma_* \leq \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3,$$

which implies that the bound $n^{1-\gamma_*}$ is more tight than the upper bounds derived in Theorem 8.6. For the other three cases for $\gamma_{\pm}$ we see that the bounds in the previous theorem still hold. Therefore, we have the following corollary:

Corollary 8.12. *Let $G_n = \mathcal{CM}(\mathcal{D}^+, \mathcal{D}^-)$, with $\mathcal{D}^+$ and $\mathcal{D}^-$ independent. Then*

$$\frac{\mathbb{E}_n[E_n]}{L_n} = \begin{cases} O_{\mathbb{P}}(n^{1-\gamma_*}) & \text{if } 1 < \gamma_+, \gamma_- < 2, \\ O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_+}-2}\right) & \text{if } \gamma_+ < 2 \text{ and } \gamma_- > 2, \\ O_{\mathbb{P}}\left(n^{\frac{2}{\gamma_-}-2}\right) & \text{if } \gamma_- < 2 \text{ and } \gamma_+ > 2, \\ O_{\mathbb{P}}(n^{-1}) & \text{if } \gamma_* > 2. \end{cases}$$

To understand where the improved upper bound in Theorem 8.11 comes from let $\psi(x)$ be as in (8.5) and define

$$\mathcal{E}_n = \sum_{i,j=1}^n \psi\left(\frac{D_i^+ D_j^-}{L_n}\right).$$

Then we obtain from Proposition 8.9 that

$$\mathbb{E}_n[E_n] \leq \mathcal{E}_n + L_n O_{\mathbb{P}}(n^{-1}).$$

By Lemma 8.7 iii) the second term is $O_{\mathbb{P}}(1)$ which is $O_{\mathbb{P}}(n^{2-\gamma_*})$, since for all $1 < \gamma_+, \gamma_- < 2$ we have that $2 - \gamma_* > 0$.

Now write

$$\mathcal{E}_n = \sum_{i,j=1}^n \psi \left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\mu n} \right) \quad (8.29)$$

$$+ \sum_{i,j=1}^n \left[\psi \left(\frac{D_i^+ D_j^-}{L_n} \right) - \psi \left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\mu n} \right) \right]. \quad (8.30)$$

and consider (8.29). Then it follows, from Theorem 2.2 and the fact that, since $\mathcal{D}^+$ and $\mathcal{D}^-$ are independent, $\mathcal{D}^+ \mathcal{D}^-$ is regularly varying with exponent γ_* , that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \left[\psi \left(\frac{\mathcal{D}^+ \mathcal{D}^-}{\mu n} \right) \right]}{n^{-\gamma+\delta}} = 0,$$

so that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \left[\sum_{i,j=1}^n \psi \left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\mu n} \right) \right]}{n^{2-\gamma+\delta}} = 0. \quad (8.31)$$

and hence (8.29) is $O_{\mathbb{P}}(n^{2-\gamma_*})$. Observe that the arguments for proving this scaling are exactly the same as those used in the proof of Proposition 8.10. The improvement is therefore obtained by a more careful analysis of the error term (8.30).

To explain where the improvement comes from, recall that, we used that

$$|\psi(x) - \psi(y)| = |x - y + e^{-x} - e^{-y}| \leq 2|x - y|,$$

for any $x, y \geq 0$, to obtain an upper bound for (8.30) in Proposition 8.10. However, since $\psi(x)$ is monotonically increasing we have that, for all $x, y \geq 0$ and $z \geq (x \vee y)$,

$$|\psi(x) - \psi(y)| \leq \psi'(z)|x - y|. \quad (8.32)$$

In addition we can bound the derivative $\psi'(x)$ by

$$\psi'(x) = 1 - e^{-x} \leq (1 \wedge x). \quad (8.33)$$

It are these bounds that will give us the desired improvement.

Proof of Theorem 8.11. Observe that due to (8.31) and the fact that

$$L_n O_{\mathbb{P}}(n^{-1}) = O_{\mathbb{P}}(1),$$

it is enough to show that

$$\left| \sum_{i,j=1}^n \psi \left(\frac{D_i^+ D_j^-}{L_n} \right) - \psi \left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\mu n} \right) \right| = O_{\mathbb{P}}(n^{2-\gamma_*}).$$

To this end fix $\delta > 0$ and define

$$\mathcal{E}_n^* = \sum_{i,j=1}^n \left[\psi \left(\frac{D_i^+ D_j^-}{L_n} \right) - \psi \left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\mu n} \right) \right].$$

In addition let $\kappa = (\gamma_* - 1)/2\gamma_*$ and define the event

$$A_n = \{|L_n - \nu n| \leq n^{1-\kappa}\}.$$

Then, since

$$1 - \kappa = \frac{1}{\gamma_*} + \kappa \quad \text{and} \quad \kappa > 1,$$

for all $1 < \gamma_+, \gamma_- < 2$, it follows from Lemma 8.7 ii) that

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n^c) = \lim_{n \rightarrow \infty} \mathbb{P}(|L_n - \nu n| > n^{\frac{1}{\gamma_*} + \kappa}) = 0.$$

We will therefore show that

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\mathcal{E}_n^*| > n^{2-\gamma_*+\delta}, A_n) = 0,$$

from which the result of the theorem follows.

We first bound $\mathcal{E}_n^*$ on the event A_n as follows:

$$\begin{aligned} |\mathcal{E}_n^*| \mathbb{1}_{A_n} &\leq \sum_{i,j=1}^n \left| \psi \left(\frac{D_i^+ D_j^-}{L_n} \right) - \psi \left(\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n} \right) \right| \mathbb{1}_{A_n} \\ &\leq \frac{1}{L_n} \sum_{i,j=1}^n \psi' \left(\frac{D_i^+ D_j^-}{\nu n - n^{1-\kappa}} \right) |D_i^+ D_j^- - \mathcal{D}_i^+ \mathcal{D}_j^-| \mathbb{1}_{A_n} \\ &\leq \frac{1}{\nu n - n^{1-\kappa}} \sum_{i,j=1}^n \psi' \left(\frac{D_i^+ D_j^-}{\nu n - n^{1-\kappa}} \right) (\mathcal{D}_i^+ \xi_j^- + \mathcal{D}_j^- \xi_i^+) \\ &\leq \frac{2}{\nu n - n^{1-\kappa}} \sum_{i,j=1}^n \psi' \left(\frac{D_i^+ D_j^-}{\nu n - n^{1-\kappa}} \right) \mathcal{D}_i^+ \mathcal{D}_j^-. \end{aligned} \quad (8.34)$$

Note that here we used the bound (8.32) for the second line and the fact that $\psi'(x)$ is monotonically increasing and that, on the event A_n ,

$$\frac{\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n}, \frac{D_i^+ D_j^-}{L_n} \leq \frac{D_i^+ D_j^-}{\nu n - n^{1-\kappa}}.$$

For the next step we note that (8.34) is zero unless $\mathcal{D}_i^+ \mathcal{D}_j^- \geq 1$ and that in this case,

$$D_i^+ D_j^- = \mathcal{D}_i^+ \mathcal{D}_j^- + \mathcal{D}_i^+ \xi_j^- + \mathcal{D}_j^- \xi_i^+ \leq 3\mathcal{D}_i^+ \mathcal{D}_j^-.$$

Therefore we have, using (8.33), that

$$\begin{aligned} &\frac{2}{\nu n - n^{1-\kappa}} \sum_{i,j=1}^n \psi' \left(\frac{D_i^+ D_j^-}{\nu n - n^{1-\kappa}} \right) \mathcal{D}_i^+ \mathcal{D}_j^- \\ &\leq \frac{2}{\nu n - n^{1-\kappa}} \sum_{i,j=1}^n \psi' \left(\frac{3\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n - n^{1-\kappa}} \right) \mathcal{D}_i^+ \mathcal{D}_j^- \end{aligned}$$

$$\leq \frac{2}{\nu n - n^{1-\kappa}} \sum_{i,j=1}^n \left(\frac{3\mathcal{D}_i^+ \mathcal{D}_j^-}{\nu n - n^{1-\kappa}} \wedge 1 \right) \mathcal{D}_i^+ \mathcal{D}_j^-.$$

To finish the proof we use, again, the fact that $\mathcal{D}^+ \mathcal{D}^-$ is regularly varying with exponent γ_* , so that by Lemma 2.6,

$$\lim_{n \rightarrow \infty} n^{\gamma_* - 1 + \delta} \mathbb{E} \left[\mathcal{D}^+ \mathcal{D}^- \left(\frac{3\mathcal{D}^+ \mathcal{D}^-}{\nu n - n^{1-\kappa}} \wedge 1 \right) \right] = 0.$$

It now follows from Markov's inequality that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} \left(|\mathcal{E}_n^*| > \frac{n^{2-\gamma_*+\delta}}{2}, A_n \right) \\ & \leq \lim_{n \rightarrow \infty} \frac{\mathbb{E} \left[|\mathcal{E}_n^{(1)}| \mathbb{1}_{A_n} \right]}{n^{2-\gamma_*+\delta}} \\ & \leq \lim_{n \rightarrow \infty} \frac{2n^{\gamma_*-1-\delta}}{\nu - n^{-\kappa}} \mathbb{E} \left[3\mathcal{D}^+ \mathcal{D}^- \left(\frac{\mathcal{D}^+ \mathcal{D}^-}{\nu n - n^{1-\kappa}} \wedge 1 \right) \right] = 0. \end{aligned}$$

□

Given that the proofs for the directed case are more involved than those for the undirected case, since in the latter we can regard the degrees as i.i.d. samples, the above proof can be adjusted in a straightforward manner to obtain the following result for the undirected erased configuration model:

Theorem 8.13. *Let $G_n = \text{CM}(\mathcal{D})$, with $1 < \gamma < 2$. Then*

$$\mathbb{E}_n[E_n] = O_{\mathbb{P}}(n^{2-\gamma}),$$

and, consequently

$$\frac{\mathbb{E}[E_n]}{L_n} = \begin{cases} O_{\mathbb{P}}(n^{1-\gamma}) & \text{if } 1 < \gamma < 2 \text{ and} \\ O_{\mathbb{P}}(n^{-1}) & \text{if } \gamma > 2. \end{cases}$$

Chapter 9

Scaling of rank-correlations measures in the erased configuration model

Based on [85]:

van der Hoorn, P. and Litvak, N.

Phase transitions for scaling of structural correlations in directed networks.

Physical Review E, 2015, 92, 022803

1

9.1 Introduction

In this chapter we revisit the directed erased configuration model for generating simple graphs with neutral mixing of degrees. In particular, we will focus on the setting where the limiting out- and in-degree distributions, F^+ and F^- , are regularly varying, with exponents $1 < \gamma_+, \gamma_- < 2$. We will use extensive numerical experiments to analyze the behavior of the three rank-correlations measures that we introduced in Chapter 4 on erased configuration model graphs with these scale-free degree distributions.

Although we have already seen (see Theorem 6.6) that the rank-correlation measures ρ_α^β , $\bar{\rho}_\alpha^\beta$ and τ_α^β converge in probability to zero as the size of the graph tends to infinity, we were unable to obtain an upper bound on the scaling, which we did achieve for the repeated configuration model in Theorem 6.5. Numerical experiments did show that the absolute value of the rank-correlation measures, in the erased configuration model, converges to zero reasonably quick in practice. However we did not investigate, for instance, how the distribution of these measures around zero behaves.

For the undirected configuration model where the degrees follow a regularly-varying distribution with exponent $1 < \gamma < 2$, it was observed [26, 55, 66] that the value of Pearson's correlation coefficient was always negative. In [55] this phenomenon is explained by observing that if one allows at most one edge between two nodes, then nodes with large degree must connect to nodes of small degree because there are simply not enough distinct large nodes to connect to. A similar explanation is given in [26]. Here, this is then related to the difference in scaling between the *natural* and *structural cut-off* of the network. The former is defined, see e.g. [39], as the degree value k_c , of which, on average, only one instance is observed. The structural cut-off is defined as the value k_s for which the ratio between the average number of edges that connect any two nodes of degree k_s , and the maximum possible number of such edges in a simple graph, is 1. The natural cut-off scales as $n^{1/\gamma}$, see [39], while the structural cut-off on networks with neutral mixing scales as $n^{1/2}$ [17]. Therefore, when $\gamma < 2$ the natural cut-off scales at a slower rate which gives rise to negative correlations.

We will show that structural negative correlations are also present for rank-correlation measures, in directed graphs generated by the erased configuration model. Interestingly, this only holds for the Out-In degree-degree correlation type. In addition, we will use numerical experiments to investigate the size dependence of the fluctuations of the value of our measures, for all four degree-degree correlation types, around their mean. We obtain a scaling for ρ_α^β , $\bar{\rho}_\alpha^\beta$ and τ_α^β in the erased configuration model, in terms of the exponents γ_+ and γ_- of the degree distributions, and show that this scaling undergoes a phase transition. Here, the different scaling terms are given by the loose upper bounds that we derived for the average number of erased edges in Chapter 8, see Theorem 8.6.

This chapter has the following structure. In Section 9.2 we will show the structural negative Out-In degree-degree correlations for the directed erased configuration model and give some insights into their origins. Then we analyze the fluctuations of the Out-In degree-degree correlation type around its mean in Section 9.3, where we show a phase-transition in the scaling behavior. The scaling for the other correlation types are discussed in Section 9.4.

9.2 Structural negative Out-In correlations in the erased configuration model

Let $\mathcal{D}^+$ and $\mathcal{D}^-$ be two independent, integer-valued regularly-varying random variables with exponents $\gamma_+ > 1$ and $\gamma_- > 1$, respectively. Then $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy (3.3) for any $\eta < (\gamma_+ \wedge \gamma_-) - 1$ and hence, if we consider graphs $G_n = \text{ECM}(\mathcal{D}^+, \mathcal{D}^-)$, it follows by Theorem 6.6 that, as $n \rightarrow \infty$,

$$\rho_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0, \quad \bar{\rho}_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \tau_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0,$$

for any $\alpha, \beta \in \{+, -\}$.

First, let us take $\gamma_+, \gamma_- > 2$, so that $\mathcal{D}^+$ and $\mathcal{D}^-$ have finite variance. Similar to the numerical experiments in Chapter 6 we have generated 10^3 graphs of different sizes with the erased configuration model, and computed all four versions of ρ_α^β on each of them. Then we computed the empirical cumulative distribution of Spearman's rho. The results are displayed in Figure 9.1. Observe that the plots for all four different degree-degree correlation types look very similar. Moreover we see that the distributions of $\rho_\alpha^\beta(G_n)$ are symmetric and centered around zero for all sizes.

Now we move to the infinite variance regime, $\gamma_+, \gamma_- < 2$. Here we have conducted the same type of experiments and have plotted the cumulative distributions of ρ_α^β in Figure 9.2. Apart from the In-Out degree-degree correlation type, the difference with Figure 9.1 is striking. Both ρ_+^+ and ρ_-^- are still centered around zero, but the variation is much larger than in the setting where $\gamma_\pm > 2$. However, the most difference, with respect to the finite-variance setting, is in the behavior of ρ_+^- . From Figure 9.2a we deduce that Spearman's rho for the Out-In degree-degree correlation type is always negative. To further illustrate this we have included the empirical averages $\mathbb{E}'[\rho_\alpha^\beta(G_n)]$ for all four degree-degree correlation types and sizes in Table 9.1. Observe that although $\rho_+^-(G_n) < 0$, it does decrease in absolute value as n increases, which reflects the results from Chapter 6.

To ensure that this behavior is not restricted to ρ_α^β we have repeated the same experiments for the other two rank-correlation measures, $\bar{\rho}_\alpha^\beta$ and τ_α^β . Again, in the finite variance regime all four degree-degree correlation types behave the same way and are centered around zero, while in the infinite variance regime the Out-In degree-degree correlation is always negative. We included Figure 9.3 to show these structural negative correlations.

The existence of these negative Out-In correlations (ρ_+^-) can be explained heuristically by first observing that, in the configuration model, multiple edges are more likely to start in a node of large out-degree and end in a node of large in-degree, since these are more likely to be sampled. Now, consider the erased configuration model where we first connect all stubs at random and then remove self-loops and merge multiple edges. By construction, immediately after the pairing the graph will have neutral mixing. However, when merging multiple edges we will often delete connections from nodes of large out-degree to nodes of large in-degree. Such edges have contributed positively to ρ_+^- , thus, deleting them will shift ρ_+^- from zero in the general configuration model, to a negative value in the erased configuration model.

Figure 9.1: Plots of the empirical cumulative distribution of ρ_α^β for all four degree-degree correlation types for graphs of different sizes, generated by the erased configuration model with $\gamma_\pm = 2.1$.

n	$\mathbb{E}'[\rho_+^-]$	$\mathbb{E}'[\rho_-^+]$	$\mathbb{E}'[\rho_+^+]$	$\mathbb{E}'[\rho_-^-]$
10000	-0.1568	-0.0001	0.0039	0.0048
50000	-0.1439	0.0001	0.0014	0.0029
100000	-0.1388	-0.0001	0.0026	0.0028
500000	-0.1198	0.0001	0.0011	0.0017
1000000	-0.1131	0.0000	0.0009	0.0002

Table 9.1: The average values for ρ_α^β for all four degree-degree correlation types, for graphs of different sizes, generated by the erased configuration model with $\gamma_\pm = 1.2$, based on 10^3 realizations of the model.

Figure 9.2: Plots of the empirical cumulative distribution of ρ_{α}^{β} for all four degree-degree correlation types for graphs of different sizes, generated by the erased configuration model with $\gamma_{\pm} = 1.2$.

Figure 9.3: Plots of the empirical cumulative distribution of ρ_+^- , $\bar{\rho}_+^-$ and τ_+^- for graphs of different sizes, generated by the erased configuration with $\gamma_{\pm} = 1.2$.

The In-Out degree-degree correlation is not affected since we took $\mathcal{D}^+$ and $\mathcal{D}^-$ to be independent, so that D_i^+ and D_i^- can also be considered as independent. Therefore, if $i \rightarrow j$ is an edge from a node i with large out-degree to a node j with large in-degree, then D_i^- and D_j^+ can be any value. Hence, removing edges does not create a bias that contributes positively or negatively towards ρ_+^+ .

Finally, for the two symmetric degree-degree correlation types, Out-Out and In-In, we note that, although there is a bias on one end of the edges, the value of the degree of the other end is independent so that, on average, ρ_+^+ and ρ_-^+ are zero. The observed increase in variance could then be due to the degree-bias on one side of the edge.

9.3 Empirical analysis of the scaling of ρ_+^- in the erased configuration model

We will analyze the scaling, in n , of the fluctuation of $\rho_+^-(G_n)$ around its average, as a function of the exponents $\gamma_\pm$. In particular we want to find coefficients $\varphi(\gamma_+, \gamma_-)$, possibly dependent on γ_+ and γ_- , such that, as $n \rightarrow \infty$,

$$\frac{\rho_+^-(G_n) - \mathbb{E}[\rho_+^-(G_n)]}{\mathcal{L}(n)n^{\varphi(\gamma_+, \gamma_-)}} \xrightarrow{d} \mathcal{R}_{\gamma_+, \gamma_-},$$

for some slowly-varying function $\mathcal{L}$ and random variable $\mathcal{R}_{\gamma_+, \gamma_-}$, which might also depend on the exponents γ_+ and γ_- . We emphasize that it is not our intention to determine the distribution of $\mathcal{R}_{\gamma_+, \gamma_-}$ here. We are only interested in the scaling coefficients $\varphi(\gamma_+, \gamma_-)$, since these will give us an indication of the variation of the rank-correlation measures in terms of the graph size.

9.3.1 Methodology of the analysis

Let $\{X_i\}_{i \geq 1}$ and X be random variables such that there exist sequences a_n and b_n , for which

$$\frac{X_n - b_n}{a_n} \xrightarrow{d} X, \quad \text{as } n \rightarrow \infty.$$

Then if we take, for different values of n , multiple samples from the random variable $Y_n := (X_n - b_n)/a_n$ and plot the empirical distribution, these should converge to the distribution of X . On the data, the plots for the empirical distribution of Y_n , for different n , will strongly overlap. Therefore, one way to empirically evaluate the sequences a_n and b_n is to plot, for each combination of candidate sequences a'_n and b'_n , the empirical distribution of $Y'_n := (X_n - b'_n)/a'_n$, for several values of n and see if these overlap. Another important property to look for is the interval to which the distribution of Y'_n is restricted. If the scaling sequence a'_n is too large then Y'_n will be very small in magnitude, concentrated around zero, while if we pick a scaling which is too small then we will "blow up" the values of Y'_n and hence, its distribution will span a large interval.

We will use this approach to determine the scaling coefficients $\varphi(\gamma_+, \gamma_-)$ of $\rho_\alpha^\beta(G_n) - \mathbb{E}[\rho_\alpha^\beta(G_n)]$. We remark that we do not consider the slowly-varying

function $\mathcal{L}$, since these are, in general, only defined implicitly. Moreover, by Potter's Theorem,

$$\lim_{n \rightarrow \infty} n^{-\delta} \mathcal{L}(n) = 0,$$

for all $\delta > 0$ so that, for large enough n , the term $n^{\varphi(\gamma_+, \gamma_-)}$ should be dominating. Therefore, we will in general be interested in the cumulative distribution of

$$Y_n^{\alpha, \beta}(\gamma_+, \gamma_-) := \frac{\rho_\alpha^\beta(G_n) - \mathbb{E}[\rho_\alpha^\beta(G_n)]}{n^{\varphi(\gamma_+, \gamma_-)}}. \quad (9.1)$$

In order to sample from $Y_n^{\alpha, \beta}(\gamma_+, \gamma_-)$ we follow the same procedure as described in the previous section, that is, we generate 10^3 graphs with the erased configuration model $\text{ECM}(\mathcal{D}^+, \mathcal{D}^-)$ for different values of n and exponents γ_+ and γ_- and compute $\rho_\alpha^\beta(G_n)$. Then, for each n , γ_+ and γ_- we calculate the empirical average $\mathbb{E}'[\rho_\alpha^\beta(G_n)]$, based on the 10^3 samples and plot the empirical cumulative distribution of

$$\frac{\rho_\alpha^\beta(G_n) - \mathbb{E}'[\rho_\alpha^\beta(G_n)]}{n^{\varphi(\gamma_+, \gamma_-)}},$$

for different choices of $\varphi(\gamma_+, \gamma_-)$. We will slightly abuse notation and use $Y_n^{\alpha, \beta}(\gamma_+, \gamma_-)$ to refer to either the above expression or to (9.1).

9.3.2 Scaling terms for Spearman's rho

Choosing the appropriate scaling coefficients $\varphi(\gamma_+, \gamma_-)$ in such empirical studies often comes down to an educated guess, lead by some insights concerning the process. We will give a short explanation for our choice of scaling coefficients.

Let $(\mathbf{X}_n, \mathbf{Y}_n)$ be a sequence of joint independent samples from integer valued random variables (X, Y) . Then it follows from [46] that Spearman's rho and Kendall's tau are normal statistics, which means that

$$\frac{\rho(\mathbf{X}_n, \mathbf{Y}_n) - \mathbb{E}[\rho(\mathbf{X}_n, \mathbf{Y}_n)]}{n^{-\frac{1}{2}}} \xrightarrow{d} \mathcal{N}(0, \xi^2),$$

where the variance ξ^2 is determined by the specific form of the distributions of X and Y .

In Chapter 4 we have introduced $\rho_+^-(G_n)$ as Spearman's rho on the joint vector $(D_i^+, D_i^-)_{i \rightarrow j}$. We have seen, for instance in Chapter 8, that although the out- and in-degrees in a bi-degree sequence $\text{IID}(\mathcal{D}^+, \mathcal{D}^-)$ are not i.i.d. samples from, respectively, $\mathcal{D}^+$ and $\mathcal{D}^-$, for many scaling results they do behave as such. Therefore we could expect that $\rho_+^-(G_n)$ in the general configuration model, where we do not change the bi-degree sequence, is also a normal statistic. Hence $-1/2$ is in the short-list of scaling coefficients $\varphi(\gamma_+, \gamma_-)$.

For other candidates we recall the intuition that the negative Out-In degree-degree correlations are a consequence of the removing of edges in the erased configuration model. Therefore, since $\rho_+^-(G_n)$ can also be seen as the average of the product of the ranks of the degrees D_i^+ and D_j^- over all edges $i \rightarrow j \in G_n$, we might expect the average number of erased edges to be involved.

Now let G_n be a graph generated by the erased configuration model and denote by $\tilde{G}_n$ the graph before the removal of edges. We then expect that

$$\rho_+^-(G_n) = \rho_E(G_n) + \rho_+^-(\tilde{G}_n),$$

Figure 9.4: Plot of the different scaling regimes for ρ_+^- . The Roman numerals indicate the three different choices of γ_+ and γ_- , used in Figure 9.5 and 9.6, to illustrated the different regimes.

where $\rho_E(G_n)$ is a term which is determined by the average number of erased edges. Consequently,

$$\begin{aligned} \rho_+^-(G_n) - \mathbb{E}[\rho_+^-(G_n)] \\ = (\rho_E(G_n) - \mathbb{E}[\rho_E(G_n)]) + \left(\rho_+^-(\tilde{G}_n) - \mathbb{E}[\rho_+^-(\tilde{G}_n)] \right). \end{aligned} \quad (9.2)$$

Since for the last term we expect, based on the results from [46], a scaling $n^{-1/2}$, the scaling of $\rho_+^-(G_n) - \mathbb{E}[\rho_+^-(G_n)]$ will be determined by the largest of the scaling for $\rho_E(G_n) - \mathbb{E}[\rho_E(G_n)]$ and $n^{-1/2}$.

For the scaling of $\rho_E(G_n) - \mathbb{E}[\rho_E(G_n)]$ we turn to the results from Chapter 8, see for instance (8.28), and consider the scaling $1 - \gamma_*$, where $\gamma_* = (\gamma_+ \wedge \gamma_-)$. Extensive numerical experiments using different combinations for the exponents $\gamma_{\pm}$ did not give good results, which we will therefore not present here. Instead we will consider the scaling coefficients presented in Theorem 8.6 for $1 < \gamma_{\pm} < 2$.

It is worth noting that, chronologically, the research done in this chapter preceded the results from Chapter 8. In particular, the upper bounds n^{1/γ_*-1} and $n^{2/\gamma_++2/\gamma_- - 3}$ presented in Theorem 8.6 were inspired by the numerical experiments presented here.

The bounds from Theorem 8.6 give rise to a phase-transition in the shape of the scaling coefficients $\varphi(\gamma_+, \gamma_-)$,

$$\varphi(\gamma_+, \gamma_-) = \begin{cases} \frac{1}{\gamma_*} - 1 & \text{if } \gamma_* \leq \frac{\gamma^*}{2(\gamma^*-1)} \\ \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3 & \text{if } \frac{\gamma^*}{2(\gamma^*-1)} < \gamma_*. \end{cases}$$

The transition is determined by the value of γ_+, γ_- for which $\frac{1}{\gamma_*} - 1 \leq \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3$, since the smallest of the two expressions will be dominating. Therefore, based

on the expression (9.2) we expect a three-stage phase-transition,

$$\varphi(\gamma_+, \gamma_-) = \begin{cases} \frac{1}{\gamma_*} - 1 & \text{if } \gamma_* \leq \frac{\gamma^*}{2(\gamma^*-1)} \\ \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3 & \text{if } \frac{\gamma^*}{2(\gamma^*-1)} < \gamma_* \text{ and } \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3 \geq -\frac{1}{2} \\ -\frac{1}{2} & \text{if } \frac{2}{\gamma_+} + \frac{2}{\gamma_-} - 3 < -\frac{1}{2}. \end{cases} \quad (9.3)$$

In Figure 9.4 we have plotted, for $1 < \gamma_{\pm} < 2$, the regions in which the different scaling terms should hold. In addition we also display three choices of the exponents, numbered with a Roman numeral, that we will use to test the scaling in our experiments.

9.3.3 Numerical results

In order to show the phase transitions we have computed the empirical cumulative distribution function of $Y_+^-(\gamma_+, \gamma_-)$ for the specific choices of $\gamma_{\pm}$, corresponding to the points I, II and III in Figure 9.4. For each of the three points we have shifted the empirical data by its average and divided it by $n^{\varphi(\gamma_+, \gamma_-)}$, for any of the three coefficients from (9.3), corresponding to the different scaling areas A, B and C. The code for these experiments uses a parallel programming scheme to optimize the computation time and was run on a multi core server.

The results are shown in Figure 9.5. As explained in Section 9.3.1, when the correct scaling is applied, the corresponding plots should almost completely overlap and resemble some limiting distribution. We observe that for each of the three choices I, II and III, this is the case when the corresponding scaling from its area, respectively A, B and C, is chosen. This is also supported by the range of the cumulative distributions. For instance when we take $\gamma_+ = 1.3 = \gamma_-$ (region I in Figure 9.4), we see that when we apply the scaling $n^{-1/2}$, which is much smaller than n^{1/γ_*-1} , the values are blown up and lie roughly in between -100 and 100 . On the other hand, the scaling $n^{2/\gamma_++2/\gamma_- - 3}$, which is too large, reduces all values to the range $[-0.02, 0.01]$.

9.4 Scaling of the In-Out, Out-Out and In-In degree-degree correlation types

In the previous section we have completely characterized the scaling behavior of the fluctuation of ρ_+^- for the erased configuration model with infinite variance of the degrees. Here, we first discuss the remaining degree-degree correlation types, ρ_+^+ , ρ_-^- and ρ_-^+ in the infinite-variance regime and then we consider all four types in the finite-variance regime.

In contrast to the Out-In degree-degree correlation type, the other three degree-degree correlation types do not show negative correlations, see Figure 9.2b-9.2d, which we argued was due to the fact that we took the out- and in-degree of a node to be independent. Nevertheless, the spread of both the Out-Out and In-In degree-degree dependency exhibits scaling with the same function as in (9.3). This is illustrated in Figure 9.6, where we have plotted the empirical cumulative distribution of $Y_+^+(\gamma_+, \gamma_-)$ for graphs $\text{ECM}(\mathcal{D}^+, \mathcal{D}^-)$, with values of the exponents $\gamma_{\pm}$ corresponding to points I, II and III from Figure 9.4 and $\varphi(\gamma_+, \gamma_-)$

Figure 9.5: Plots of the empirical cumulative distribution function of $Y_+^-(\gamma_+, \gamma_-)$ using different scaling and for different choices of $\gamma_{\pm}$. The left column is scaled by n^{1/γ_*-1} , the center column by $n^{2/\gamma_++2/\gamma_- - 3}$ and the right column by $n^{-1/2}$. The first row is for graphs $\text{ECM}(\mathcal{D}^+, \mathcal{D}^-)$ with $\gamma_{\pm} = 1.3$, the second for $\gamma_+ = 1.9$ and $\gamma_- = 1.3$, and the third for $\gamma_+ = 1.9$ and $\gamma_- = 1.5$. These correspond to, respectively, the points I, II and III in Figure 9.4.

Figure 9.6: Plots of the empirical cumulative distribution function of $Y_+^+(\gamma_+, \gamma_-)$ for choices of $\gamma_{\pm}$ corresponding to points I, II and III from Figure 9.4, using the three different scaling coefficients.

Figure 9.7: Plots of the empirical cumulative distribution function of $Y_-^+(\gamma_+, \gamma_-)$ for choices of $\gamma_{\pm}$ corresponding to points I, II and III from Figure 9.4, and $\varphi(\gamma_+, \gamma_-) = -1/2$.

Figure 9.8: Plots of the empirical cumulative distribution of $Y_{\alpha}^{\beta}(\gamma_+, \gamma_-)$ for all four degree-degree dependency types for graphs of different sizes, generated by the erased configuration model with $\gamma_{\pm} = 2.1$ and $\varphi(\gamma_+, \gamma_-) = -1/2$.

Figure 9.9: Plot of the empirical cumulative distribution function of $(\rho_+^-(G_n) - \mathbb{E}[\rho_+^-(G_n)]) / n^{-1/2}$ for graphs size 10^6 generated by the erased configuration model, with $\gamma_{\pm} = 2.1$ and a normal cumulative distribution with $\mu = 0$ and $\sigma^2 = 0.8$.

as in (9.3). An explanation of this behavior is that $\rho_+^+(G_n)$ again depends on the number of erased edges, through the out-degree of their source nodes. By symmetry, the scaling for the In-In dependency is similar and we do not show the plot here.

Recall that we expect Spearman's rho to have scaling $n^{-1/2}$ when the values where (approximately) i.i.d. samples. This is exactly what we observe for the In-Out degree-degree correlation type, which, in contrast to the other three, is not biased towards removed edges. As we expect, here we have normal, square root, scaling for any choice of $\gamma_{\pm}$. This can clearly be observed in Figure 9.7, where we plotted the empirical cumulative distributions of $Y_-^+(\gamma_+, \gamma_-)$ with $\varphi(\gamma_+, \gamma_-) = n^{-1/2}$.

For the degree-degree correlation types in the finite variance regime we plotted, for all $\alpha, \beta \in \{+, -\}$, the empirical cumulative distributions of $Y_{\alpha}^{\beta}(\gamma_+, \gamma_-)$ again with $\varphi(\gamma_+, \gamma_-) = n^{-1/2}$, in Figure 9.8. Since these are all completely similar, we took the plot for ρ_+^- for graphs of size 10^6 and compared it to a fitted normal distribution with $\mu = 0$ and $\sigma^2 = 0.8$, see Figure 9.9. These plots strongly overlap, enforcing the claim that for the erased configuration model with finite degree variance all three measures ρ_{α}^{β} , $\bar{\rho}_{\alpha}^{\beta}$ and τ_{α}^{β} are normal statistics for all four degree-degree correlation types. The exact characterization of the scaling of rank-correlations in the infinite variance regime remains an open problem.

Chapter 10

Distances in the directed configuration model

Based on [87]:

van der Hoorn, P and Olvera-Cravioto, M

Distances in the directed configuration model

arXiv preprint arXiv:1511.04553

1

10.1 Introduction

In this chapter, we provide an analysis of the distance between two randomly chosen nodes in the directed configuration model under the assumption that the covariance between out- and in-degree is finite. We show that the hopcount, i.e., the length of the shortest directed path between two nodes, grows logarithmically with the graph size and characterize the fluctuations around this term.

The directed nature of the graphs introduces some subtle differences compared to the undirected case. An important difference between the undirected and directed cases is that the dependence between the out- and in-degree in the latter plays an important role in the behavior of the distance between nodes. More precisely, unlike the undirected case [80], the distances in the directed model can scale logarithmically even when the variance of the degrees is infinite. Intuitively, the length of the shortest directed path between any two nodes will always be larger than the shortest undirected path. However, what is surprising, is that this distance does not necessarily get shorter as the variability of the degrees grows larger, and whether it gets shorter or not depends on the level of dependence between the in- and out-degree. Together with prior results on the existence and the size of a strongly connected component in random directed graphs [31, 68], our results provide valuable insights into the differences and similarities between the directed and undirected models.

We use numerical examples on large configuration model graphs to show that our characterization of the hopcount gives a very good approximation of distances in finite size graphs. Here we also show the influence of the covariance between the out- and in-degrees on the hopcount, by considering different way of sampling the bi-degree sequences. These include, for example, the d -regular bi-degree sequence and a bi-degree sequence where the out- and in-degree of a node are the same.

In order to characterize the hopcount we introduce a novel coupling between a breadth-first graph exploration process and a Galton-Watson tree. This coupling is based on the Kantorovich-Rubinstein distance between two probability measures (see, e.g., [91]), and has the advantage of being uniformly accurate for a considerably longer time than existing constructions. Specifically, the coupling holds for a number of steps in the graph exploration process equivalent to discovering $n^{1-\epsilon}$ nodes, for arbitrarily small $\epsilon > 0$, compared to a constant number of nodes in [68], $n^{1/2-\epsilon}$ nodes in [65] and [40, Theorem 2.2.2], or $n^{1/2+\delta}$ nodes, for a very small $\delta > 0$, in [77, 80]. Moreover, the coupled branching process has a deterministic offspring distribution that does not depend on n or the degree sequences, avoiding the need to consider intermediate tree constructions. The generality of our main coupling result, and the wide range of applications where a so-called branching process argument is used, makes it of independent interest.

This chapter is organized as follows. We start by describing the coupling and proving our main coupling theorem in Section 10.2. Then we use this coupling in Section 10.3 to analyze the hopcount in the directed configuration model. Finally, Section 10.4 contains numerical examples for the hopcount in large configuration model graphs, with different types of bi-degree sequences.

10.2 Coupling the exploration of a graph with a branching process

In this section we introduce a direct coupling between a breadth-first exploration of graphs, generated by the directed configuration model, and a delayed branching process. This is a well-known approach for analyzing the properties of random graphs, also referred to as a branching process argument.

The main result of this section is Theorem 10.1, along with its more immediately useful corollary (Corollary 10.2), which is the key ingredient in the proof of the main result for the distances in the directed configuration model, Theorem 10.9.

We will first introduce a specific set of regularity assumptions needed for the proofs in this section and the remainder of this chapter.

10.2.1 Regularity assumptions on the bi-degree sequence

The results in this chapter will hold for all bi-degree sequences that satisfy the regularity condition that is similar to Assumption 3.3. In this specific case however we do not require the convergence of $f_n^{+,+}$ and $f_n^{-,-}$ and we only need that the sum

$$\sum_{i=1}^n ((D_i^+)^{\eta} + (D_i^-)^{\eta}) D_i^+ D_i^-,$$

scales as n , which is less than the convergence of the empirical joint $1+\eta$ moment of the out- and in-degrees.

Assumption 10.1. There exist probability mass functions f^+ , f^- , $f^{+,-}$ and $f^{-,+}$ on the non-negative integers that satisfy

$$\begin{aligned} \nu &:= \sum_{k=0}^{\infty} k f^+(k) = \sum_{k=0}^{\infty} k f^-(k) < \infty \quad \text{and} \\ \mu &:= \sum_{k=0}^{\infty} k f^{+,-}(k) = \sum_{k=0}^{\infty} k f^{-,+}(k) < \infty, \end{aligned}$$

and there exists $\varepsilon > 0$, $0 < \eta \leq 1$ and constant $K_{\eta} < \infty$, such that, if

$$\begin{aligned} \Xi_n &= \left\{ \max \left\{ d_1(f_n^+, f^+), d_1(f_n^-, f^-), d_1(f_n^{+,-}, f^{+,-}), d_1(f_n^{-,+}, f^{-,+}) \right\} \leq n^{-\varepsilon} \right\} \\ &\cap \left\{ \sum_{i=1}^n ((D_i^+)^{\eta} + (D_i^-)^{\eta}) D_i^+ D_i^- \leq K_{\eta} n \right\}, \end{aligned} \tag{10.1}$$

then

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Xi_n) = 1.$$

Note that

$$\sum_{i=1}^n ((D_i^+)^{\eta} + (D_i^-)^{\eta}) D_i^+ D_i^- \leq K_{\eta} n$$

implies that

$$\sum_{k=0}^{\infty} k^{1+\eta} f^{+,-}(k) \leq \liminf_{n \rightarrow \infty} \sum_{k=0}^{\infty} k^{1+\eta} f_n^{+,-}(k) \leq K_{\eta}/\nu < \infty,$$

and similarly for $f^{-,+}$, i.e., $f^{+,-}$ and $f^{-,+}$ have finite moments of order $1 + \eta$. Moreover, recall that

$$\nu_n = \frac{L_n}{n} \quad \text{and} \quad \mu_n = \frac{1}{L_n} \sum_{i=1}^n D_i^+ D_i^-,$$

are the common mean of f^+ , f^- , and $f^{+,-}$, $f^{-,+}$, respectively. Hence, it follows that on the event Ξ_n

$$|\nu - \nu_n| \leq n^{-\varepsilon} \quad \text{and} \quad |\mu - \mu_n| \leq n^{-\varepsilon}. \quad (10.2)$$

Remark 10.1. Throughout the remainder of this chapter, $\varepsilon > 0$ and $0 < \eta \leq 1$ are those from Assumption 10.1.

10.2.2 Exploration of new nodes in a graph

Let G_n be a directed graph. We will describe two breadth-first exploration processes, $\{Z_k^+ : k \geq 0\}$ and $\{Z_k^- : k \geq 0\}$, that start in an arbitrary node and explore the graph by going over, respectively, the outbound and inbound edges. In addition, we will keep track of the newly discovered outbound(inbound) stubs using the processes $\{A_k^{\pm} : k \geq 0\}$.

Similarly to the construction in [80], we start by labeling all the n nodes with a 1, meaning they have not been explored yet, and set $Z_0^{\pm} = 1$ (note that in [80] it is the stubs themselves that are labeled, not the nodes). Let $\emptyset$ denote the fictional first stub, and set $A_0^{\pm} = \{\emptyset\}$; call this initialization step 0. The process $\{Z_k^{\pm} : k \geq 0\}$ will keep track of the number of outbound (inbound) stubs discovered during the k -th step of the exploration process, as we will now describe. The super-index $\pm$ refers to whether the exploration follows the outbound stubs (for which we use the super-index $+$), or the inbound stubs (for which we use the super-index $-$).

In step 1 we randomly select a node and set $Z_1^{\pm} = j$ if it has j outbound (inbound) stubs; we label the node itself with a 2, meaning it has already been seen. To identify each of the outbound (inbound) stubs we index them 1 through j and let $A_1^{\pm} = \{1, \dots, j\}$ be the set of the indices of the newly discovered stubs. For the second step of the exploration process we will need to traverse all $Z_1^{\pm}$ outbound (inbound) stubs, which we do sequentially and in lexicographic order with respect to their indices. Here, we say that we have traversed an outbound (inbound) stub if we have identified the node it leads to and discovered how many outbound (inbound) stubs this new node has. If the stub is pointing to a node with label 1, then relabel it with a 2, index all its outbound (inbound) stubs with a name of the form (i, j) , $j \geq 1$, and finally proceed to explore the next outbound (inbound) stub. If the stub is pointing to a node with label 2 no new outbound (inbound) stubs are discovered. Once we are done exploring all

$Z_1^\pm$ outbound (inbound) stubs we set $Z_2^\pm$ to be the number of newly discovered outbound (inbound) stubs and let $A_2^\pm$ denote the set of their indices.

In general, in step k we will traverse all $Z_{k-1}^\pm$ outbound (inbound) stubs, in lexicographic order, discovering new nodes and hence new outbound (inbound) stubs. If outbound (inbound) stub $\mathbf{i} = (i_1, \dots, i_{k-1})$ is paired with an inbound (outbound) stub belonging to a node with label 1, then the outbound (inbound) stubs of the newly discovered node receive an index of the form $(i_1, \dots, i_{k-1}, i_k)$, $i_k \geq 1$; if outbound (inbound) stub $\mathbf{i}$ is paired with an inbound (outbound) stub belonging to a node having label 2, then no new outbound (inbound) stubs are discovered. Once we have traversed all $Z_{k-1}^\pm$ outbound (inbound) stubs we set $Z_k^\pm$ to be the number of new outbound (inbound) stubs discovered in step k . The process continues until all L_n outbound (inbound) stubs have been traversed.

Note that the process $\{Z_k^\pm : k \geq 0\}$ defines a labeled tree, where the "individuals" are the outbound (inbound) stubs discovered in step k ($Z_0^\pm = 1$), and not the nodes of the graph themselves. In addition to keeping track of $Z_k^\pm$, we will also keep track of "time" in the exploration process, where time T means we have traversed T outbound (inbound) stubs.

10.2.3 Construction of the coupling

To study the distance between two randomly chosen nodes in a graph generated by the directed erased configuration model, we will couple the exploration of the graph described above with a branching process. To do this we first note that the exploration process is equivalent to assigning to outbound stub $\mathbf{i} \neq \emptyset$ a number of offspring $\chi_{\mathbf{i}}^+$ chosen according to the (random) probability mass function

$$h_{\mathbf{i}}^+(t) = \begin{cases} \frac{1}{L_n - T_{\mathbf{i}}^+} \sum_{r=1}^n \mathbb{1}_{\{D_r^+ = t\}} D_r^- \mathcal{I}_r(T_{\mathbf{i}}^+), & t = 1, 2, \dots, \\ \frac{1}{L_n - T_{\mathbf{i}}^+} \left\{ \sum_{r=1}^n \mathbb{1}_{\{D_r^+ = 0\}} D_r^- \mathcal{I}_r(T_{\mathbf{i}}^+) + V_{\mathbf{i}}^- \right\}, & t = 0, \end{cases} \quad (10.3)$$

where $T_{\mathbf{i}}^+$ is the number of outbound stubs that have been traversed up until the moment outbound stub $\mathbf{i}$ is about to be explored,

$$\mathcal{I}_r(T) = \mathbb{1}_{\{\text{node } r \text{ has label 1 after having explored } t \text{ stubs}\}} \quad \text{and} \quad (10.4)$$

$$V_{\mathbf{i}}^- = L_n - \sum_{r=1}^n N_r \mathcal{I}_r(T_{\mathbf{i}}^+) - T_{\mathbf{i}}^+ \quad (10.5)$$

is the number of unexplored inbound stubs belonging to nodes with label 2 at time $T_{\mathbf{i}}^+$. Note that $T_{\mathbf{i}}^+$ is also the number of inbound stubs that already belong to edges in the graph up until the moment outbound stub $\mathbf{i}$ is about to be explored. Symmetrically, we assign to inbound stub $\mathbf{i}$ a number of offspring $\chi_{\mathbf{i}}^-$ distributed according to

$$h_{\mathbf{i}}^-(t) = \begin{cases} \frac{1}{L_n - T_{\mathbf{i}}^-} \sum_{r=1}^n \mathbb{1}_{\{D_r^- = t\}} D_r^+ \mathcal{I}_r(T_{\mathbf{i}}^-), & t = 1, 2, \dots, \\ \frac{1}{L_n - T_{\mathbf{i}}^-} \left\{ \sum_{r=1}^n \mathbb{1}_{\{D_r^- = 0\}} D_r^+ \mathcal{I}_r(T_{\mathbf{i}}^-) + V_{\mathbf{i}}^+ \right\}, & t = 0, \end{cases} \quad (10.6)$$

with $T_{\mathbf{i}}^-$ the number of inbound stubs that have been traversed up until the moment inbound stub $\mathbf{i}$ is about to be explored, and

$$V_{\mathbf{i}}^+ = L_n - \sum_{r=1}^n D_r \mathcal{I}_r(T_{\mathbf{i}}^-) - T_{\mathbf{i}}^- \quad (10.7)$$

is the number of unexplored outbound stubs belonging to nodes with label 2 at time $T_{\mathbf{i}}^-$. As before, we have that $T_{\mathbf{i}}^-$ is also the number of outbound stubs that already belong to edges in the graph up until the moment inbound stub $\mathbf{i}$ is about to be explored.

Note that the number of outbound (inbound) stubs of the first node, i.e., $Z_1^\pm$, is distributed according to $f_n^\pm$.

Remark 10.2. Note that we have used the notation $h^{\alpha,\beta}$ before, to indicate the limiting joint-degree distribution, while here we use h for the probability densities $h_{\mathbf{i}}^\pm$. However, since this chapter does not deal with degree-degree correlations no confusion should arise.

The key idea behind the coupling is that sampling from $h_{\mathbf{i}}^\pm$ and sampling from $f_n^{\mp,\pm}$ should be roughly equivalent as long as $T_{\mathbf{i}}^\pm$ is not too large. In turn, Assumption 10.1 implies that for large n , $f_n^{\mp,\pm}$ is very close to $f^{\mp,\pm}$. It follows that the process $\{Z_k^\pm : k \geq 0\}$ should be very close to a suitably constructed (delayed) branching process $\{\hat{Z}_k^\pm : k \geq 0\}$ having offspring distributions $(f^\pm, f^{\mp,\pm})$, where $f^\pm$ is the distribution of $\hat{Z}_1^\pm$ and all other nodes have offspring according to $f^{\mp,\pm}$.

To construct the coupling define $U = \bigcup_{k=0}^\infty \mathbb{N}_+^k$, with the convention that $\mathbb{N}_+^0 = \{\emptyset\}$, and let $\{U_{\mathbf{i}}\}_{\mathbf{i} \in U}$ be a sequence of i.i.d. uniform distributed random variables on $(0, 1)$. Next recall, for a non-decreasing function ϕ , the definition of pseudo-inverse $\phi^{-1}(t) = \inf\{x \in \mathbb{R} : \phi(x) \geq t\}$. Now set the number of outbound (inbound) stubs of $\mathbf{i}$ in the graph to be

$$\chi_{\mathbf{i}}^\pm = (H_{\mathbf{i}}^\pm)^{-1}(U_{\mathbf{i}}), \quad \mathbf{i} \neq \emptyset, \quad \chi_\emptyset^\pm = (F_n^\pm)^{-1}(U_\emptyset),$$

where $H_{\mathbf{i}}^\pm$ is the cumulative distribution function of $h_{\mathbf{i}}^\pm$, and the number of offspring of individual $\mathbf{i}$ in the outbound (inbound) branching process to be

$$\hat{\chi}_{\mathbf{i}}^\pm = (F^{\mp,\pm})^{-1}(U_{\mathbf{i}}), \quad \mathbf{i} \neq \emptyset, \quad \hat{\chi}_\emptyset^\pm = (F^\pm)^{-1}(U_\emptyset).$$

In addition we let $\hat{A}_r^\pm$ denote the set of individuals in the tree, corresponding to the process $\{\hat{Z}_k^\pm : k \geq 0\}$, at distance r from the root.

Note that $\chi_{\mathbf{i}}$ and $\hat{\chi}_{\mathbf{i}}$ are now coupled through the same $U_{\mathbf{i}}$, and this coupling minimizes the Kantorovich-Rubinstein distance between the distributions $h_{\mathbf{i}}^\pm$ and $f^{\mp,\pm}$. Moreover, although the $\chi_{\mathbf{i}}^\pm$ are only defined for stubs $\mathbf{i}$ that have been created through the pairing process, the $\hat{\chi}_{\mathbf{i}}^\pm$ are well defined regardless of whether $\mathbf{i}$ belongs to the tree or not. Furthermore, the sequence $\{U_{\mathbf{i}}\}_{\mathbf{i} \in U}$ defines the entire branching process $\{\hat{Z}_k^\pm : k \geq 0\}$, even after the graph has been fully explored.

The last thing we need to take care of is the observation that knowing $\chi_{\mathbf{i}}^\pm$ in the exploration of the graph does not necessarily tell us the identity of the node that stub $\mathbf{i}$ leads to, since there may be more than one node with $\chi_{\mathbf{i}}^\pm$ outbound (inbound) stubs, which is problematic if they do not also have the same

number of inbound (outbound) stubs. To fix this problem, given $\chi_i^\pm = t > 0$, we randomly select an inbound (outbound) stub from the set of unexplored inbound (outbound) stubs belonging to nodes with label 1 and having exactly t outbound (inbound) stubs; if $\chi_i^\pm = 0$ sample the inbound (outbound) stub from the set of unexplored inbound (outbound) stubs belonging to either nodes with label 2 or to nodes with label 1 having no outbound (inbound) stubs.

Summarizing the notation, we have:

- A_r^+ (A_r^-): set of outbound (inbound) stubs created during the r -th step of the exploration process on the graph.
- $\hat{A}_r^+$ ($\hat{A}_r^-$): set of individuals in the outbound (inbound) tree at distance r of the root.
- Z_r^+ (Z_r^-): number of outbound (inbound) stubs created during the r -th step of the exploration process.
- $\hat{Z}_r^+$ ($\hat{Z}_r^-$): number of individuals in the r -th generation of the outbound (inbound) tree.

The main observation upon which the analysis of the coupling is based is that if $|A|$ denotes the cardinality of set A , then

$$Z_k^\pm = |A_k^\pm| = |A_k^\pm \cap \hat{A}_k^\pm| + |A_k^\pm \cap (\hat{A}_k^\pm)^c|$$

which implies that

$$\hat{Z}_k^\pm - |\hat{A}_k^\pm \cap (A_k^\pm)^c| \leq Z_k^\pm \leq \hat{Z}_k^\pm + |A_k^\pm \cap (\hat{A}_k^\pm)^c|. \quad (10.8)$$

Note that as long as the exploration process and the branching process agree,

$$|\hat{A}_k^\pm \cap (A_k^\pm)^c| = 0 = |A_k^\pm \cap (\hat{A}_k^\pm)^c|.$$

Hence (10.8) gives an approximation of exploration process $Z_k^\pm$ in terms of the branching process $\hat{Z}_k^\pm$, with error terms $|\hat{A}_k^\pm \cap (A_k^\pm)^c|$ and $|A_k^\pm \cap (\hat{A}_k^\pm)^c|$.

10.2.4 Coupling results

We now present our main result on the coupling between the exploration process $\{Z_k^\pm : k \geq 1\}$ and the delayed branching process $\{\hat{Z}_k^\pm : k \geq 1\}$ described above. Theorem 10.1 gives a uniform bound on the errors

$$|\hat{A}_k^\pm \cap (A_k^\pm)^c| \quad \text{and} \quad |A_k^\pm \cap (\hat{A}_k^\pm)^c|,$$

in terms of $n^{-\kappa} \hat{Z}_k^\pm$, for some $\kappa > 0$ up until a certain number of steps in the exploration process. As mentioned earlier, the strength of this new coupling is that this number of steps is equivalent to having discovered $n^{1-\delta}$ nodes for arbitrarily small $0 < \delta < 1$. Moreover, the coupled branching process is independent of the bi-degree sequence and of the number of nodes.

Theorem 10.1. *Let $G_n = \mathcal{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$ and η, ε be as in Assumption 10.1. Then, for any $0 < \delta < 1$ and any $0 < \kappa < \min\{\delta\eta, \varepsilon\}$, there exist finite constants $K, a > 0$ such that for all $1 \leq k \leq (1 - \delta) \log_\mu n$, on the event Ξ_n ,*

$$\mathbb{P}_n \left(\bigcap_{m=1}^k \left\{ \left| \hat{A}_m^\pm \cap (A_m^\pm)^c \right| \leq \hat{Z}_m^\pm n^{-\kappa}, \left| A_m^\pm \cap (\hat{A}_m^\pm)^c \right| \leq \hat{Z}_m^\pm n^{-\kappa} \right\} \right) \geq 1 - Kn^{-a}.$$

As an immediate corollary, relation (10.8) gives:

Corollary 10.2. *Let $G_n = \mathcal{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$ and η, ε be as in Assumption 10.1. Then, for any $0 < \delta < 1$, any $0 < \kappa < \min\{\delta\eta, \varepsilon\}$, there exist finite constants $K, a > 0$ such that for all $1 \leq k \leq (1 - \delta) \log_\mu n$, on the event Ξ_n ,*

$$\mathbb{P}_n \left(\bigcap_{m=1}^k \left\{ \hat{Z}_m^\pm (1 - n^{-\kappa}) \leq Z_m^\pm \leq \hat{Z}_m^\pm (1 + n^{-\kappa}) \right\} \right) \geq 1 - Kn^{-a}.$$

The proof of Theorem 10.1 is given in Section 10.2.7. The proof uses several technical results that require conditioning on carefully selected events. We will, therefore, first give a high-level explanation of the proof strategy.

10.2.5 Road map for the proof of Theorem 10.1

Fix $0 < \delta < 1$, $0 < \kappa < \min\{\delta\eta, \varepsilon\}$, and define for $m \geq 1$

$$C_m = \bigcap_{r=1}^m \left\{ \left| \hat{A}_r^\pm \cap (A_r^\pm)^c \right| \leq \hat{Z}_r^\pm n^{-\kappa}, \left| A_r^\pm \cap (\hat{A}_r^\pm)^c \right| \leq \hat{Z}_r^\pm n^{-\kappa} \right\}. \quad (10.9)$$

For the proof of Theorem 10.1 we will show that for all $1 \leq k \leq (1 - \delta) \log_\mu n$, on the event Ξ_n , $\mathbb{P}_n(C_k^c) \leq Kn^{-a}$, for some constants $K, a > 0$.

First note that $W_k^\pm = \hat{Z}_k^\pm / (\nu \mu^{k-1})$ is a mean one martingale with respect to the filtration generated by the process $\{\hat{Z}_k^\pm : k \geq 1\}$. Hence, by the martingale convergence theorem,

$$W^\pm = \lim_{k \rightarrow \infty} \hat{Z}_k^\pm / (\nu \mu^{k-1}) \quad \text{a.s.}$$

exists and satisfies $\mathbb{E}[W^\pm] \leq 1$. To see that under the conditions in Assumption 10.1 W^+ and W^- are non-trivial, it is useful to define first $\{\mathcal{Z}_k^+ : k \geq 0\}$ and $\{\mathcal{Z}_k^- : k \geq 0\}$ to be (non-delayed) Galton-Watson processes having offspring distribution $f^{-,+}$ and $f^{+,-}$, respectively, and let

$$\mathcal{W}^\pm = \lim_{k \rightarrow \infty} \mathcal{Z}_k^\pm / \mu^k \quad \text{a.s.}$$

be its corresponding martingale limit. Now recall that $f^{-,+}$ and $f^{+,-}$ have finite moments of order $1 + \eta > 1$, which implies that

$$\sum_{t=1}^{\infty} t \log t f^{-,+}(t) < \infty \quad \text{and} \quad \sum_{t=1}^{\infty} t \log t f^{+,-}(t) < \infty,$$

a necessary and sufficient condition for $\mathcal{W}^\pm$ to be non-trivial (see, e.g., [4]). More precisely, if $q^\pm = \mathbb{P}(\mathcal{Z}_m^\pm = 0 \text{ for some } m)$ denotes the probability of extinction of $\{\mathcal{Z}_k^\pm : k \geq 0\}$, then $q^\pm < 1$ and $\mathbb{P}(\mathcal{W}^\pm = 0) = q^\pm$. It follows from these observations that $\mathbb{E}[W^\pm] = 1$.

The first step of the proof is to split the probability based on the survival of the branching process as follows

$$\mathbb{P}_n(C_k^c) \leq \mathbb{P}_n(C_k^c, W^\pm = 0) + \mathbb{P}_n(C_k^c, W^\pm > 0).$$

We deal with $\mathbb{P}_n(C_k^c, W^\pm = 0)$ in Proposition 10.8. For the other probability we will condition on the event that the branching process $\hat{Z}_k^\pm$ behaves typically, i.e. does not deviate too much from its mean. For this we introduce for $0 < \tau < 1$ and $m \geq 1$ the event

$$I_m = \bigcap_{r=1}^m \left\{ \hat{Z}_r^\pm / \mu^r \leq n^\tau \right\}, \quad (10.10)$$

so that by Doob's inequality $\mathbb{P}(I_k^c) \leq (\mu/\nu)n^{-\tau}$.

For the last steps of the proof of Theorem 10.1 we use an induction argument of the form

$$\mathbb{P}_n(C_k^c \cap I_k, W^\pm > 0) \leq \mathbb{P}_n(C_{k-1} \cap C_k^c \cap I_k, W^\pm > 0) + \mathbb{P}_n(C_{k-1}^c \cap I_k, W^\pm > 0),$$

for which we use bounds established in Lemma 10.7.

We remark that there are some technical details that will require us to condition on additional events. We deal with this in the proofs themselves. However this does not change the main line of reasoning that we present here.

Remark 10.3. We will use the events C_m and I_m , as defined by, respectively, (10.9) and (10.10), in several places in the proofs leading to Theorem 10.1. Therefore, for the remainder of this chapter, we will reserve the constants $\kappa > 0$ and $0 < \tau < 1$ for the scaling $n^{-\kappa}$ and n^τ in the events C_m and I_m , respectively.

For both Lemma 10.7 and Proposition 10.8 we need to control the difference $|\hat{\chi}_i^\pm - \chi_i^\pm|$, between the offspring of the exploration process $Z_k^\pm$ and the branching process $\hat{Z}_k^\pm$. For this define the function

$$\Phi(T) = \frac{4}{\nu n} \sum_{r=1}^n (1 - \mathcal{I}_r(T)) D_r^+ D_r^- + \frac{4\mu t}{\nu n} + 3n^{-\varepsilon}, \quad (10.11)$$

with $\mathcal{I}_r(T)$ defined as in (10.4). We will show in Lemma 10.5 that $\Phi(T)$ gives an upper bound on the difference $\mathbb{E}[|\hat{\chi}_i^\pm - \chi_i^\pm|]$, that holds up until a certain time T_n which depends on n . We then establish bounds on the expectations of $\Phi(T)$ for the case where $\hat{Z}_k^\pm$ behaves typically, i.e. conditioned on the event I_m .

In addition to these technical results we will also use some general results on delayed branching processes that we will present first.

10.2.6 Some results for delayed branching processes

Our first result for a general delayed branching process is an expression for its extinction probability in terms of the probability of extinction of the corresponding non-delayed process, as well as for the distribution of its number of offspring

conditional on extinction. Since these results are independent of the coupling with the graph, we do not use the super-scripts $+$ and $-$. Moreover, in this section g and g^* denote arbitrary probability mass functions, unrelated to the bi-degree sequence.

Lemma 10.3. *Let $\{\mathcal{Z}_k : k \geq 0\}$ denote a (non-delayed) branching process having offspring distribution g^* and extinction probability q and let $\{\hat{\mathcal{Z}}_k : k \geq 1\}$ be a delayed branching process having offspring distributions (g, g^*) , with finite mean (ν, μ) . Suppose $q > 0$. Then, conditioned on extinction, $\{\hat{\mathcal{Z}}_k : k \geq 1\}$ is a delayed branching process with offspring distributions $(\tilde{g}, \tilde{g}^*)$ with*

$$\tilde{g}(i) = \frac{g(i)q^i}{\sum_{t=0}^{\infty} g(t)q^t} \quad \text{and} \quad \tilde{g}^*(i) = g^*(i)q^{i-1}, \quad i \geq 0. \quad (10.12)$$

Moreover, $\mathbb{P}(\hat{\mathcal{Z}}_k = 0 \text{ for some } k \geq 1) = \sum_{t=0}^{\infty} g(t)q^t$.

Proof. Let $\hat{\chi}_\emptyset$ have distribution g and let $\{\mathcal{Z}_{k-1,i}\}_{i \geq 1}$ be a sequence of i.i.d. copies of $\mathcal{Z}_{k-1}$, independent of $\hat{\chi}_\emptyset$. In addition let $W = \lim_{k \rightarrow \infty} \hat{\mathcal{Z}}_k / (\nu \mu^{k-1})$ denote the a.s. limit of the martingale $\hat{\mathcal{Z}}_k / (\nu \mu^{k-1})$. Computing the probability generating function of $\hat{\mathcal{Z}}_k$ we obtain

$$\begin{aligned} & \mathbb{P}(W = 0) \mathbb{E} \left[s^{\hat{\mathcal{Z}}_k} \mid W = 0 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^{\hat{\chi}_\emptyset} \mathcal{Z}_{k-1,i} \mathbb{1}_{\{W=0\}} \right] \\ &= \mathbb{E} \left[s^{\sum_{i=1}^{\hat{\chi}_\emptyset} \mathcal{Z}_{k-1,i}} \mathbb{1}_{\{W=0, \hat{\chi}_\emptyset \geq 1\}} \right] + \mathbb{P}(\hat{\chi}_\emptyset = 0) \\ &= \mathbb{E} \left[\mathbb{E} \left[s^{\sum_{i=1}^{\hat{\chi}_\emptyset} \mathcal{Z}_{k-1,i}} \prod_{i=1}^{\hat{\chi}_\emptyset} \mathbb{1}_{\{\mathcal{W}_i = 0\}} \mid \hat{\chi}_\emptyset \right] \mathbb{1}_{\{\hat{\chi}_\emptyset \geq 1\}} \right] + \mathbb{P}(\hat{\chi}_\emptyset = 0) \\ &= \mathbb{E} \left[\left(\mathbb{E} \left[s^{\mathcal{Z}_{k-1}} \mathbb{1}_{\{\mathcal{W} = 0\}} \right] \right)^{\hat{\chi}_\emptyset} \mathbb{1}_{\{\hat{\chi}_\emptyset \geq 1\}} \right] + \mathbb{P}(\hat{\chi}_\emptyset = 0) \\ &= \mathbb{E} \left[\left(\mathbb{E} \left[s^{\mathcal{Z}_{k-1}} \mid \mathcal{W} = 0 \right] q \right)^{\hat{\chi}_\emptyset} \right], \end{aligned}$$

where $\mathcal{W}_i$ is the a.s. limit of the martingale $\{\mathcal{Z}_{k,i}/\mu^k : k \geq 0\}$ that has as root the i th individual in the first generation of $\{\hat{\mathcal{Z}}_k : k \geq 1\}$. Also,

$$\mathbb{P}(W = 0) = \mathbb{P}(\hat{\chi}_\emptyset = 0) + \mathbb{P} \left(\hat{\chi}_\emptyset \geq 1, \bigcap_{i=1}^{\hat{\chi}_\emptyset} \{\mathcal{W}_i = 0\} \right) = \sum_{j=0}^{\infty} g(j)q^j.$$

Hence,

$$\mathbb{E} \left[s^{\hat{\mathcal{Z}}_k} \mid W = 0 \right] = \sum_{j=0}^{\infty} \left(\mathbb{E} \left[s^{\mathcal{Z}_{k-1}} \mid \mathcal{W} = 0 \right] \right)^j \tilde{g}(j),$$

where

$$\tilde{g}(j) = \frac{q^j g(j)}{\sum_{t=0}^{\infty} g(t)q^t}, \quad j \geq 0.$$

Since conditionally on extinction $\{\mathcal{Z}_k : k \geq 0\}$ is a subcritical (non-delayed) branching process with offspring distribution $\tilde{g}^*(j) = g^*(j)q^{j-1}$, $j \geq 0$, see, e.g., [4, pp. 52], the result follows. $\square$

The second result we show is in some sense the counterpart of Doob's maximal martingale inequality. It states that provided the limiting martingale is strictly positive, the branching process itself cannot grow too slowly.

Lemma 10.4. *Suppose $\{\hat{Z}_k : k \geq 1\}$ is a delayed branching process with offspring distributions (g, g^*) , where g^* has finite $1 + \eta \in (1, 2]$ moment and mean $\mu > 1$, and g has finite mean $\nu > 0$. Let $W = \lim_{k \rightarrow \infty} \hat{Z}_k / (\nu \mu^{k-1})$. Then, for any $1 < u < \mu$, there exists a constant $K < \infty$ such that for any $k \geq 1$,*

$$\mathbb{P} \left(\inf_{r \geq k} \frac{\hat{Z}_r}{u^r} < 1, W > 0 \right) \leq K (u^{-\eta k} + (u/\mu)^{\omega k} \mathbb{1}_{\{q > 0\}}),$$

where q is the extinction probability of a branching process having offspring distribution g^* , $\lambda = \sum_{i=1}^{\infty} g^*(i) i q^{i-1}$, and $\omega = -\log \lambda / \log \mu > 0$ if $q > 0$.

Proof. We start by defining for $r \geq k$ the event

$$J_r = \left\{ \min_{k \leq j \leq r} \hat{Z}_j / u^j \geq 1 \right\}$$

and letting $p_r = \mathbb{P}(W > 0, (J_r)^c)$. Let $\{\hat{\chi}, \hat{\chi}_i\}$ be a sequence of i.i.d. random variables having distribution f and use the technical Lemma 3.5, applied conditionally on $\hat{Z}_{r-1}$, to obtain, for some constant $K_1 < \infty$,

$$\begin{aligned} p_r &\leq \mathbb{P} \left(J_{r-1}, \hat{Z}_r \leq u^r \right) + p_{r-1} \\ &\leq \mathbb{P} \left(\hat{Z}_{r-1} \geq u^{r-1}, \sum_{i=1}^{\hat{Z}_{r-1}} \hat{\chi}_i \leq u \hat{Z}_{r-1} \right) + p_{r-1} \\ &\leq \mathbb{P} \left(\hat{Z}_{r-1} \geq u^{r-1}, \sum_{i=1}^{\hat{Z}_{r-1}} (\mu - \hat{\chi}_i) \geq (\mu - u) \hat{Z}_{r-1} \right) + p_{r-1} \\ &\leq \mathbb{E} \left[\mathbb{1}_{\{\hat{Z}_{r-1} \geq u^{r-1}\}} \frac{K_1 \mathbb{E} [|\hat{\chi} - \mu|^{1+\eta}]}{(\mu - u)^{1+\eta} (\hat{Z}_{r-1})^\eta} \right] + p_{r-1} \\ &\leq K_2 u^{-\eta(r-1)} + p_{r-1}, \end{aligned}$$

where $K_2 = K_1 \mathbb{E} [|\hat{\chi} - \mu|^{1+\eta}] / (\mu - u)^{1+\eta}$ and $\mathbb{E} [(\hat{\chi})^{1+\eta}] < \infty$ since g^* has finite $1 + \eta$ moment. It follows from iterating the inequality derived above that

$$p_r \leq K_2 \sum_{j=k}^{r-1} \frac{1}{u^{\eta j}} + p_k \leq \frac{K_2}{(u^\eta - 1) u^{\eta(k-1)}} + \mathbb{P} \left(W > 0, \hat{Z}_k < u^k \right)$$

for all $r \geq k$. It remains to bound the last probability.

Let $\{\mathcal{Z}_k : k \geq 0\}$ be a (non-delayed) branching process having offspring distribution g^* , and let $\mathcal{W} = \lim_{k \rightarrow \infty} \mathcal{Z}_k / \mu^k$. It is well known, see [4, pp. 52], that conditional on non-extinction, $\mathcal{W}$ has an absolutely continuous distribution on $(0, \infty)$. Note also that, for any $m \geq 1$,

$$W_{m+k} = \frac{\hat{Z}_{m+k}}{\nu \mu^{m+k-1}} = \frac{1}{\nu \mu^{m+k-1}} \sum_{i \in \hat{A}_k} \mathcal{Z}_{m,i},$$

where the $\{\mathcal{Z}_{m,i}\}$ are i.i.d. copies of $\mathcal{Z}_m$ and $\hat{A}_k$ is the set of individuals in the k th generation of $\{\hat{Z}_k : k \geq 1\}$, and therefore, for any $k \geq 1$,

$$W_{m+k} - W_k = \frac{1}{\nu\mu^{k-1}} \sum_{i \in \hat{A}_k} \left(\frac{\mathcal{Z}_{m,i}}{\mu^m} - 1 \right).$$

Now define $\mathcal{W}_i = \lim_{m \rightarrow \infty} \mathcal{Z}_{m,i}/\mu^m$ to obtain that

$$W - W_k = \frac{1}{\nu\mu^{k-1}} \sum_{i \in \hat{A}_k} (\mathcal{W}_i - 1), \quad (10.13)$$

where the $\{\mathcal{W}_i\}$ are i.i.d. copies of $\mathcal{W}$, independent of the history of the tree up to generation k . It follows that for $x_k = 2u^k/(\nu\mu^{k-1})$,

$$\begin{aligned} \mathbb{P}(\hat{Z}_k < u^k, W > 0) &\leq \mathbb{P}(\hat{Z}_k < u^k, W \geq x_k) + \mathbb{P}(0 < W < x_k) \\ &\leq \mathbb{P}(\hat{Z}_k < u^k, W - W_k \geq x_k - u^k/(\nu\mu^{k-1})) \\ &\quad + \mathbb{P}(0 < W < x_k) \\ &= \mathbb{E} \left[\mathbb{1}_{\{\hat{Z}_k < u^k\}} \mathbb{P} \left(\sum_{i \in \hat{A}_k} (\mathcal{W}_i - 1) \geq u^k \middle| \hat{Z}_k \right) \right] \\ &\quad + \mathbb{P}(0 < W < x_k). \end{aligned}$$

Now note that $\mathbb{E}[\hat{\chi}^{1+\eta}] < \infty$ implies that $\mathbb{E}[\mathcal{W}^{1+\eta}] < \infty$. Therefore, by Lemma 3.5, applied conditionally on $\hat{Z}_k$, we obtain

$$\mathbb{P} \left(\sum_{i \in \hat{A}_k} (\mathcal{W}_i - 1) \geq u^k \middle| \hat{Z}_k \right) \leq K_3 \mathbb{E}[|\mathcal{W} - 1|^{1+\eta}] \frac{\hat{Z}_k}{u^{(1+\eta)k}},$$

for some other constant $K_3 < \infty$. Hence,

$$\mathbb{P}(\hat{Z}_k < u^k, W > 0) \leq \frac{K_3 \mathbb{E}[|\mathcal{W} - 1|^{1+\eta}]}{u^{\eta k}} + \mathbb{P}(0 < W < x_k).$$

Finally, to bound $\mathbb{P}(0 < W < x_k)$, note that W admits the representation

$$W = \frac{1}{\nu} \sum_{i=1}^{\hat{\chi}_\emptyset} \mathcal{W}_i,$$

where the $\{\mathcal{W}_i\}$ are i.i.d. copies of $\mathcal{W}$, independent of $\hat{\chi}_\emptyset$, with $\hat{\chi}_\emptyset$ distributed according to g . Note that $W > 0$ implies that at least one of the $\mathcal{W}_i$ is strictly positive. Let $N(t)$ be the number of non-zero random variables among $\{\mathcal{W}_1, \dots, \mathcal{W}_t\}$. It follows that if we let $\{V_i\}$ be i.i.d. random variables having the same distribution as $\mathcal{W}$ given $\mathcal{W} > 0$, then

$$\mathbb{P}(0 < W < x_k) = \mathbb{P} \left(\frac{1}{\nu} \sum_{i=1}^{N(\hat{\chi}_\emptyset)} V_i < x_k, N(\hat{\chi}_\emptyset) \geq 1 \right) \leq \mathbb{P}(V_1 < \nu x_k).$$

Hence, if $w(t)$ denotes the density of $\mathcal{W}$ conditional on non-extinction,

$$\mathbb{P}(V_1 < \nu x_k) = \mathbb{P}(\mathcal{W} < \nu x_k | \mathcal{W} > 0) = \int_0^{\nu x_k} w(t) dt.$$

By [74, Theorem 1] (see also [12, Theorem 4]) we have that if $\lambda = \sum_{i=1}^{\infty} g^*(i) i q^{i-1} > 0$, which under the assumptions of the lemma occurs whenever $q > 0$, then there exists a constant $K_4 < \infty$ such that

$$\int_0^{\nu x_k} w(t) dt \leq K_4 (\nu x_k)^\omega$$

for $\omega = -\log \lambda / \log \mu$; whereas if $g^*(0) + g^*(1) = 0$, then [12, Theorem 3] gives that for every $a > 0$ there exists a constant $Q_a < \infty$ such that

$$\int_0^{\nu x_k} w(t) dt \leq Q_a (\nu x_k)^a.$$

We conclude that for $a^* = \eta \log u / \log(\mu/u)$ and $K_5 = Q_{a^*}$,

$$\begin{aligned} \mathbb{P}\left(\min_{r \geq k} \frac{\hat{Z}_r}{u^r} < 1, W > 0\right) &\leq \frac{K_2}{(u^\eta - 1)u^{\eta(k-1)}} + \frac{K_3 \mathbb{E}[|\mathcal{W} - 1|^{1+\eta}]}{u^{\eta k}} \\ &\quad + K_4 (\nu x_k)^\omega \mathbb{1}_{\{q > 0\}} + K_5 (\nu x_k)^{a^*} \\ &\leq K \left(u^{-\eta k} + \left(\frac{u}{\mu} \right)^{\omega k} \mathbb{1}_{\{q > 0\}} \right), \end{aligned}$$

for some constant $K < \infty$. □

10.2.7 Proving the coupling theorem

In this section we prove Theorem 10.1. As mentioned in Section 10.2, the coupling we have constructed is based on minimizing the Kantorovich-Rubinstein distance between the distributions $H_i^\pm$ and $F^\pm$, and the main difficulty lies in the fact that this distance grows as the number of explored stubs in the graph grows. The proof of the main theorem is based on four technical results, Lemmas 10.5, 10.7, 10.6 and Proposition 10.8, which we state and prove below.

Throughout this section let

$$Y_k^\pm = \sum_{r=1}^k Z_r^\pm, \quad k \geq 1; \quad Y_0^\pm = 0.$$

The first of the technical lemmas gives us an upper bound for the Kantorovich-Rubinstein distance conditionally on the history of the graph exploration process and its coupled tree up to the moment that the stub $\mathbf{i}$ is about to be traversed.

Lemma 10.5. *Let $\mathcal{F}_i$ denote the sigma-algebra generated by the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ and the graph exploration process up to the time that outbound (inbound) stub $\mathbf{i}$ is about to be traversed. Then, on the event Ξ_n , for all n sufficiently large, and for $T_i^\pm \leq (\nu/2)n$, we have*

$$\mathbb{E}[|\hat{\chi}_i^\pm - \chi_i^\pm| | \mathcal{F}_i] \leq \Phi(T_i^\pm),$$

where $\Phi(T)$ is defined as in (10.11).

Proof. We first point out that for $\mathbf{i} = \emptyset$ the result trivially holds, since on Ξ_n

$$\mathbb{E} [|\hat{\chi}_\emptyset^\pm - \chi_\emptyset^\pm|] = d_1(F_n^\pm, F^\pm) \leq n^{-\varepsilon}.$$

For $\mathbf{i} \neq \emptyset$ we have

$$\mathbb{E} [|\hat{\chi}_\mathbf{i}^\pm - \chi_\mathbf{i}^\pm| | \mathcal{F}_\mathbf{i}] = d_1(H_\mathbf{i}^\pm, F^\pm) \leq d_1(H_\mathbf{i}^\pm, F_n^\pm) + d_1(F_n^\pm, F^\pm).$$

Since on the event Ξ_n we have that the second distance is smaller or equal than $n^{-\varepsilon}$, we only need to analyze the first one, for which we start with the $+$ case. To this end write

$$\begin{aligned} d_1(H_\mathbf{i}^+, F_n^+) &= \sum_{k=0}^{\infty} \left| \sum_{j=0}^k (h_\mathbf{i}^+(j) - f_n^+(j)) \right| = \sum_{k=0}^{\infty} \left| \sum_{j=k+1}^{\infty} (f_n^+(j) - h_\mathbf{i}^+(j)) \right| \\ &= \sum_{k=0}^{\infty} \left| \sum_{j=k+1}^{\infty} \sum_{r=1}^n \mathbb{1}_{\{D_r^+ = j\}} D_r^- \left(\frac{\mathcal{I}_r(T_\mathbf{i}^+)}{L_n - T_\mathbf{i}^+} - \frac{1}{L_n} \right) \right| \\ &\leq \sum_{k=0}^{\infty} \sum_{r=1}^n \left| \frac{\mathcal{I}_r(T_\mathbf{i}^+)}{L_n - T_\mathbf{i}^+} - \frac{1}{L_n} \right| D_r^- \mathbb{1}_{\{D_r^+ > k\}} \\ &= \sum_{r=1}^n \left| \frac{(\mathcal{I}_r(T_\mathbf{i}^+) - 1)L_n + T_\mathbf{i}^+}{(L_n - T_\mathbf{i}^+)L_n} \right| D_r^+ D_r^- \\ &\leq \frac{1}{L_n - T_\mathbf{i}^+} \sum_{r=1}^n (1 - \mathcal{I}_r(T_\mathbf{i}^+)) D_r^+ D_r^- + \frac{T_\mathbf{i}^+ \mu_n}{L_n - T_\mathbf{i}^+}, \end{aligned}$$

where $\mu_n = L_n^{-1} \sum_{r=1}^n D_r^+ D_r^-$ is the common mean of $F_n^{-,+}$ and $F_n^{+,-}$. Symmetrically,

$$d_1(H_\mathbf{i}^-, F_n^-) \leq \frac{1}{L_n - T_\mathbf{i}^-} \sum_{r=1}^n (1 - \mathcal{I}_r(T_\mathbf{i}^-)) D_r^+ D_r^- + \frac{T_\mathbf{i}^- \mu_n}{L_n - T_\mathbf{i}^-}.$$

Now note that on the event Ξ_n

$$\mu_n \leq \mu + d_1(F_n^\pm, F^\pm),$$

and if ν_n denotes the common mean of G_n^+ and G_n^- , then

$$\frac{L_n}{n} = \nu_n \geq \nu - d_1(G_n^\pm, G^\pm),$$

which in turn implies that

$$(L_n - T_\mathbf{i}^\pm)^{-1} \leq (\nu n - T_\mathbf{i}^\pm - n d_1(G_n^\pm, G^\pm))^{-1}.$$

We conclude that, on the event Ξ_n and for $T_\mathbf{i}^\pm \leq (\nu/2)n$,

$$\begin{aligned} d_1(H_\mathbf{i}^\pm, F^\pm) &\leq \frac{1}{(\nu/2)n - n^{1-\varepsilon}} \sum_{r=1}^n (1 - \mathcal{I}_r(T_\mathbf{i}^\pm)) D_r^+ D_r^- + \frac{T_\mathbf{i}^\pm (\mu + n^{-\varepsilon})}{(\nu/2)n - n^{1-\varepsilon}} + n^{-\varepsilon} \\ &\leq \frac{4}{\nu n} \sum_{r=1}^n (1 - \mathcal{I}_r(T_\mathbf{i}^\pm)) D_r^+ D_r^- + \frac{4\mu T_\mathbf{i}^\pm}{\nu n} + 3n^{-\varepsilon} \end{aligned}$$

for all $n \geq (4/\nu)^{1/\varepsilon}$. □

The second preliminary result provides an estimate for the expected value of the bound obtained in the previous lemma on the set where $\{\hat{Z}_k^\pm : k \geq 0\}$ behaves typically, i.e., without exhibiting large deviations from its mean.

Lemma 10.6. *Define $\Phi(T)$ according to (10.11), fix $0 < \tau < 1$ and let the event I_m be as in (10.10). Then, on the event Ξ_n , there exists a constant $K < \infty$ such that for any $0 \leq T \leq \nu n/2$ and any $k \geq 1$,*

$$\begin{aligned} \mathbb{E}_n [\Phi(T)] &\leq K \left(\frac{T^\eta}{n^\tau} + n^{-\varepsilon} \right) \quad \text{and} \\ \mathbb{E}_n \left[\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(T) \right] &\leq K \mu^k \left(\frac{T^\eta}{n^{\eta(1-\tau)}} + n^{-\varepsilon} \right) \end{aligned}$$

where $0 < \varepsilon < 1$ and $0 < \eta \leq 1$ are those from Assumption 10.1.

Proof. We start by proving the bound for $\mathbb{E}_n [\Phi(T)]$. Let X_r denote either D_r^+ or D_r^- depending on whether we are exploring outbound stubs or inbound stubs, respectively. Recall that $\mathcal{I}_r(T)$ is the indicator of node r having label 1 at time T in the graph exploration process. Next, note that $\mathcal{I}_r(0) = 1$, $\mathbb{E}_n [\mathcal{I}_r(1)] = 1 - 1/n$, and for any $2 \leq T < L_n$,

$$\begin{aligned} \mathbb{E}_n [\mathcal{I}_r(T)] &= \left(1 - \frac{1}{n} \right) \prod_{s=1}^{T-1} \left(1 - \frac{X_r}{L_n - s} \right) \\ &\geq \left(1 - \frac{1}{n} \right) \left(1 - \frac{X_r}{L_n - T} \right)^{T-1}, \end{aligned}$$

from where it follows that $\mathbb{E}_n [1 - \mathcal{I}_r(0)] = 0$, $\mathbb{E}_n [1 - \mathcal{I}_r(1)] = 1/n$, and for $2 \leq T < L_n$,

$$\begin{aligned} \mathbb{E}_n [1 - \mathcal{I}_r(T)] &\leq 1 - \left(1 - \frac{1}{n} \right) \left(1 - \frac{X_r}{L_n - T} \right)^{T-1} \\ &= \frac{1}{n} + \left(1 - \frac{1}{n} \right) \frac{X_r}{L_n - T} \sum_{s=0}^{T-2} \left(1 - \frac{X_r}{L_n - T} \right)^s \\ &\leq \frac{1}{n} + \frac{(T-1)X_r}{L_n - T}. \end{aligned}$$

Now let $0 < \eta \leq 1$ be as in Assumption 10.1 and note that since $\mathbb{E}_n [1 - \mathcal{I}_r(T)] \leq 1$, then,

$$\mathbb{E}_n [1 - \mathcal{I}_r(T)] \leq (\mathbb{E}_n [1 - \mathcal{I}_r(T)])^\eta \leq \left(\frac{1}{n} \right)^\eta + \left(\frac{(T-1)X_r}{L_n - T} \right)^\eta,$$

where we have used the inequality $(\sum_i y_i)^\beta \leq \sum_i y_i^\beta$ for $y_i \geq 0$ and $0 < \beta \leq 1$. It follows that on the event Ξ_n , with K_η defined by Assumption 10.1, and for any $2 \leq T < \nu n - n^{1-\varepsilon}$,

$$\begin{aligned}
& \frac{1}{n} \sum_{r=1}^n (\mathbb{E}_n [1 - \mathcal{I}_r(T)])^\eta D_r^+ D_r^- \\
& \leq \frac{1}{n^{1+\eta}} \sum_{r=1}^n D_r^+ D_r^- + \frac{(T-1)^\eta}{n(L_n - T)^\eta} \sum_{r=1}^n ((D_r^+)^\eta + (D_r^-)^\eta) D_r^+ D_r^- \\
& \leq \frac{\nu_n \mu_n}{n^\eta} + \frac{K_\eta T^\eta}{(L_n - T)^\eta} \\
& \leq \frac{(\nu + n^{-\varepsilon})(\mu + n^{-\varepsilon})}{n^\eta} + \frac{K_\eta T^\eta}{(\nu n - n^{1-\varepsilon} - T)^\eta} \leq K_0 n^{-\eta} (1 + T^\eta) \quad (10.14)
\end{aligned}$$

for some constant $K_0 < \infty$. It follows that

$$\mathbb{E}_n [\Phi(0)] \leq 3n^{-\varepsilon}, \quad \mathbb{E}_n [\Phi(1)] \leq 4\mu_n/(\nu n) + 3n^{-\varepsilon},$$

and for any $2 \leq T \leq \nu n/2$ we have, using (10.14),

$$\begin{aligned}
\mathbb{E}_n [\Phi(T)] & \leq \frac{4}{\nu n} \sum_{r=1}^n \mathbb{E}_n [1 - \mathcal{I}_r(T)] D_r^+ D_r^- + \frac{4\mu T}{\nu n} + 3n^{-\varepsilon} \\
& \leq \frac{4}{\nu n} \sum_{r=1}^n \mathbb{E}_n [1 - \mathcal{I}_r(T)]^\eta D_r^+ D_r^- + \frac{4\mu T}{\nu n} + 3n^{-\varepsilon} \\
& \leq K_0 n^{-\eta} + \frac{4\mu T}{\nu n} + 3n^{-\varepsilon} \\
& \leq K_1 \left(\frac{T^\eta}{n^\eta} + n^{-\varepsilon} \right),
\end{aligned}$$

for some other constant $K_1 < \infty$.

Next, to compute a bound for $\mathbb{E}_n [\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(T)]$ let $q = 1/\eta$ and $p = q/(q-1)$, with $p = \infty$ if $q = 1$, and use Hölder's inequality to obtain, for $1 \leq T \leq \nu n/2$,

$$\begin{aligned}
\mathbb{E}_n [\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(T)] & = \frac{4}{\nu n} \sum_{r=1}^n \mathbb{E}_n [\mathbb{1}_{I_k} \hat{Z}_k^\pm (1 - \mathcal{I}_r(T))] D_r^+ D_r^- \\
& \quad + \frac{4\mu T}{\nu n} \mathbb{E} [\mathbb{1}_{I_k} \hat{Z}_k^\pm] + 3n^{-\varepsilon} \mathbb{E} [\mathbb{1}_{I_k} \hat{Z}_k^\pm] \\
& \leq \frac{4}{\nu n} \sum_{r=1}^n \left(\mathbb{E} [\mathbb{1}_{I_k} (\hat{Z}_k^\pm)^p] \right)^{1/p} (\mathbb{E}_n [1 - \mathcal{I}_r(T)])^{1/q} D_r^+ D_r^- \\
& \quad + \frac{4\mu T}{\nu n} \mathbb{E} [\hat{Z}_k^\pm] + 3n^{-\varepsilon} \mathbb{E} [\hat{Z}_k^\pm].
\end{aligned}$$

Now note that, since $1/p = 1 - \eta$,

$$\begin{aligned}
\left(\mathbb{E} [\mathbb{1}_{I_k} (\hat{Z}_k^\pm)^p] \right)^{1/p} & \leq \left((\mu^k n^\tau)^{p-1} \mathbb{E} [\hat{Z}_k^\pm] \right)^{1/p} \\
& = \mu^k n^\tau \left(\frac{\nu}{\mu n^\tau} \right)^{1/p} = \left(\frac{\mu}{\nu} \right)^\eta \mathbb{E} [\hat{Z}_k^\pm] n^{\eta\tau}.
\end{aligned}$$

Combining this inequality with (10.14) gives, for $1 \leq T \leq \nu n/2$,

$$\begin{aligned}
& \mathbb{E}_n \left[\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(T) \right] \\
& \leq \mathbb{E} \left[\hat{Z}_k^\pm \right] \left(\frac{\mu^\eta n^{\eta\tau}}{\nu^\eta} \frac{4}{\nu n} \sum_{r=1}^n (\mathbb{E}_n [1 - \mathcal{I}_r(T)])^\eta D_r^+ D_r^- + \frac{4\mu T}{\nu n} + 3n^{-\varepsilon} \right) \\
& \leq \mathbb{E} \left[\hat{Z}_k^\pm \right] \left(\frac{\mu^\eta n^{\eta\tau}}{\nu^\eta} K_0 n^{-\eta} (1 + T^\eta) + \frac{4\mu T}{\nu n} + 3n^{-\varepsilon} \right) \\
& \leq K_2 \mu^k \left(\frac{T^\eta}{n^{\eta(1-\tau)}} + n^{-\varepsilon} \right),
\end{aligned}$$

for some constant $K_2 < \infty$. Noting that

$$\mathbb{E}_n \left[\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(0) \right] \leq \mathbb{E} \left[\hat{Z}_k^\pm \right] 3n^{-\varepsilon} = O(\mu^k n^{-\varepsilon}),$$

and taking $K = K_1 \vee K_2$ completes the proof. $\square$

The third technical lemma provides an estimate for the expected number of stubs that are discovered during step $k + 1$ of the graph exploration process, on the set where the coupling holds uniformly well up to step k . This bound is the key component that will enable the induction step in the proof Theorem 10.1:

Lemma 10.7. *Fix $0 < \delta < 1$, $0 < \tau < \min\{1/2, \delta\}$ and $0 < \kappa < \min\{\delta\eta, \varepsilon\}$. Let I_m be defined according to (10.10) and C_m according to (10.9). Then, on the event Ξ_n , there exists a constant $K < \infty$ such that for all $1 \leq k \leq (1 - \delta) \log n / \log \mu$,*

$$\begin{aligned}
& \mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} \left(\left| \hat{A}_{k+1}^\pm \cap (A_{k+1}^\pm)^c \right| + \left| A_{k+1}^\pm \cap (\hat{A}_{k+1}^\pm)^c \right| \right) \right] \\
& \leq K \mu^k \left(\mu^{\eta k} n^{-\eta(1-2\tau)} + k n^{-\varepsilon} \right),
\end{aligned}$$

where $0 < \varepsilon < 1$ and $0 < \eta \leq 1$ are those from Assumption 10.1.

Proof. Define

$$u_{k+1} = \mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} \left(\left| \hat{A}_{k+1}^\pm \cap (A_{k+1}^\pm)^c \right| + \left| A_{k+1}^\pm \cap (\hat{A}_{k+1}^\pm)^c \right| \right) \right],$$

The idea of the proof is to establish a recursive relation for u_{k+1} of the form

$$u_{k+1} = u_k + \mathbb{E}_n \left[\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(T_k) \right],$$

for values T_k , depending on k , and $\Phi(T)$ according to (10.11). Then we use the bound from Lemma 10.6 to get a recursive relation for u_{k+1} involving μ^k , which we can iterate to obtain the bound in the statement of the lemma.

Let $\mathcal{F}_m$ denote the sigma-algebra generated by the bi-degree sequence and the history of the exploration process up until step $m - 1$ is completed and note that this includes the value of $Z_m^\pm$. Then we condition on $\mathcal{F}_k$ to obtain that

$$u_{k+1} = \mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} \mathbb{E} \left[\sum_{\mathbf{i} \in \hat{A}_k^\pm \cap A_k^\pm} (\hat{\chi}_\mathbf{i}^\pm - \chi_\mathbf{i}^\pm)^+ + \sum_{\mathbf{i} \in \hat{A}_k^\pm \cap (A_k^\pm)^c} \hat{\chi}_\mathbf{i}^\pm \middle| \mathcal{F}_k \right] \right] \quad (10.15)$$

$$+ \mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} \mathbb{E} \left[\sum_{\mathbf{i} \in A_k^\pm \cap \hat{A}_k^\pm} (\chi_{\mathbf{i}}^\pm - \hat{\chi}_{\mathbf{i}}^\pm)^+ + \sum_{\mathbf{i} \in A_k^\pm \cap (\hat{A}_k^\pm)^c} \chi_{\mathbf{i}}^\pm \middle| \mathcal{F}_k \right] \right]. \quad (10.16)$$

To analyze the conditional expectations recall that $Y_k = \sum_{r=1}^k Z_k^\pm$ while $T_{\mathbf{i}}^\pm$ denotes the number of outbound(inbound) stubs that have been traversed in the exploration process up until we discover stub $\mathbf{i}$. Hence, for $\mathbf{i} \in \hat{A}_k^\pm \cap A_k^\pm$ we must have that $T_{\mathbf{i}}^\pm \leq Y_k^\pm$. Moreover, on the event $C_k \cap I_k$ we have that

$$Y_k^\pm \leq \sum_{i=1}^k Z_k^\pm \leq (1 + n^{-\kappa}) \sum_{i=1}^k \hat{Z}_k^\pm \leq 2n^\tau \sum_{i=1}^k \mu^i \leq 2\mu^{k+1} n^\tau / (\mu - 1).$$

Define $y_k = 2\mu^{k+1} n^\tau / (\mu - 1)$ and observe that for the range of values of k and τ in the lemma we have $y_k = o(n)$ as $n \rightarrow \infty$. It follows by Lemma 10.5 and the tower property that on the event $C_k \cap I_k$, the conditional expectation on $\mathcal{F}_k$ in (10.15) is bounded from above by

$$\begin{aligned} & \sum_{\mathbf{i} \in \hat{A}_k^\pm \cap A_k^\pm} \mathbb{E}_n [\Phi(T_{\mathbf{i}}^\pm) | \mathcal{F}_k] + \left| \hat{A}_k^\pm \cap (A_k^\pm)^c \right| \mu \\ & \leq \hat{Z}_k^\pm \mathbb{E}_n [\Phi(Y_k^\pm) | \mathcal{F}_k] + \left| \hat{A}_k^\pm \cap (A_k^\pm)^c \right| \mu, \\ & \leq \hat{Z}_k^\pm \mathbb{E}_n [\Phi(y_k) | \mathcal{F}_k] + \left| \hat{A}_k^\pm \cap (A_k^\pm)^c \right| \mu, \end{aligned}$$

where we have used the observation that $\Phi(T)$ is non-decreasing and $T_{\mathbf{i}}^\pm \leq Y_k^\pm \leq y_k$. Hence, (10.15) is bounded from above by

$$\mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} \hat{Z}_k^\pm \mathbb{E}_n [\Phi(y_k) | \mathcal{F}_k] \right] + \mu \mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} \left| \hat{A}_k^\pm \cap (A_k^\pm)^c \right| \right] \quad (10.17)$$

Similarly, the conditional expectation on $\mathcal{F}_k$ in (10.16) is bounded from above by

$$\begin{aligned} & \mathbb{E}_n \left[\sum_{\mathbf{i} \in A_k^\pm} (\chi_{\mathbf{i}}^\pm - \hat{\chi}_{\mathbf{i}}^\pm)^+ + \sum_{\mathbf{i} \in A_k^\pm \cap (\hat{A}_k^\pm)^c} \hat{\chi}_{\mathbf{i}}^\pm \middle| \mathcal{F}_k \right] \\ & \leq \sum_{\mathbf{i} \in A_k^\pm} \mathbb{E}_n [\Phi(T_{\mathbf{i}}^\pm) | \mathcal{F}_k] + \left| A_k^\pm \cap (\hat{A}_k^\pm)^c \right| \mu \\ & \leq Z_k^\pm \mathbb{E}_n [\Phi(Y_k^\pm) | \mathcal{F}_k] + \left| A_k^\pm \cap (\hat{A}_k^\pm)^c \right| \mu \\ & \leq (1 + n^{-\kappa}) \hat{Z}_k^\pm \mathbb{E}_n [\Phi(y_k) | \mathcal{F}_k] + \left| A_k^\pm \cap (\hat{A}_k^\pm)^c \right| \mu, \end{aligned}$$

where in the last step we also used that $Z_k \leq (1 + n^{-\kappa}) \hat{Z}_k$ on the event C_k . Therefore we have the following upper bound for (10.16):

$$\mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} (1 + n^{-\kappa}) \hat{Z}_k^\pm \mathbb{E}_n [\Phi(y_k) | \mathcal{F}_k] \right] + \mu \mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} \left| A_k^\pm \cap (\hat{A}_k^\pm)^c \right| \right]. \quad (10.18)$$

Combining (10.17) and (10.18), and using that $C_k \cap I_k \subseteq C_{k-1} \cap I_{k-1}$ gives

$$u_{k+1} \leq \mathbb{E}_n \left[\mathbb{1}_{C_k \cap I_k} (2 + n^{-\gamma}) \hat{Z}_k^\pm \mathbb{E}_n [\Phi(y_k) | \mathcal{F}_k] \right] + \mu u_k$$

$$\leq 3\mathbb{E}_n \left[\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(y_k) \right] + \mu u_k.$$

For the final step, we use Lemma 10.6 to get

$$\begin{aligned} 3\mathbb{E}_n \left[\mathbb{1}_{I_k} \hat{Z}_k^\pm \Phi(y_k) \right] &\leq 3K_1 \mu^k \left(\frac{y_k^\eta}{n^{\eta(1-\tau)}} + n^{-\varepsilon} \right) \\ &\leq K_2 \left(\mu^{(1+\eta)k} n^{-\eta(1-2\tau)} + \mu^k n^{-\varepsilon} \right) \end{aligned}$$

for some constants $K_1, K_2 < \infty$. Let $r_k = K_2 \mu^k (\mu^{\eta k} n^{-\eta(1-2\tau)} + n^{-\varepsilon})$ and iterate the inequality $u_{k+1} \leq r_k + \mu u_k$ to obtain

$$\begin{aligned} u_{k+1} &\leq \sum_{j=1}^k \mu^{j-1} r_{k+1-j} + \mu^k u_1 \\ &= \sum_{j=1}^k \mu^{j-1} K_2 \mu^{k+1-j} \left(\mu^{\eta(k+1-j)} n^{-\eta(1-2\tau)} + n^{-\varepsilon} \right) + \mu^k u_1 \\ &= K_2 \mu^k \left(\sum_{j=1}^k \mu^{\eta(k+1-j)} n^{-\eta(1-2\tau)} + k n^{-\varepsilon} \right) + \mu^k u_1 \\ &= K_2 \mu^k \left(n^{-\eta(1-2\tau)} \frac{\mu^\eta (\mu^{\eta k} - 1)}{\mu^\eta - 1} + k n^{-\varepsilon} \right) + \mu^k u_1. \end{aligned}$$

Noting that on the event Ξ_n

$$u_1 = \mathbb{E}_n \left[|\hat{\chi}_\emptyset^\pm - \chi_\emptyset^\pm| \right] \leq n^{-\varepsilon}$$

completes the proof. $\square$

The last preliminary result before the proof of Theorem 10.1 is an analysis of the coupling when the branching process becomes extinct, which occurs most likely when the out- or in-component of the node being explored in the graph is small.

Proposition 10.8. *Fix $0 < \delta < 1$, $0 < \kappa < \min\{\delta\eta, \varepsilon\}$ and define for $k \geq 1$ the event C_k according to (10.9). Let $W^\pm = \lim_{k \rightarrow \infty} \hat{Z}_k^\pm / (\nu \mu^{k-1})$. Then, on the event Ξ_n , there exists a constant $K < \infty$ such that for all $1 \leq k \leq (1 - \delta) \log n / \log \mu$,*

$$\mathbb{P}_n (C_k^c, W^\pm = 0) \leq K n^{-\varepsilon\eta/(1+\eta+\varepsilon)}.$$

Proof. Let $\{\mathcal{Z}_k^\pm : k \geq 0\}$ denote a non-delayed branching process with offspring distribution $f^{\mp, \pm}$ and let $q^\pm$ denote its extinction probability. We start by pointing out that if $q^\pm = 0$ the probability that the branching process $\{\hat{Z}_k^\pm : k \geq 1\}$ becomes extinct is zero unless $\hat{Z}_1^\pm = 0$, see [4, Theorem 4]). Since in our construction of the coupling $Z_1^\pm = \hat{Z}_1^\pm$, we may assume from now on that $q^\pm > 0$.

Analogously to the events I_m and C_m defined in (10.10) and (10.9), we now define for $m \geq 1$, $\epsilon > 0$,

$$B_m = \bigcap_{r=1}^m \left\{ \left| \hat{A}_r^\pm \cap (A_r^\pm)^c \right| = 0, \left| A_r^\pm \cap (\hat{A}_r^\pm)^c \right| = 0 \right\}, \quad (10.19)$$

$$J_m = \bigcap_{r=1}^m \left\{ \hat{Z}_r^\pm / (\lambda^\pm)^r \leq n^\epsilon \right\},$$

where $\lambda^\pm = \sum_{j=1}^\infty j f^\pm(j) (q^\pm)^{j-1} < 1$ and $\epsilon = \eta / (1 + \eta + \epsilon)$. Now note that

$$\mathbb{P}_n (C_k^c, W^\pm = 0) \leq \mathbb{P}_n (C_k^c \cap J_k) + \mathbb{P} (J_k^c, W^\pm = 0),$$

where the last probability is independent of the bi-degree sequence. Next, define,

$$\tilde{f}^\pm(i) = \frac{f^\pm(i) q^i}{\sum_{t=0}^\infty f^\pm(t) q^t} \quad \text{and} \quad \tilde{f}^{\mp, \pm}(i) = f^{\mp, \pm}(i) q^{i-1}, \quad i \geq 0,$$

according to (10.12) and use Lemma 10.3 to see that conditionally on $\{W^\pm = 0\}$, $\hat{Z}_k^\pm / (\nu^\pm (\lambda^\pm)^{k-1})$ is a mean one martingale, where $\nu^\pm$ is the mean of $\tilde{f}^\pm$ and $\lambda^\pm$ is the mean of $\tilde{f}^{\mp, \pm}$. It then follows from Doob's inequality that

$$\mathbb{P} (J_k^c, W^\pm = 0) \leq \mathbb{P} (J_k^c | W^\pm = 0) \leq (\nu^\pm / \lambda^\pm) n^{-\epsilon}.$$

Hence, we are left with analyzing $\mathbb{P}_n (C_k^c \cap J_k)$. To this end write

$$\mathbb{P}_n (C_k^c \cap J_k) \leq \mathbb{P}_n (B_k^c \cap J_k) \leq \sum_{r=1}^k \mathbb{P}_n (B_{r-1} \cap B_r^c \cap J_k)$$

and note that the event B_{r-1} implies that $Z_i^\pm = \hat{Z}_i^\pm$ for all $1 \leq i \leq r-1$. In addition define $r_n = \lfloor \epsilon \log n / |\log \lambda^\pm| \rfloor$ and observe that on the event $B_{r-1} \cap J_k$, $Z_{r-1}^\pm = \hat{Z}_{r-1}^\pm = 0$ for all $r-1 > r_n$. Since for any $r \geq 1$

$$\begin{aligned} \left| \hat{A}_r^\pm \cap (A_r^\pm)^c \right| &= \sum_{\mathbf{i} \in \hat{A}_{r-1}^\pm \cap A_{r-1}^\pm} \sum_{j=1}^{\hat{\chi}_i^\pm} \mathbb{1}_{\{(i,j) \in (A_r^\pm)^c\}} + \sum_{\mathbf{i} \in \hat{A}_{r-1}^\pm \cap (A_{r-1}^\pm)^c} \hat{\chi}_i^\pm \\ &= \sum_{\mathbf{i} \in \hat{A}_{r-1}^\pm \cap A_{r-1}^\pm} (\hat{\chi}_i^\pm - \chi_i^\pm)^+ + \sum_{\mathbf{i} \in \hat{A}_{r-1}^\pm \cap (A_{r-1}^\pm)^c} \hat{\chi}_i^\pm, \end{aligned} \quad (10.20)$$

and, similarly,

$$\left| A_r^\pm \cap (\hat{A}_r^\pm)^c \right| = \sum_{\mathbf{i} \in A_{r-1}^\pm \cap \hat{A}_{r-1}^\pm} (\chi_i^\pm - \hat{\chi}_i^\pm)^+ + \sum_{\mathbf{i} \in A_{r-1}^\pm \cap (\hat{A}_{r-1}^\pm)^c} \chi_i^\pm \quad (10.21)$$

so that $\mathbb{P}_n (B_{r-1} \cap B_r^c \cap J_k) = 0$ for all $r_n + 2 \leq r \leq k$. For $1 \leq r \leq r_n + 1$ we obtain using (10.20) and (10.21) that

$$\begin{aligned} \mathbb{P}_n (B_{r-1} \cap B_r^c \cap J_k) &\leq \mathbb{P}_n \left(\left| \hat{A}_r^\pm \cap (A_r^\pm)^c \right| + \left| A_r^\pm \cap (\hat{A}_r^\pm)^c \right| \geq 1, B_{r-1} \cap J_{r-1} \right) \\ &= \mathbb{P}_n \left(\sum_{\mathbf{i} \in \hat{A}_{r-1}^\pm} |\chi_i^\pm - \hat{\chi}_i^\pm| \geq 1, B_{r-1} \cap J_{r-1} \right). \end{aligned}$$

Now let $\mathcal{F}_r$ denote the sigma-algebra generated by the bi-degree sequence and the history of the exploration process up until step $r-1$ is completed, which

includes the value of $Z_r^\pm$. Also define $\mathcal{F}_i$ to be the sigma-algebra generated by the bi-degree sequence and the exploration process up to the time that outbound (inbound) stub i is about to be traversed and note that for $i \in \hat{A}_{r-1}^\pm$ we have $\mathcal{F}_{r-1} \subseteq \mathcal{F}_i$. Applying Markov's inequality conditionally on $\mathcal{F}_{r-1}$ we obtain

$$\begin{aligned} \mathbb{P}_n \left(\sum_{i \in \hat{A}_{r-1}^\pm} |\chi_i^\pm - \hat{\chi}_i^\pm| \geq 1, B_{r-1} \cap J_{r-1} \right) \\ \leq \mathbb{E}_n \left[1(B_{r-1} \cap J_{r-1}) \sum_{i \in \hat{A}_{r-1}^\pm} \mathbb{E}_n [|\chi_i^\pm - \hat{\chi}_i^\pm| | \mathcal{F}_{r-1}] \right]. \end{aligned}$$

To analyze the conditional expectation involving $\mathcal{F}_{r-1}$, recall that $T_i^\pm$ is the number of stubs (outbound for $+$ and inbound for $-$) that have been seen up until it is stub i 's turn to be traversed, and use Lemma 10.5 to obtain

$$\begin{aligned} \mathbb{E}_n [|\chi_i^\pm - \hat{\chi}_i^\pm| | \mathcal{F}_{r-1}] &= \mathbb{E}_n [\mathbb{E}_n [|\chi_i^\pm - \hat{\chi}_i^\pm| | \mathcal{F}_i] | \mathcal{F}_{r-1}] \\ &\leq \mathbb{E}_n [\mathbb{E} [\Phi(T_i^\pm)] | \mathcal{F}_{r-1}]. \end{aligned}$$

Recall that on the event B_{r-1} we have $Z_i^\pm = \hat{Z}_i^\pm$ for all $1 \leq i \leq r-1$, and therefore, for $i \in \hat{A}_{r-1}^\pm$, $T_i^\pm \leq Y_{r-1}^\pm = \hat{Y}_{r-1}^\pm = \sum_{i=1}^{r-1} \hat{Z}_i^\pm$. Moreover, on the event J_{r-1} we have $\hat{Y}_{r-1}^\pm \leq \sum_{i=1}^{r-1} (\lambda^\pm)^{r-1} n^\epsilon \leq n^\epsilon / (1 - \lambda^\pm)$, from where it follows from Lemma 10.6 that

$$\begin{aligned} \mathbb{P}_n (B_{r-1} \cap B_r^c \cap J_{r-1}) &\leq \mathbb{E}_n \left[\mathbb{1}_{\{J_{r-1}\}} \sum_{i \in \hat{A}_{r-1}^\pm} \mathbb{E}_n [\mathbb{E} [\Phi(n^\epsilon / (1 - \lambda^\pm))] | \mathcal{F}_{r-1}] \right] \\ &= \mathbb{E}_n \left[\mathbb{1}_{\{J_{r-1}\}} \hat{Z}_{r-1}^\pm \mathbb{E} [\Phi(n^\epsilon / (1 - \lambda^\pm))] \right] \\ &\leq (\lambda^\pm)^{r-1} n^\epsilon \mathbb{E}_n [\mathbb{E} [\Phi(n^\epsilon / (1 - \lambda^\pm))]] \\ &\leq K(\lambda^\pm)^{r-1} n^\epsilon \left(\frac{(n^\epsilon / (1 - \lambda^\pm))^\eta}{n^\eta} + n^{-\epsilon} \right) \\ &= O((\lambda^\pm)^r n^{-\epsilon}), \end{aligned}$$

as $n \rightarrow \infty$, where the last equality is due to our choice of ϵ . Thus, we have shown that

$$\mathbb{P}_n (C_k^c, W^\pm = 0) = O(n^{-\epsilon} + n^{-\epsilon}) = O(n^{-\epsilon\eta/(1+\eta+\epsilon)}),$$

as $n \rightarrow \infty$, which implies the result. $\square$

We are now ready to prove Theorem 10.1, which in view of Proposition 10.8, reduces to analyzing the event that the error in the coupling is large conditionally on the branching process surviving:

Proof of Theorem 10.1. Let $W^\pm = \lim_{k \rightarrow \infty} \hat{Z}_k^\pm / (\nu \mu^{k-1})$ and for each $m \geq 1$ define the event C_m according to (10.9). Now note that

$$\mathbb{P}_n (C_k^c) \leq \mathbb{P}_n (C_k^c, W^\pm = 0) + \mathbb{P}_n (C_k^c, W^\pm > 0),$$

where by Proposition 10.8 we have

$$\mathbb{P}_n((C_k^c, W^\pm = 0) \leq K_0 n^{-\varepsilon\kappa/(1+\kappa+\varepsilon)})$$

for some constant $K_0 < \infty$.

To analyze $\mathbb{P}_n(C_k^c, W^\pm > 0)$ we proceed similarly to the proof of Proposition 10.8 by setting $\tau = (\delta\eta - \kappa)/(4\eta) \in (0, \delta/4)$ and defining the events I_m and B_m , $m \geq 1$, according to (10.10) and (10.19), respectively. Define also $s_n = \min\{k, \lceil c \log n / \log \mu \rceil\}$, with $0 < c < \min\{\eta(1 - 2\tau)/(1 + \eta), \varepsilon\}$, and note that $C_r^c \subset B_r^c$ for any $r \geq 1$. We start by deriving an upper bound as follows:

$$\begin{aligned} \mathbb{P}_n(C_k^c, W^\pm > 0) &\leq \mathbb{P}_n(C_k^c \cap I_k, W^\pm > 0) + \mathbb{P}(I_k^c) \\ &\leq \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap I_k, W^\pm > 0) \\ &\quad + \mathbb{P}(C_{s_n-1}^c \cap I_k, W^\pm > 0) + \mathbb{P}(I_k^c) \\ &\leq \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap I_k, W^\pm > 0) \\ &\quad + \mathbb{P}_n(B_{s_n-1}^c \cap I_k) + \mathbb{P}(I_k^c) \\ &\leq \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap I_k, W^\pm > 0) \\ &\quad + \sum_{r=1}^{s_n-1} \mathbb{P}_n(B_{r-1} \cap B_r^c \cap I_k) + \mathbb{P}(I_k^c). \end{aligned}$$

Note that Doob's inequality gives $\mathbb{P}(I_k^c) \leq (\mu/\nu)n^{-\tau}$. Also, if we again let $\mathcal{F}_r$ denote the sigma-algebra generated by the bi-degree sequence and the history of the exploration process up until step $r - 1$ is completed, the same steps used in the proof of Proposition 10.8 give that for $1 \leq r \leq s_n - 1$,

$$\begin{aligned} \mathbb{P}_n(B_{r-1} \cap B_r^c \cap I_k) &\leq \mathbb{E}_n \left[\mathbb{1}_{I_{r-1}} \sum_{i \in \hat{A}_{r-1}^\pm} \mathbb{E}_n [\Phi(\hat{Y}_{r-1}^\pm)] | \mathcal{F}_{r-1} \right] \\ &= \mathbb{E}_n \left[\mathbb{1}_{I_{r-1}} \hat{Z}_{r-1}^\pm \Phi(\hat{Y}_{r-1}^\pm) \right] \\ &\leq \mathbb{E}_n \left[\mathbb{1}_{I_{r-1}} \hat{Z}_{r-1}^\pm \Phi(n^\tau \mu^r / (\mu - 1)) \right] \\ &\leq K_1 \mu^{r-1} \left(\frac{(n^\tau \mu^r / (\mu - 1))^\eta}{n^{\eta(1-\tau)}} + n^{-\varepsilon} \right), \end{aligned}$$

for some other constant $K_1 < \infty$. Here we have used that on the event I_{r-1} we have $\hat{Y}_{r-1}^\pm \leq \sum_{i=1}^{r-1} \mu^i n^\tau \leq n^\tau \mu^r / (\mu - 1)$ and that $\Phi(T)$ is non decreasing, followed by an application of Lemma 10.6. We then obtain that, as $n \rightarrow \infty$,

$$\mathbb{P}_n(B_{r-1} \cap B_r^c \cap I_k) = O \left(\frac{\mu^{r(1+\eta)}}{n^{\eta(1-2\tau)}} + \mu^r n^{-\varepsilon} \right),$$

which implies that

$$\sum_{r=1}^{s_n-1} \mathbb{P}_n(B_{r-1} \cap B_r^c \cap I_k) = O \left(\frac{\mu^{s_n(1+\eta)}}{n^{\eta(1-2\tau)}} + \mu^{s_n} n^{-\varepsilon} \right)$$

$$= O\left(n^{c(1+\eta)-\eta(1-2\tau)} + n^{c-\varepsilon}\right),$$

as $n \rightarrow \infty$.

We have thus shown that, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{P}_n(C_k^c, W^\pm > 0) &\leq \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap E_k, W^\pm > 0) \\ &\quad + O\left(n^{-\tau} + n^{c(1+\eta)-\eta(1-2\tau)} + n^{c-\varepsilon}\right), \end{aligned}$$

with all the exponents of n inside the big-O term strictly negative, since $c(1+\eta) < \eta(1-2\tau)$. To analyze the remaining probability we first introduce one last conditioning event. Let

$$b = \min\{(1-\delta)/2, (\varepsilon-\kappa)/2, (\eta\delta-\kappa)/4\}/(1-\delta),$$

and note that, since the conditions of Theorem 10.1 state that $\kappa < \min(\delta\eta, \varepsilon)$, that $0 < b < 1/2$. Set $1 < u = \mu^{1-b}$ and define

$$M_{s_n} = \left\{ \inf_{r \geq s_n} \hat{Z}_r^\pm / u^r \geq 1 \right\}.$$

Now write

$$\begin{aligned} \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap I_k, W^\pm > 0) &\leq \mathbb{P}(C_{s_n-1} \cap C_k^c \cap I_k \cap M_{s_n}) \\ &\quad + \mathbb{P}(M_{s_n}^c, W^\pm > 0) \\ &\leq \sum_{r=s_n}^k \mathbb{P}_n(C_{r-1} \cap C_r^c \cap I_k \cap M_{s_n}) \\ &\quad + \mathbb{P}(M_{s_n}^c, W^\pm > 0). \end{aligned}$$

By Lemma 10.4 we have, for some $K < \infty$,

$$\begin{aligned} \mathbb{P}(M_{s_n}^c, W^\pm > 0) &\leq K \left(u^{-\eta s_n} + (u/\mu)^{\omega^\pm s_n} \mathbb{1}_{\{q^\pm > 0\}} \right) \\ &= O\left(n^{-\eta c(1-b)} + n^{-\omega^\pm cb} \mathbb{1}_{\{q^\pm > 0\}}\right), \end{aligned}$$

as $n \rightarrow \infty$, where $\omega^\pm = |\log \lambda^\pm| / \log \mu > 0$.

To bound each of the remaining probabilities, $\mathbb{P}_n(C_{r-1} \cap C_r^c \cap I_k \cap M_{s_n})$, use the union bound followed by Markov's inequality applied conditionally on $\mathcal{F}_{k-1}$, to obtain, for $s_n \leq r \leq k$,

$$\begin{aligned} &\mathbb{P}_n(C_{r-1} \cap C_r^c \cap I_k \cap M_{s_n}) \\ &\leq \mathbb{P}_n\left(\left|\hat{A}_r^\pm \cap (A_r^\pm)^c\right| > \hat{Z}_r^\pm n^{-\kappa}, C_{r-1} \cap I_k \cap M_{s_n}\right) \\ &\quad + \mathbb{P}_n\left(\left|\hat{A}_r^\pm \cap (\hat{A}_r^\pm)^c\right| > \hat{Z}_r^\pm n^{-\kappa}, C_{r-1} \cap I_k \cap M_{s_n}\right) \\ &\leq \mathbb{P}_n\left(\left|\hat{A}_r^\pm \cap (A_r^\pm)^c\right| > u^r n^{-\kappa}, C_{r-1} \cap I_{r-1}\right) \\ &\quad + \mathbb{P}_n\left(\left|A_r^\pm \cap (\hat{A}_r^\pm)^c\right| > u^r n^{-\kappa}, C_{r-1} \cap I_{r-1}\right) \end{aligned}$$

$$\leq \frac{n^\kappa}{u^r} \mathbb{E}_n \left[\mathbb{1}_{C_{r-1} \cap I_{r-1}} \mathbb{E}_n \left[\left| \hat{A}_r^\pm \cap (A_r^\pm)^c \right| + \left| A_r^\pm \cap (\hat{A}_r^\pm)^c \right| \middle| \mathcal{F}_{r-1} \right] \right].$$

It follows from Lemma 10.7 that

$$\mathbb{P}_n(C_{r-1} \cap C_r^c \cap I_k \cap M_{s_n}) \leq K \frac{n^\kappa \mu^r}{u^r} \left(\mu^{\eta r} n^{-\eta(1-2\tau)} + r n^{-\varepsilon} \right),$$

which in turn implies that, as $n \rightarrow \infty$,

$$\begin{aligned} & \sum_{r=s_n}^k \mathbb{P}_n(C_{r-1} \cap C_r^c \cap I_k \cap M_{s_n}) \\ &= O \left(n^{\kappa-\eta(1-2\tau)} \sum_{r=s_n}^k (\mu^{1+\eta}/u)^r + n^{\kappa-\varepsilon} \sum_{r=s_n}^k r (\mu/u)^r \right) \\ &= O \left(n^{\kappa-\eta(1-2\tau)} (\mu^{1+\eta}/u)^k + n^{\kappa-\varepsilon} k (\mu/u)^k \right) \\ &= O \left(n^{\kappa-\eta(1-2\tau)+(1-\delta)(b+\eta)} + n^{\kappa-\varepsilon+(1-\delta)b} \log n \right) \\ &= O \left(n^{(\kappa-\eta\delta)/2+(1-\delta)b} + n^{\kappa-\varepsilon+(1-\delta)b} \log n \right). \end{aligned}$$

Since $(\kappa - \eta\delta)/2 + (1 - \delta)b < (\kappa - \eta\delta)/4 < 0$ and $\kappa - \varepsilon + (1 - \delta)b < (\kappa - \varepsilon)/2 < 0$ all the exponents inside the big-O term are strictly negative, which completes the proof. $\square$

10.3 Distribution of the hopcount

In this section we use the coupling result to analyze the distribution of the hopcount of graphs generated by the directed configuration model from bi-degree sequences that satisfy Assumption 10.1.

10.3.1 The main result

We will show in Theorem 10.9 below, that the distance between two randomly chosen nodes grows logarithmically in the size of the graph, and characterize the spread around the logarithmic term.

In the statement of our results, we use $\mathcal{H}_n$ to denote the hopcount, or distance, between two randomly chosen nodes in a graph G_n . Since the graph is directed, we say that the hopcount between node i and node j is k if there exists a shortest directed path of length k from i to j ; if there is no directed path from i to j we say that the hopcount is infinite. Since the two nodes are chosen at random, we can assume without loss of generality that $\mathcal{H}_n$ is the hopcount from the first node to the second one.

Before we can state Theorem 10.9 we need to introduce the limiting random variables appearing in the characterization of the hopcount. To this end, let $f^\pm$, $f^{+,-}$ and $f^{-,+}$ be the probability mass functions from Assumption 10.1 and let $\{\hat{Z}_k^\pm : k \geq 0\}$, $\hat{Z}_0^\pm = 1$ denote the delayed branching process having offspring distributions $(f^\pm, f^{\mp,\pm})$, where $f^\pm$ is the distribution of $\hat{Z}_1^\pm$ and all other nodes

have offspring according to $f^{\mp, \pm}$. Recall from Section 10.2.5 that under the conditions in Assumption 10.1 $W^\pm = \lim_{k \rightarrow \infty} \hat{Z}_k^\pm / (\nu \mu^{k-1})$ are non-trivial and $\mathbb{E}[W^\pm] = 1$, see Lemma 10.3 for an expression for $\mathbb{P}(W^\pm = 0)$.

We are now ready to state the main result of the chapter; $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x .

Theorem 10.9. *Let the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfy Assumption 10.1 and $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$. Let $\mathcal{H}_n$ denote the hopcount between two randomly chosen nodes in G_n . Then, there exist random variables Υ_n such that for each (fixed) $t \in \mathbb{Z}$,*

$$\lim_{n \rightarrow \infty} |\mathbb{P}(\mathcal{H}_n - \lfloor \log_\mu n \rfloor = t | \mathcal{H}_n < \infty) - \mathbb{P}(\Upsilon_n = t)| = 0, \quad (10.22)$$

where Υ_n has distribution

$$\mathbb{P}(\Upsilon_n \leq x) = 1 - \mathbb{E} \left[\exp \left\{ - \frac{\nu \mu^{\lfloor \log_\mu n \rfloor + \lfloor x \rfloor}}{(\mu - 1)n} W^+ W^- \right\} \middle| W^+ W^- > 0 \right]. \quad (10.23)$$

Theorem 10.9 shows that the directed distance between two randomly chosen nodes in the DCM scales logarithmically in the size of the graph, which is consistent with existing results for the undirected configuration model under the assumption that the degree distribution has finite variance [80]. The interesting difference between the directed and undirected cases lies in the observation that Assumption 10.1 can hold with high probability for degree sequences having infinite variance, as shown in Chapter 3 (see Theorem 3.4). This shows that the distance remains logarithmic even when in its undirected counterpart it becomes of order $\log \log n$ [81]. To explain this, note that distances in the undirected model get smaller as the degree distribution gets heavier (i.e., more variable) because of the appearance of nodes with extremely large degrees, that have the effect of creating shortcuts between nodes in their connected component. In contrast, when the graph is directed, increasing the variability of the out- and in-degree distributions does mean that there will be more nodes with very large out-degrees or very large in-degrees, but this does not necessarily imply the appearance of more shortcuts, e.g., when the out-degree is independent of the in-degree, it is unlikely that a node has both large out-degree and large in-degree. Our results are consistent with the intuition that if the nodes with very large out-degrees are the same as those with very large in-degrees (i.e., positively correlated out- and in-degrees), then more shortcuts will be created and the distance should get smaller.

To complement this theorem we provide below a limit theorem for the probability of the hopcount being finite, which can be expressed in terms of the survival properties of the delayed branching processes $\{\hat{Z}_k^+ : k \geq 1\}$ and $\{\hat{Z}_k^- : k \geq 1\}$.

Proposition 10.10. *Let the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfy Assumption 10.1 and let $G_n = \text{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$. Then,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{H}_n < \infty) = s^+ s^-, \quad (10.24)$$

where $s^\pm = \mathbb{P}(W^\pm > 0)$.

10.3.2 Ideas and heuristic explanation of the proof

Given a directed graph G_n and a node i , a straightforward way of determining the shortest path from i to all other nodes is to start a breadth-first exploration rooted at i . In the first step we visit all outbound neighbors of i , i.e., the nodes to which the edges of i point. These are the nodes that can be reached from i in one hop. Next, we continue by exploring the edges of the nodes discovered in the previous round. These will then be on a path of length 2 from i . Note that since shortest paths do not contain any cycles, we only need to continue the exploration from nodes that we have not seen before. Once we run out of new nodes we stop. All nodes that we have not uncovered through this process will be unreachable from i . Similarly we can determine the shortest path from all nodes to i by doing a breadth-first exploration over the inbound edges.

The idea behind the proof of Theorem 10.9 is to start with two nodes 1 and 2, selected uniformly at random. Then we explore the graph from node 1 over the outbound edges and from node 2 over the inbound edges and use Theorem 10.1 to couple the exploration of the graph from node 1 and 2, with the branching process $\{\hat{Z}_k^+ : k \geq 0\}$ and $\{\hat{Z}_k^- : k \geq 0\}$, respectively. We can then describe the distribution of the hopcount in terms of the coupled branching process and use martingale limits to express the limit of this distribution. The road map of this approach is as follows:

Fix two nodes 1 and 2 and let $\{Z_k^+ : k \geq 0\}$ and $\{Z_k^- : k \geq 0\}$ denote the exploration process of new nodes on the graph, starting from node 1 and 2, respectively, where Z^+ runs over the outbound edges and Z^- over the inbound edges. We will find an expression for the tail probability of the hopcount in terms of $\{Z_k^\pm : k \geq 0\}$. Then we use the coupling to replace these processes with $\{\hat{Z}_k^\pm : k \geq 0\}$, as described in Section 10.3.1. We will show that on the event Ξ_n , see Proposition 10.12,

$$\mathbb{P}_n(\mathcal{H}_n > k) \approx \mathbb{E} \left[\exp \left\{ -\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^- \hat{Z}_{\lfloor i/2 \rfloor}^+ \right\} \right].$$

Next, see Proposition 10.13, we will transition from the $\hat{Z}^\pm$ to their martingale limits $W^\pm$,

$$\mathbb{E} \left[\exp \left\{ -\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^+ \hat{Z}_{\lfloor i/2 \rfloor}^- \right\} \right] \approx \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k}{(\mu - 1)n} W^+ W^- \right\} \right].$$

It will follow from the proofs that these results hold for all $k = \lfloor \log_\mu n \rfloor + t$, for any fixed $t \in \mathbb{Z}$. Therefore, since

$$1 - \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^{\lfloor \log_\mu n \rfloor + t}}{(\mu - 1)n} W^+ W^- \right\} \middle| W^+ W^- > 0 \right]$$

defines a cumulative distribution function and $\mathbb{P}(\Xi_n) \rightarrow 1$, by Assumption 10.1, we have that

$$\mathbb{P}(\mathcal{H}_n > \lfloor \log_\mu n \rfloor + t | \mathcal{H}_n < \infty) \approx \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^{\lfloor \log_\mu n \rfloor + t}}{(\mu - 1)n} W^+ W^- \right\} \middle| W^+ W^- > 0 \right],$$

from which Theorem 10.9 follows.

10.3.3 Proof of the main result

As described in the previous section, we compute the hopcount of a graph by selecting two nodes at random, say 1 and 2, and then start two independent breadth-first exploration processes. One such process follows the outgoing edges of node 1 while the other uses the incoming edges of node 2. At each step we explore one generation of the out-component of node 1 and the corresponding generation of the in-component of node 2, starting with node 1. The hopcount between the two nodes is then the sum of the number of generations explored up until the time the two exploration processes have a node in common. In terms of the two nodes, $\{Z_k^+ : k \geq 1\}$ denotes the number of outbound stubs discovered during the k -th step of exploration of the out-component of node 1, while $\{Z_k^- : k \geq 1\}$ denotes the number of inbound stubs discovered during the k -th step of the exploration of the in-component of node 2. An expression for the distribution of the hopcount is then obtained by computing the probability that there are no nodes in common given the current number of stubs explored so far in each of the two processes.

The first step in the analysis is a recursive relation for $\mathbb{P}_n(\mathcal{H}_n > k)$. For this we denote by $\mathcal{F}^{l,m} = \sigma(Z_i^-, Z_j^+ : 0 \leq i \leq l, 0 \leq j \leq m)$ the sigma algebra generated by the Z_i^- and Z_j^+ of the first l and m generations, respectively. By a straightforward adaption of the analysis done in [80, Lemma 4.1] to the directed setting, we obtain that for all $k \geq 1$,

$$\mathbb{P}_n(\mathcal{H}_n > k) = \mathbb{E}_n \left[\prod_{i=2}^{k+1} \mathbb{P}_n \left(\mathcal{H}_n > i-1 \mid \mathcal{H}_n > i-2, \mathcal{F}^{\lceil i/2 \rceil, \lfloor i/2 \rfloor} \right) \right], \quad (10.25)$$

where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x . The ceiling and floor functions are here because we iteratively advance the exploration process from node 1 and 2, starting with 1.

Note that $\mathcal{H}_n > 1$ holds if and only if none of the Z_1^+ outgoing edges points towards node 2. From the definition of the model this occurs if and only if none of the Z_1^+ outbound stubs have been paired with one of the Z_1^- inbound stubs. Hence,

$$\mathbb{P}_n(\mathcal{H}_n > 1 \mid \mathcal{F}^{1,1}) = p(Z_1^+, Z_1^-, L_n),$$

where $p(A, B, L)$ denote the probability that none of the outbound stubs from a set of size A connect to one of the inbound stubs from a set of size B , given that there are L outbound/inbound stubs left in total. It now follows from Chapter 8, see (8.23), that

$$\mathbb{P}_n(\mathcal{H}_n > 1 \mid \mathcal{F}^{1,1}) = \mathbb{1}_{\{Z_1^+ + Z_1^- \leq L_n\}} \prod_{s=0}^{Z_1^+ - 1} \left(1 - \frac{Z_1^-}{L_n - s} \right).$$

Similarly,

$$\mathbb{P}_n(\mathcal{H}_n > 2 \mid \mathcal{H}_n > 1, \mathcal{F}^{2,1}) = \mathbb{1}_{\{Z_2^+ + Z_1^- \leq L_n - Z_1^+\}} \prod_{s=0}^{Z_2^+ - 1} \left(1 - \frac{Z_1^-}{L_n - Z_1^+ - s} \right).$$

In order to write the full formula we first define $\{\mathcal{S}_k\}_{k \geq 0}$ as follows:

$$\mathcal{S}_0 = 0, \quad \mathcal{S}_1 = Z_1^+, \quad \mathcal{S}_k = \sum_{j=1}^{\lceil k/2 \rceil} Z_j^+ + \sum_{j=1}^{\lfloor k/2 \rfloor} Z_j^- \quad \text{for } k \geq 2. \quad (10.26)$$

We then obtain, for $i \geq 2$,

$$\begin{aligned} \mathbb{P}_n \left(\mathcal{H}_n > i-1 \mid \mathcal{H}_n > i-2, \mathcal{F}^{\lceil i/2 \rceil, \lfloor i/2 \rfloor} \right) \\ = 1(\mathcal{S}_i \leq L_n) \prod_{s=0}^{Z_{\lceil i/2 \rceil}^+ - 1} \left(1 - \frac{Z_{\lfloor i/2 \rfloor}^-}{L_n - \mathcal{S}_{i-2} - s} \right). \end{aligned}$$

Substituting this expression into (10.25) yields

$$\mathbb{P}_n(\mathcal{H}_n > k) = \mathbb{E}_n \left[\mathbb{1}_{\{\mathcal{S}_{k+1} \leq L_n\}} \prod_{i=2}^{k+1} \prod_{s=0}^{Z_{\lceil i/2 \rceil}^+ - 1} \left(1 - \frac{Z_{\lfloor i/2 \rfloor}^-}{L_n - \mathcal{S}_{i-2} - s} \right) \right] \quad (10.27)$$

For the analysis of (10.27) the following technical lemma will be useful:

Lemma 10.11. *For any non-negative $x, x_0 > 0$, $y_i, z_i \geq 0$ with $z_i < x$ for all i , and any $m \geq 1$,*

$$\begin{aligned} -\frac{x_0}{x^2}(x_0 - x)^+ - \frac{x_0}{2} \max_{1 \leq i \leq m} \frac{z_i}{(x - z_i)^2} \\ \leq \prod_{i=1}^m \left(1 - \frac{z_i}{x} \right)^{y_i} - \exp \left\{ -\frac{1}{x_0} \sum_{i=1}^m y_i z_i \right\} \leq \frac{|x - x_0|}{(x \wedge x_0)}. \end{aligned}$$

Proof. For the upper bound note that

$$\begin{aligned} \prod_{i=1}^m \left(1 - \frac{z_i}{x} \right)^{y_i} &= \exp \left\{ \sum_{i=1}^m y_i \log \left(1 - \frac{z_i}{x} \right) \right\} \leq \exp \left\{ -\frac{1}{x} \sum_{i=1}^m y_i z_i \right\} \\ &\leq \exp \left\{ -\frac{1}{x_0} \sum_{i=1}^m y_i z_i \right\} + \frac{|x - x_0|}{x_0 \wedge x}, \end{aligned}$$

where in the second step we have used the inequality $\log(1-t) \leq -t$ for $t \in [0, 1]$ and in the third step we have used the inequality

$$|e^{-S/x} - e^{-S/x_0}| \leq \frac{S}{\xi^2} e^{-S/\xi} |x - x_0| \leq \sup_{t \geq 0} t e^{-t} \cdot \frac{|x - x_0|}{\xi} \leq \frac{|x - x_0|}{(x \wedge x_0)}$$

for any $S, x, x_0 \geq 0$ and some ξ between x and x_0 .

Similarly, using the first order Taylor expansion for $\log(1-t)$ we obtain

$$\log(1 - c/x) = -\frac{c}{x} - \frac{c^2}{2x^2(1 - \xi')^2} = -\frac{c}{x_0} + \frac{c}{(\xi'')^2}(x - x_0) - \frac{c^2}{2x^2(1 - \xi')^2}$$

for any $c < x$, $0 < \xi' < c/x$, and ξ'' between x and x_0 , which in turn yields the inequality

$$\log(1 - c/x) \geq -\frac{c}{x_0} - \frac{c}{x^2}(x_0 - x)^+ - \frac{c^2}{2(x - c)^2}.$$

It follows that

$$\begin{aligned} \prod_{i=1}^m \left(1 - \frac{z_i}{x}\right)^{y_i} &\geq \exp \left\{ -\sum_{i=1}^m \left(\frac{y_i z_i}{x_0} + \frac{y_i z_i}{x^2} (x_0 - x)^+ + \frac{y_i z_i^2}{2(x - z_i)^2} \right) \right\} \\ &\geq \exp \left\{ -\frac{1}{x_0} \sum_{i=1}^m y_i z_i \right\} \\ &\quad - \exp \left\{ -\frac{1}{x_0} \sum_{i=1}^m y_i z_i \right\} \sum_{i=1}^m \left(\frac{y_i z_i}{x^2} (x_0 - x)^+ + \frac{y_i z_i^2}{2(x - z_i)^2} \right) \\ &\geq \exp \left\{ -\frac{1}{x_0} \sum_{i=1}^m y_i z_i \right\} - \frac{x_0}{x^2} (x_0 - x)^+ - \frac{x_0}{2} \max_{1 \leq i \leq m} \frac{z_i}{(x - z_i)^2}. \end{aligned}$$

□

The first result for the hopcount uses (10.27) combined with the coupling result, see Corollary 10.2, to obtain an expression in terms of the branching processes $\{\hat{Z}_k^+ : k \geq 1\}$ and $\{\hat{Z}_k^- : k \geq 1\}$.

Proposition 10.12. *Let $G_n = \mathbf{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$. Then, for any $0 < \delta < 1$ and for any $0 \leq k \leq 2(1 - \delta) \log_\mu n$, there exists a constant $a > 0$ such that, on the event Ξ_n ,*

$$\left| \mathbb{P}_n(\mathcal{H}_n > k) - \mathbb{E} \left[\exp \left\{ -\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^- \hat{Z}_{\lfloor i/2 \rfloor}^+ \right\} \right] \right| = O(n^{-a}),$$

as $n \rightarrow \infty$, where $\{\hat{Z}_i^+ : i \geq 1\}$ and $\{\hat{Z}_i^- : i \geq 1\}$ are independent delayed branching processes having offspring distributions $(f^+, f^{-,+})$ and $(f^-, f^{+,-})$, respectively.

Proof. Let $0 < \kappa < \min\{\eta\delta, \varepsilon\}$, and construct the pairs of processes $\{(Z_i^+, \hat{Z}_i^+) : i \geq 0\}$ and $\{(Z_i^-, \hat{Z}_i^-) : i \geq 0\}$ according to the coupling described in Section 10.2.3. Now define the events

$$\begin{aligned} \mathcal{C}_m^\pm &= \left\{ \hat{Z}_m^\pm (1 - n^{-\kappa}) \leq Z_m^\pm \leq \hat{Z}_m^\pm (1 + n^{-\kappa}) \right\} \quad \text{and} \\ \mathcal{C}_k &= \bigcap_{m=1}^{\lceil k/2 \rceil + 1} (\mathcal{C}_m^+ \cap \mathcal{C}_m^-), \end{aligned}$$

and note that since $\{(Z_i^-, \hat{Z}_i^+) : i \geq 0\}$ and $\{(Z_i^+, \hat{Z}_i^-) : i \geq 0\}$ are independent, Corollary 10.2 gives $\mathbb{P}_n(\mathcal{C}_k^c) = O(n^{-a_1})$ as $n \rightarrow \infty$ for some $a_1 > 0$.

Next, let $S_k = \sum_{i=2}^{k+1} Z_{\lceil i/2 \rceil}^+ Z_{\lfloor i/2 \rfloor}^-$ and use the triangle's inequality to get

$$\begin{aligned} & \left| \mathbb{P}_n(\mathcal{H}_n > k) - \mathbb{E} \left[\exp \left\{ -\frac{\sum_{i=2}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^-}{\nu n} \right\} \right] \right| \\ & \leq \left| \mathbb{P}_n(\mathcal{H}_n > k) - \mathbb{E}_n \left[\exp \left\{ -\frac{S_k}{\nu n} \right\} \right] \right| \end{aligned} \quad (10.28)$$

$$+ \left| \mathbb{E}_n \left[\exp \left\{ -\frac{S_k}{\nu n} \right\} \right] - \mathbb{E} \left[\exp \left\{ -\frac{\sum_{i=2}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^-}{\nu n} \right\} \right] \right|. \quad (10.29)$$

We start by bounding (10.29), for which we use the independence of $\{\hat{Z}_i^+\}$ and $\{\hat{Z}_i^-\}$ from the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$, and the inequality $|e^{-x} - e^{-y}| \leq e^{-(x \wedge y)}|x - y|$ for $x, y \geq 0$ to obtain

$$\begin{aligned} & \left| \mathbb{E}_n \left[\exp \left\{ -\frac{S_k}{\nu n} \right\} \right] - \mathbb{E} \left[\exp \left\{ -\frac{\sum_{i=2}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^-}{\nu n} \right\} \right] \right| \\ & \leq \mathbb{E}_n \left[\left| \exp \left\{ -\frac{S_k}{\nu n} \right\} - \exp \left\{ -\frac{\sum_{i=2}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^-}{\nu n} \right\} \right| \right] \\ & \leq \frac{1}{n\nu} \mathbb{E}_n \left[\mathbb{1}_{\mathcal{C}_k} \exp \left\{ -\frac{S_k(1 - n^{-\kappa})^2}{\nu n} \right\} \left| S_k - \sum_{i=2}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- \right| \right] + \mathbb{P}_n(\mathcal{C}_k^c). \end{aligned}$$

On the event $\mathcal{C}_k$ and for all $2 \leq i \leq k+1$,

$$(1 - 2n^{-\kappa} + n^{-2\kappa}) \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- \leq Z_{[i/2]}^+ Z_{[i/2]}^- \leq (1 + 2n^{-\kappa} + n^{-2\kappa}) \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^-,$$

so that $n \geq 1$,

$$\left| Z_{[i/2]}^+ Z_{[i/2]}^- - \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- \right| \leq 3n^{-\kappa} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- \leq \frac{3n^{-\kappa}}{(1 - n^{-\kappa})^2} Z_{[i/2]}^+ Z_{[i/2]}^-.$$

It follows that (10.29) is bounded from above by

$$\begin{aligned} & \frac{3n^{-\kappa}}{(1 - n^{-\kappa})^2} \mathbb{E}_n \left[\exp \left\{ -\frac{S_k(1 - n^{-\kappa})^2}{\nu n} \right\} \frac{1}{n\nu} S_k \right] + \mathbb{P}_n(\mathcal{C}_k^c) \\ & \leq \frac{3e^{-1}n^{-\kappa}}{(1 - n^{-\kappa})^4} + \mathbb{P}_n(\mathcal{C}_k^c), \end{aligned}$$

where we have used the observation that $\sup_{x \geq 0} e^{-x}x = e^{-1}$.

We now proceed to bound (10.28). From (10.27), we have that

$$\mathbb{P}_n(\mathcal{H}_n > k) = \mathbb{E}_n \left[\mathbb{1}_{\{\mathcal{S}_{k+1} \leq L_n\}} \prod_{i=2}^{k+1} \prod_{s=0}^{Z_{[i/2]}^+ - 1} \left(1 - \frac{Z_{[i/2]}^-}{L_n - \mathcal{S}_{i-2} - s} \right) \right],$$

where $\mathcal{S}_k$ is as in (10.26). Recall that on the event Ξ_n we have $|L_n - n\nu| \leq n^{1-\varepsilon}$, and therefore $\{\mathcal{S}_{k+1} \leq n^b\} \subseteq \{\mathcal{S}_{k+1} \leq L_n\}$ for any $1 - \delta < b < 1$ and all

sufficiently large n . Now note that

$$\begin{aligned} \mathbb{E}_n \left[\mathbb{1}_{\{\mathcal{S}_{k+1} \leq n^b\}} \prod_{i=2}^{k+1} \left(1 - \frac{Z_{\lfloor i/2 \rfloor}^-}{L_n - \mathcal{S}_{k+1}} \right)^{Z_{\lceil i/2 \rceil}^+} \right] \\ \leq \mathbb{P}_n(\mathcal{H}_n > k) \leq \mathbb{E}_n \left[\prod_{i=2}^{k+1} \left(1 - \frac{Z_{\lfloor i/2 \rfloor}^-}{L_n} \right)^{Z_{\lceil i/2 \rceil}^+} \right]. \end{aligned}$$

Using Lemma 10.11 with $x = L_n$ and $x_0 = \nu n$ gives

$$\mathbb{P}_n(\mathcal{H}_n > k) \leq \mathbb{E}_n \left[e^{-S_k/(\nu n)} \right] + \frac{|L_n - \nu n|}{L_n \wedge (\nu n)} = \mathbb{E}_n \left[e^{-S_k/(\nu n)} \right] + O(n^{-\varepsilon}).$$

Similarly, using Lemma 10.11 with $x = L_n - \mathcal{S}_{k+1}$ and $x_0 = \nu n$ we obtain

$$\begin{aligned} \mathbb{P}_n(\mathcal{H}_n > k) &\geq \mathbb{E}_n \left[e^{-S_k/(\nu n)} \mathbb{1}_{\{\mathcal{S}_{k+1} \leq n^b\}} \right] \\ &\quad - \mathbb{E}_n \left[\mathbb{1}_{\{\mathcal{S}_{k+1} \leq n^b\}} \frac{\nu n}{(L_n - \mathcal{S}_{k+1})^2} (\nu n - L_n + \mathcal{S}_{k+1})^+ \right] \\ &\quad - \mathbb{E}_n \left[\mathbb{1}_{\{\mathcal{S}_{k+1} \leq n^b\}} \frac{\nu n}{2} \max_{2 \leq i \leq k+1} \frac{Z_{\lfloor i/2 \rfloor}^-}{(L_n - \mathcal{S}_{k+1} - Z_{\lfloor i/2 \rfloor}^-)^2} \right] \\ &\geq \mathbb{E}_n \left[e^{-S_k/(\nu n)} \right] - \mathbb{P}_n(\mathcal{S}_{k+1} > n^b) - \frac{(n^{1-\varepsilon} + n^b)^+}{(L_n - n^b)^2} \\ &\quad - \frac{\nu n^{1+b}}{2(L_n - 2n^b)^2} \\ &= \mathbb{E}_n \left[e^{-S_k/(\nu n)} \right] - \mathbb{P}_n(\mathcal{S}_{k+1} > n^b) - O(n^{-(1-b)}). \end{aligned}$$

Finally, note that using Markov's inequality we obtain

$$\begin{aligned} \mathbb{P}_n(\mathcal{S}_{k+1} > n^b) &\leq \mathbb{P}_n(\mathcal{C}_k^c) + \mathbb{P}_n(\mathcal{S}_{k+1} > n^b, \mathcal{C}_k) \\ &\leq \mathbb{P}_n(\mathcal{C}_k^c) + \mathbb{P} \left(\sum_{j=1}^{\lfloor k/2 \rfloor} (\hat{Z}_j^+ + \hat{Z}_j^-)(1 + n^{-\kappa}) > n^b \right) \\ &\leq \mathbb{P}_n(\mathcal{C}_k^c) + \frac{1 + n^{-\kappa}}{n^b} \sum_{j=1}^{\lfloor k/2 \rfloor} \mathbb{E} \left[\hat{Z}_j^+ + \hat{Z}_j^- \right] = O(n^{-a_1} + \mu^{\lfloor k/2 \rfloor} n^{-b}), \end{aligned}$$

where in the last step we have used the observation that

$$\sum_{j=1}^{\lfloor k/2 \rfloor} \mathbb{E} \left[\hat{Z}_j^+ + \hat{Z}_j^- \right] = 2 \sum_{j=1}^{\lfloor k/2 \rfloor} \nu \mu^{j-1} = O(\mu^{\lfloor k/2 \rfloor}).$$

Since $k \leq 2(1 - \delta) \log n / (\log \mu)$, we have that $\mu^{\lfloor k/2 \rfloor} = O(n^{1-\delta})$ and the result follows, with $a = \min\{a_1, b + \delta - 1\} > 0$. $\square$

The next result shows a simplified expression for the limit in Proposition 10.12 in terms of the martingale limits W^+ and W^- . This result is independent of the coupling, and follows from the properties of the (delayed) branching processes $\{\hat{Z}_k^+ : k \geq 1\}$ and $\{\hat{Z}_k^- : k \geq 1\}$. We state it here since it plays an important role in establishing both Theorem 10.9 and Proposition 10.10.

Proposition 10.13. *Suppose $\{\hat{Z}_i^+ : i \geq 1\}$ and $\{\hat{Z}_i^- : i \geq 1\}$ are independent delayed branching processes having offspring distributions $(f^+, f^{-,+})$ and $(f^-, f^{+,-})$, respectively. Suppose that $f^{-,+}$ and $f^{+,-}$ have finite moments of order $1 + \eta \in (1, 2]$ with common mean $\mu > 1$, and f^+, f^- have common mean ν . Then, there exists $b > 0$ such that, as $n \rightarrow \infty$,*

$$\left| \mathbb{E} \left[\exp \left\{ -\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^+ \hat{Z}_{\lfloor i/2 \rfloor}^- \right\} \right] - \exp \left\{ -\frac{\nu \mu^k}{(\mu - 1)n} W^+ W^- \right\} \right| = O(n^{-b}),$$

uniformly for all $k \in \mathbb{N}_+$, where $W^\pm = \lim_{k \rightarrow \infty} \hat{Z}_k^\pm / (\nu \mu^{k-1})$.

Proof. Fix $0 < \varepsilon < \eta/(2 + 2\eta)$ and set $m_n = \lfloor (1 - \varepsilon) \log n / \log \mu \rfloor$. We start by noting that for $1 \leq k \leq m_n$ the inequality $|e^{-x} - e^{-y}| \leq |x - y|$ for $x, y \geq 0$ gives

$$\begin{aligned} & \left| \mathbb{E} \left[\exp \left\{ -\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^+ \hat{Z}_{\lfloor i/2 \rfloor}^- \right\} \right] - \exp \left\{ -\frac{\nu \mu^k}{(\mu - 1)n} W^+ W^- \right\} \right| \\ & \leq \mathbb{E} \left[\left| \frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^+ \hat{Z}_{\lfloor i/2 \rfloor}^- - \frac{\nu \mu^k}{(\mu - 1)n} W^+ W^- \right| \right] \\ & \leq \frac{1}{\nu n} \sum_{i=2}^{k+1} \mathbb{E} [\hat{Z}_{\lceil i/2 \rceil}^+] \mathbb{E} [\hat{Z}_{\lfloor i/2 \rfloor}^-] + \frac{\nu \mu^k}{(\mu - 1)n} \mathbb{E} [W^+] \mathbb{E} [W^-] \\ & = \frac{1}{\nu n} \sum_{i=2}^{k+1} \nu^2 \mu^{\lceil i/2 \rceil + \lfloor i/2 \rfloor - 2} + \frac{\nu \mu^k}{(\mu - 1)n} \mathbb{E} [W^+] \mathbb{E} [W^-] \\ & = \frac{\nu}{\mu^2 n} \left(\sum_{i=2}^{k+1} \mu^i + \frac{\mu^{k+2}}{\mu - 1} \right) \leq \frac{2\nu \mu^k}{n(\mu - 1)} \leq \frac{2\nu \mu^{m_n}}{n(\mu - 1)} = O(n^{-\varepsilon}) \end{aligned}$$

as $n \rightarrow \infty$, where in the second equality we have used the observation that $\lceil i/2 \rceil + \lfloor i/2 \rfloor = i$ for all $i \in \mathbb{N}$, and $\mathbb{E} [W^+] = \mathbb{E} [W^-] = 1$ (since f^+ and f^- have finite moments of order $1 + \eta$). It remains to consider the case $k > m_n$.

Suppose from now on that $k > m_n$ and note that

$$\frac{\mu^k}{\mu - 1} = \sum_{i=m_n}^{k+1} \mu^{i-2} + \frac{\mu^{m_n-2}}{\mu - 1},$$

and therefore, using $|e^{-x} - e^{-y}| \leq |x - y|$ for $x, y \geq 0$ and the triangle inequality we obtain

$$\left| \mathbb{E} \left[\exp \left\{ -\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- \right\} - \exp \left\{ -\frac{\nu \mu^k}{(\mu-1)n} W^+ W^- \right\} \right] \right|$$

$$\leq \mathbb{E} \left[\left| \exp \left\{ -\frac{1}{\nu n} \sum_{i=m_n}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- \right\} - \exp \left\{ -\frac{\nu}{n} \sum_{i=m_n}^{k+1} \mu^{i-2} W^+ W^- \right\} \right| \right] \quad (10.30)$$

$$+ \mathbb{E} \left[\frac{1}{\nu n} \sum_{i=2}^{m_n-1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- + \frac{\nu \mu^{m_n-2}}{(\mu-1)n} W^+ W^- \right], \quad (10.31)$$

where (10.31) is of order $O(n^{-\varepsilon})$ as shown above. To bound (10.30), let $W_k^\pm = \hat{Z}_k^\pm / (\nu \mu^{k-1})$, set $\kappa = \varepsilon/2$, and define the events

$$\mathcal{C}_r^\pm = \left\{ \max_{\lfloor m_n/2 \rfloor \leq j \leq r} \mathbb{1}_{\{W_j^\pm > 0\}} |W_j^\pm - W^\pm| / W_j^\pm \leq n^{-\kappa} \right\}, \quad \text{for } r \geq \lfloor m_n/2 \rfloor,$$

$$\mathcal{C}_k = \mathcal{C}_{\lfloor (k+1)/2 \rfloor}^+ \cap \mathcal{C}_{\lceil (k+1)/2 \rceil}^- \quad \text{and}$$

$$\mathcal{B} = \{W^+ W^- > 0\}.$$

The idea is to split up (10.30) over the events $\mathcal{B} \cap \mathcal{C}_k$, $\mathcal{B} \cap \mathcal{C}_k^c$ and $\mathcal{B}^c$. We will analyze each term separately.

On the event $\mathcal{B} \cap \mathcal{C}_k$

Use the inequality $|e^{-y} - e^{-x}| \leq e^{-(x \wedge y)} |x - y|$ for $x, y \geq 0$ to obtain

$$\mathbb{E} \left[\mathbb{1}_{\mathcal{B} \cap \mathcal{C}_k} \left| \exp \left\{ -\frac{1}{\nu n} \sum_{i=m_n}^{k+1} \hat{Z}_{[i/2]}^+ \hat{Z}_{[i/2]}^- \right\} - \exp \left\{ -\frac{\nu}{n} \sum_{i=m_n}^{k+1} \mu^{i-2} W^+ W^- \right\} \right| \right]$$

$$\leq \mathbb{E} \left[\mathbb{1}_{\mathcal{B} \cap \mathcal{C}_k} \exp \left\{ -\frac{\nu}{\mu^2 n} \sum_{i=m_n}^{k+1} \mu^i W_{[i/2]}^+ W_{[i/2]}^- (1 - n^{-\kappa}) \right\} \right]$$

$$\times \left| \frac{\nu}{\mu^2 n} \sum_{i=m_n}^{k+1} \mu^i (W_{[i/2]}^+ W_{[i/2]}^- - W^+ W^-) \right|$$

$$\leq \frac{3\nu n^{-\kappa}}{\mu^2 n} \mathbb{E} \left[\exp \left\{ -\frac{\nu}{\mu^2 n} \sum_{i=m_n}^{k+1} \mu^i W_{[i/2]}^+ W_{[i/2]}^- (1 - n^{-\kappa}) \right\} \sum_{i=m_n}^{k+1} \mu^i W_{[i/2]}^+ W_{[i/2]}^- \right]$$

$$\leq \frac{3n^{-\kappa}}{\mu^2 (1 - n^{-\kappa})} \sup_{x>0} e^{-x} x \leq \frac{3n^{-\kappa}}{\mu^2 (1 - n^{-\kappa})},$$

where we have used the observation that on the event $\mathcal{B} \cap \mathcal{C}_k$ we have

$$\left| W_{[i/2]}^+ W_{[i/2]}^- - W^+ W^- \right| \leq \left| W_{[i/2]}^+ - W^+ \right| W_{[i/2]}^- + W^+ \left| W_{[i/2]}^- - W^- \right|$$

$$\leq W_{[i/2]}^+ W_{[i/2]}^- n^{-\kappa} + (1 + n^{-\kappa}) n^{-\kappa} W_{[i/2]}^+ W_{[i/2]}^-$$

$$\leq 3n^{-\kappa} W_{[i/2]}^+ W_{[i/2]}^-,$$

and that $\sup_{x \geq 0} x e^{-x} = e^{-1}$.

On the event $\mathcal{B} \cap \mathcal{C}_k^c$

Now, using that $|e^{-x} - e^{-y}| \leq 1$ for $x, y \geq 0$ we obtain

$$\begin{aligned} & \mathbb{E} \left[\mathbb{1}_{\mathcal{B} \cap \mathcal{C}_k^c} \left| \exp \left\{ -\frac{1}{\nu n} \sum_{i=m_n}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^+ \hat{Z}_{\lfloor i/2 \rfloor}^- \right\} - \exp \left\{ -\frac{\nu}{\mu^2 n} \sum_{i=m_n}^{k+1} \mu^i W^+ W^- \right\} \right| \right] \\ & \leq \mathbb{P} \left(W^+ > 0, (\mathcal{C}_{\lceil (k+1)/2 \rceil}^+)^c \right) + \mathbb{P} \left(W^- > 0, (\mathcal{C}_{\lfloor (k+1)/2 \rfloor}^-)^c \right). \end{aligned}$$

To bound the last two probabilities set $u = \mu^{\varepsilon(2+\eta)/(\eta-\eta\varepsilon)} \in (1, \mu)$, define the event $J^\pm = \{\inf_{j \geq \lfloor m_n/2 \rfloor} \hat{Z}_j^\pm / u^j \geq 1\}$, and note that for any $r \geq \lfloor m_n/2 \rfloor$,

$$\begin{aligned} \mathbb{P}(W^\pm > 0, (\mathcal{C}_r^\pm)^c) & \leq \mathbb{P}(W^\pm > 0, (\mathcal{C}_r^\pm)^c \cap J^\pm) + \mathbb{P}(W^\pm > 0, (J^\pm)^c) \\ & \leq \sum_{j=\lfloor m_n/2 \rfloor}^r \mathbb{P}(|W_j^\pm - W^\pm| > n^{-\eta} W_j^\pm, J^\pm) \\ & \quad + \mathbb{P}(W^\pm > 0, (J^\pm)^c). \end{aligned}$$

By Lemma 10.4

$$\mathbb{P}(W^\pm > 0, (J^\pm)^c) \leq K \left(u^{-\eta \lfloor m_n/2 \rfloor} + (u/\mu)^{\omega^\pm \lfloor m_n/2 \rfloor} \mathbb{1}_{\{q^\pm > 0\}} \right),$$

for some constant $K < \infty$, where $\lambda^\pm = \sum_{i=1}^\infty i f^\pm(i) (q^\pm)^{i-1} \in [0, 1)$ and $\omega^\pm = |\log \lambda^\pm| / \log \mu > 0$ if $q^\pm > 0$; and for the remaining probabilities we use the representation (10.13) for $W^\pm - W_j^\pm$ and Lemma 3.5, applied conditionally on $\hat{Z}_j^\pm$, to obtain

$$\begin{aligned} & \mathbb{P}(|W_j^\pm - W^\pm| > n^{-\eta} W_j^\pm, J^\pm) \\ & \leq \mathbb{P} \left(\left| \sum_{i \in \hat{A}_j^\pm} (\mathcal{W}_i^\pm - 1) \right| > n^{-\eta} \hat{Z}_j^\pm, \hat{Z}_j^\pm \geq u^j \right) \\ & \leq \mathbb{E} \left[\mathbb{1}_{\{\hat{Z}_j^\pm \geq u^j\}} \mathbb{P} \left(\left| \sum_{i \in \hat{A}_j^\pm} (\mathcal{W}_i^\pm - 1) \right| > n^{-\eta} \hat{Z}_j^\pm \middle| \hat{Z}_j^\pm \right) \right] \\ & \leq Q_{1+\eta} \mathbb{E}[|\mathcal{W}^\pm - 1|^{1+\eta}] \mathbb{E} \left[1(\hat{Z}_j^\pm \geq u^j) \frac{\hat{Z}_j^\pm}{(\hat{Z}_j^\pm n^{-\eta})^{1+\eta}} \right] \\ & \leq Q_{1+\eta} \mathbb{E}[|\mathcal{W}^\pm - 1|^{1+\eta}] \frac{n^{\eta(1+\eta)}}{u^{\eta j}}, \end{aligned}$$

where $\{\mathcal{W}_i^\pm\}$ are i.i.d. random variables having the same distribution as $\mathcal{W}^\pm = \lim_{r \rightarrow \infty} \mathcal{Z}_r^\pm / \mu^r$, and $Q_{1+\eta}$ is a finite constant that depends only on η . It follows from our choice of u and η that, as $n \rightarrow \infty$,

$$\begin{aligned} & \mathbb{P}(W^\pm > 0, (\mathcal{C}_r^\pm)^c) \\ & = O \left(\sum_{j=\lfloor m/2 \rfloor}^r \frac{n^{\eta(1+\eta)}}{u^{\eta j}} + u^{-\eta \lfloor m/2 \rfloor} + (u/\mu)^{\omega^\pm \lfloor m/2 \rfloor} \mathbb{1}_{\{q^\pm > 0\}} \right) \end{aligned}$$

$$\begin{aligned}
&= O\left(n^{\eta(1+\eta)}u^{-\eta\lfloor m/2\rfloor} + (u/\mu)^{\omega^\pm\lfloor m/2\rfloor}\mathbb{1}_{\{q^\pm>0\}}\right) \\
&= O\left(n^{-\eta} + n^{-\omega^\pm(1/2-\varepsilon(1+\eta)/\eta)}\mathbb{1}_{\{q^\pm>0\}}\right).
\end{aligned}$$

Hence, as $n \rightarrow \infty$,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\mathcal{B}}\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1}\hat{Z}_{\lceil i/2\rceil}^+\hat{Z}_{\lfloor i/2\rfloor}^-\right\}-\exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1}\mu^iW^+W^-\right\}\right|\right] \\
&= O\left(n^{-\eta} + n^{-\omega^+(1/2-\varepsilon(1+\eta)/\eta)}\mathbb{1}_{\{q^+>0\}} + n^{-\omega^-(1/2-\varepsilon(1+\eta)/\eta)}\mathbb{1}_{\{q^->0\}}\right).
\end{aligned}$$

On the event $\mathcal{B}^c$

Here we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\mathcal{B}^c}\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1}\hat{Z}_{\lceil i/2\rceil}^+\hat{Z}_{\lfloor i/2\rfloor}^-\right\}-\exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1}\mu^iW^-W^+\right\}\right|\right] \quad (10.32) \\
&= \mathbb{E}\left[\mathbb{1}_{\mathcal{B}^c\cap\{\hat{Z}_{\lceil m_n/2\rceil}^+\hat{Z}_{\lfloor m_n/2\rfloor}^->0\}}\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1}\hat{Z}_{\lceil i/2\rceil}^+\hat{Z}_{\lfloor i/2\rfloor}^-\right\}-1\right|\right] \\
&\leq \mathbb{P}\left(\mathcal{B}^c, \hat{Z}_{\lceil m_n/2\rceil}^+\hat{Z}_{\lfloor m_n/2\rfloor}^->0\right) \\
&\leq \mathbb{P}\left(W^+=0, \hat{Z}_{\lceil m_n/2\rceil}^+>0\right) + \mathbb{P}\left(W^-=0, \hat{Z}_{\lfloor m_n/2\rfloor}^->0\right).
\end{aligned}$$

By Lemma 10.3, conditionally on $W^\pm = 0$, $\{\hat{Z}_k^\pm : k \geq 1\}$ is a delayed branching process with offspring distributions $(\tilde{g}^\pm, \tilde{f}^\pm)$, as defined in the lemma, having means $\nu^\pm$ and $\lambda^\pm < 1$, respectively. Moreover, by [4, Theorem 4], $W^\pm = 0$ implies that either $q^\pm > 0$ or $\hat{Z}_1^\pm = 0$. Therefore, from Markov's inequality we obtain, as $n \rightarrow \infty$,

$$\begin{aligned}
&\mathbb{P}\left(W^\pm = 0, \hat{Z}_{\lceil m_n/2\rceil}^\pm > 0\right) \\
&\leq \mathbb{E}\left[\hat{Z}_{\lceil m_n/2\rceil}^\pm \mid W^\pm = 0\right]\mathbb{1}_{\{q^\pm>0\}} \\
&= \nu^\pm(\lambda^\pm)^{\lfloor m_n/2\rfloor-1}\mathbb{1}_{\{q^\pm>0\}} \\
&= O\left(n^{-(1-\varepsilon)|\log \lambda^\pm|/\log \mu}\mathbb{1}_{\{q^\pm>0\}}\right) \\
&= O\left(n^{-(1-\varepsilon)\omega^\pm}\mathbb{1}_{\{q^\pm>0\}}\right).
\end{aligned}$$

We therefore conclude that as $n \rightarrow \infty$,

$$\begin{aligned}
&\mathbb{E}\left[\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1}\hat{Z}_{\lceil i/2\rceil}^+\hat{Z}_{\lfloor i/2\rfloor}^-\right\}-\exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1}\mu^iW^+W^-\right\}\right|\right] \\
&= O\left(n^{-\varepsilon/2} + n^{-\omega^+(1/2-\varepsilon(1+\eta)/\eta)}\mathbb{1}_{\{q^+>0\}} + n^{-\omega^-(1/2-\varepsilon(1+\eta)/\eta)}\mathbb{1}_{\{q^->0\}}\right).
\end{aligned}$$

□

Combining Propositions 10.12 and 10.13, we immediately obtain the following result, which expresses the tail distribution of the hopcount in terms of limiting random variables related to the branching processes $\{\hat{Z}_k^\pm : k \geq 1\}$ introduced in Section 10.3.1.

Theorem 10.14. *Let the bi-degree sequence $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfy Assumption 10.1 and $G_n = \mathbf{CM}(\mathbf{D}_n^+, \mathbf{D}_n^-)$. Then, for any $0 < \delta < 1$ and for any $0 \leq k \leq 2(1 - \delta) \log_\mu n$, there exists a constant $c > 0$ such that, on the event Ξ_n ,*

$$\left| \mathbb{P}_n(\mathcal{H}_n > k) - \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k}{(\mu - 1)n} W^- W^+ \right\} \right] \right| = O(n^{-c}),$$

as $n \rightarrow \infty$, where $W^\pm = \lim_{k \rightarrow \infty} \hat{Z}_k^\pm / (\nu \mu^{k-1})$, with W^+ and W^- independent of each other.

As a corollary of Theorem 10.14 we obtain the following result for the probability that there exists a directed path between two randomly chosen nodes, which implies Proposition 10.10:

Corollary 10.15. *Suppose $(\mathbf{N}_n, \mathbf{D}_n)$ satisfies Assumption 10.1. Then, there exists a constant $c > 0$ such that, as $n \rightarrow \infty$,*

$$|\mathbb{P}_n(\mathcal{H}_n < \infty) - s^+ s^-| = O(n^{-c}),$$

where $s^\pm = \mathbb{P}(W^\pm > 0)$.

Proof of Corollary 10.15. Let $0 < \delta < 1/4$ and set $\omega_n = 2(1 - \delta) \log n / (\log \mu)$. Our analysis is based on splitting the probability that the hopcount is finite into two terms:

$$\mathbb{P}_n(\mathcal{H}_n < \infty) = \mathbb{P}_n(\mathcal{H}_n \leq \omega_n) + \mathbb{P}_n(\omega_n < \mathcal{H}_n < \infty),$$

where for the first probability we will use the approximation provided by Theorem 10.14. Intuitively, the second term corresponds to an event that should be negligible in the limit, since it is unlikely that if there exists a directed path between the two randomly chosen nodes it will not have been discovered after ω_n steps of the exploration process.

First, let $\{\mathcal{Z}_k^+ : k \geq 0\}$ and $\{\mathcal{Z}_k^- : k \geq 0\}$ denote two independent non-delayed branching processes with, respectively, offspring distribution $f^{-,+}$ and $f^{+,-}$ and note that from Lemma 10.3 we have

$$s^\pm = 1 - \sum_{t=0}^{\infty} f^\pm(t) (q^\pm)^t \quad \text{with} \quad q^\pm = \mathbb{P}(\mathcal{Z}_k^\pm = 0 \text{ for some } k \geq 1),$$

and since W^+ and W^- are independent, $s^+ s^- = \mathbb{P}(W^+ W^- > 0)$. As in the previous proof, let $\mathcal{B} = \{W^+ W^- > 0\}$ and write

$$\mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k}{(\mu - 1)n} W^+ W^- \right\} \right] = \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k}{(\mu - 1)n} W^+ W^- \right\} \mathbb{1}_{\mathcal{B}} \right] + \mathbb{P}(\mathcal{B}^c).$$

Next, use the triangle's inequality followed by an application of Theorem 10.14 to obtain

$$|\mathbb{P}_n(\mathcal{H}_n < \infty) - s^+ s^-| = |\mathbb{P}_n(\mathcal{H}_n \leq \omega_n) + \mathbb{P}_n(\omega_n < \mathcal{H}_n < \infty) - \mathbb{P}(\mathcal{B})|$$

$$\begin{aligned}
& \leq |\mathbb{P}_n(\mathcal{H}_n \leq \omega_n) - \mathbb{P}(\mathcal{B})| + \mathbb{P}_n(\omega_n < \mathcal{H}_n < \infty) \\
& \leq \left| \mathbb{P}_n(\mathcal{H}_n \leq \omega_n) - 1 + \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^{\omega_n}}{(\mu-1)n} W^+ W^- \right\} \right] \right| \\
& \quad + \left| 1 - \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^{\omega_n}}{(\mu-1)n} W^+ W^- \right\} \right] - \mathbb{P}(\mathcal{B}) \right| \\
& \quad + \mathbb{P}_n(\omega_n < \mathcal{H}_n < \infty) \\
& \leq \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^{\omega_n}}{(\mu-1)n} W^+ W^- \right\} 1(\mathcal{B}) \right] \\
& \quad + \mathbb{P}_n(\omega_n < \mathcal{H}_n < \infty) + O(n^{-c_1})
\end{aligned}$$

for some $c_1 > 0$, as $n \rightarrow \infty$.

To analyze $\mathbb{P}_n(\omega_n < \mathcal{H}_n < \infty)$ use the expression in (10.27) to see that for any $k \geq 0$,

$$\begin{aligned}
& \mathbb{P}_n(k < \mathcal{H}_n < \infty) \\
& \leq \mathbb{E}_n \left[\mathbb{1}_{\{Z_{\lceil (k+1)/2 \rceil}^+ Z_{\lfloor (k+1)/2 \rfloor}^- > 0\}} \prod_{i=2}^{k+1} \left(1 - \frac{Z_{\lfloor i/2 \rfloor}^-}{L_n} \right)^{Z_{\lceil i/2 \rceil}^+} \right]. \quad (10.33)
\end{aligned}$$

Note that the same steps in the proofs of Propositions 10.12 and 10.13 give that (10.33) is equal to

$$\begin{aligned}
& \mathbb{E} \left[\mathbb{1}_{\{\hat{Z}_{\lceil (k+1)/2 \rceil}^+ \hat{Z}_{\lfloor (k+1)/2 \rfloor}^- > 0\}} \exp \left\{ -\frac{\nu \mu^k}{(\mu-1)n} W^+ W^- \right\} \right] + O(n^{-a} + n^{-b}) \\
& \leq \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k}{(\mu-1)n} W^- W^+ \right\} \mathbb{1}_{\mathcal{B}} \right] + \mathbb{P} \left(\hat{Z}_{\lceil (k+1)/2 \rceil}^+ \hat{Z}_{\lfloor (k+1)/2 \rfloor}^- > 0, \mathcal{B}^c \right) \\
& \quad + O(n^{-a} + n^{-b}),
\end{aligned}$$

as $n \rightarrow \infty$, for some constants $a, b > 0$ and all $0 \leq k \leq \omega_n$. It follows that, as $n \rightarrow \infty$,

$$|\mathbb{P}_n(\mathcal{H}_n < \infty) - s^+ s^-| \leq 2\mathbb{E} \left[\exp \left\{ -\frac{\nu n^{1-2\delta}}{\mu-1} W^+ W^- \right\} \mathbb{1}_{\mathcal{B}} \right] \quad (10.34)$$

$$\begin{aligned}
& + \mathbb{P} \left(\hat{Z}_{\lceil (\omega_n+1)/2 \rceil}^+ \hat{Z}_{\lfloor (\omega_n+1)/2 \rfloor}^- > 0, \mathcal{B}^c \right) \quad (10.35) \\
& + O \left(n^{-\min\{a, b, c_1\}} \right),
\end{aligned}$$

where we have used the observation that $\mu^{\omega_n} \geq n^{2(1-\delta)}$.

We start with analyzing (10.35). For this we use Lemma 10.3 to see that conditionally on $W^\pm = 0$ the process $\{\hat{Z}_k^\pm : k \geq 1\}$ is a subcritical delayed branching process with offspring distributions $(\tilde{f}^\pm, \tilde{f}^{\mp, \pm})$, defined in the lemma, having means $\nu^\pm$ and $\lambda^\pm < 1$, respectively. It then follows from the union bound and Markov's inequality that,

$$\begin{aligned}
& \mathbb{P} \left(\hat{Z}_{\lceil (\omega_n+1)/2 \rceil}^+ \hat{Z}_{\lfloor (\omega_n+1)/2 \rfloor}^- > 0, \mathcal{B}^c \right) \\
& \leq \mathbb{P} \left(\hat{Z}_{\lceil (\omega_n+1)/2 \rceil}^+ \geq 1, W^+ = 0 \right) + \mathbb{P} \left(\hat{Z}_{\lfloor (\omega_n+1)/2 \rfloor}^- \geq 1, W^- = 0 \right) \\
& \leq (1-s^+) \nu^+(\lambda^+) \lceil (\omega_n+1)/2 \rceil^{-1} + (1-s^-) \nu^-(\lambda^-) \lfloor (\omega_n+1)/2 \rfloor^{-1} \\
& = O \left(n^{-(1-\delta)\omega^+} \mathbb{1}_{\{q^+ > 0\}} + n^{-(1-\delta)\omega^-} \mathbb{1}_{\{q^- > 0\}} \right),
\end{aligned}$$

as $n \rightarrow \infty$, where $\omega^\pm = |\log \lambda^\pm| / \log \mu$ provided $q^\pm > 0$.

Finally, to analyze the remaining expectation (10.34) define the event

$$\mathcal{M} = \{W^+ > n^{-1/4}, W^- > n^{-1/4}\}$$

and note that

$$\begin{aligned}
& \mathbb{E} \left[\exp \left\{ -\frac{\nu n^{1-2\delta}}{\mu-1} W^+ W^- \right\} \mathbb{1}_{\mathcal{B}} \right] \\
& \leq \mathbb{E} \left[\exp \left\{ -\frac{\nu n^{1-2\delta}}{\mu-1} W^+ W^- \right\} \mathbb{1}_{\mathcal{M}} \right] + \mathbb{P}(\mathcal{B} \cap \mathcal{M}^c) \\
& \leq \exp \left\{ -\frac{\nu}{(\mu-1)} n^{1/2-2\delta} \right\} + \mathbb{P}(\mathcal{B} \cap \mathcal{M}^c) \\
& \leq \mathbb{P}(0 < W^+ \leq n^{-1/4}) + \mathbb{P}(0 < W^- \leq n^{-1/4}) + o(n^{-1}),
\end{aligned}$$

as $n \rightarrow \infty$. The proof of Lemma 10.4 gives that

$$\mathbb{P}(0 < W^\pm \leq n^{-1/4}) = O \left(n^{-1} \mathbb{1}_{\{q^\pm = 0\}} + n^{-\alpha^\pm/4} \mathbb{1}_{\{q^\pm > 0\}} \right),$$

which completes the proof. $\square$

Theorem 10.14 together with Corollary 10.15 imply the main result, Theorem 10.9:

Proof of Theorem 10.9. Note that

$$\mathbb{P}_n(\mathcal{H}_n > k) = \mathbb{P}_n(\mathcal{H}_n > k | \mathcal{H}_n < \infty) \mathbb{P}_n(\mathcal{H}_n < \infty) + \mathbb{P}_n(\mathcal{H}_n = \infty).$$

Hence, defining $\mathcal{B} = \{W^+ W^- > 0\}$, and using Theorem 10.14 and Corollary 10.15 gives, on the event Ξ_n ,

$$\begin{aligned}
& \mathbb{P}_n(\mathcal{H}_n > k | \mathcal{H}_n < \infty) \\
& = \frac{\mathbb{P}_n(\mathcal{H}_n > k) - \mathbb{P}_n(\mathcal{H}_n = \infty)}{\mathbb{P}_n(\mathcal{H}_n < \infty)} \\
& = \frac{1}{\mathbb{P}(\mathcal{B})} \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k}{(\mu-1)n} W^- W^+ \right\} \right] - \frac{\mathbb{P}(\mathcal{B}^c)}{\mathbb{P}(\mathcal{B})} + O(n^{-c}) \\
& = \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k}{(\mu-1)n} W^- W^+ \right\} \middle| W^+ W^- > 0 \right] + O(n^{-c})
\end{aligned}$$

as $n \rightarrow \infty$, for some $c > 0$ and for the range of values of k indicated in the theorems. In particular, this holds for all $\lfloor \log_\mu n \rfloor + t$, for fixed $t \in \mathbb{Z}$.

Now define for $x \in \mathbb{R}$,

$$V_n(x) = 1 - \mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^{\lfloor \log_\mu n \rfloor + \lfloor x \rfloor}}{(\mu - 1)n} W^+ W^- \right\} \middle| W^+ W^- > 0 \right].$$

That $V_n(x)$ is a cumulative distribution function for each fixed n follows from noting that it is non-decreasing with $\lim_{x \rightarrow -\infty} V_n(x) = 0$ and $\lim_{x \rightarrow \infty} V_n(x) = 1$. If we let $\mathcal{H}_n$ be a random variable having distribution V_n then, since $(\mathbf{D}_n^+, \mathbf{D}_n^-)$ satisfies Assumption 10.1, we have for each fixed t ,

$$\begin{aligned} & |\mathbb{P}(\mathcal{H}_n > \lfloor \log_\mu n \rfloor + t | \mathcal{H}_n < \infty) - \mathbb{P}(\mathcal{H}_n > t)| \\ & \leq \frac{1}{\mathbb{P}(\Xi_n)} \mathbb{E} [|\mathbb{P}_n(\mathcal{H}_n > \lfloor \log_\mu n \rfloor + t | \mathcal{H}_n < \infty) - (1 - V_n(t))| \mathbb{1}_{\Xi_n}] \\ & = O(n^{-c}), \end{aligned}$$

as $n \rightarrow \infty$, for some constant $c > 0$. Hence

$$\lim_{n \rightarrow \infty} |\mathbb{P}(\mathcal{H}_n > \lfloor \log_\mu n \rfloor + t | \mathcal{H}_n < \infty) - \mathbb{P}(\mathcal{H}_n > t)| = 0,$$

which implies Theorem 10.9. $\square$

10.4 Numerical examples

To illustrate the accuracy of the approximation for the hopcount between two randomly chosen nodes provided by Theorem 10.9, we give in this section several numerical examples for different choices of the bi-degree sequence. For this we use the directed IID algorithm as described in Chapter 3, Section 3.2.

10.4.1 Computing the hopcount distribution

In order to compute the hopcount distribution, we have constructed 20 graphs of size $n = 10^6$, using the DCM for different choices of bi-degree sequence. For each of these graphs we have computed the neighborhood function, which gives for each $s > 0$ the number of pairs of nodes at distance at least s . For the computation of the neighborhood function we have used the HyperBall algorithm [22], which is part of the Webgraph Framework [21]. We have used HyperBall since it implements the HyperANF algorithm [20], which is designed to give a tight approximation of the neighborhood function of large graphs. From the neighborhood function we have determined, for all finite s , the number of shortest paths of length s . In this way, we compute the distance between all pairs of nodes, with finite distance, in 20 independently generated graphs. We then took the empirical distribution of these values as an approximation of the hopcount distribution.

We point out that since $\mathcal{H}_n$ was defined as the hopcount between two randomly selected nodes, the natural unbiased estimator for the distribution of $\mathcal{H}_n$ is the one obtained from randomly selecting pairs of nodes in independent graphs

and using the corresponding empirical distribution function. However, this approach is computationally too intensive considering the amount of effort needed to generate one graph. We use the HyperBall algorithm since it is considerably more efficient, and although the empirical distribution function it generates does not consist of i.i.d. samples (samples from the same graph are positively correlated), it produces results that are in close agreement with the theoretical approximation in Theorem 10.9. Additional experiments not included in this thesis showed that the two approaches produce similar results, with the method used in this chapter exhibiting smaller variance.

10.4.2 Results for different bi-degree sequences

The three examples below illustrate the accuracy of the approximation provided by Theorem 10.9 for different choices of bi-degree sequences. All three examples are special cases of the IID algorithm, where the random variables $\mathcal{D}^+$ and $\mathcal{D}^-$ satisfy Assumption 3.4 and thus, by Theorem 3.4, satisfy Assumption 10.1.

d -regular bi-degree sequence

A d -regular bi-degree sequence satisfies $D_i^+ = d = D_i^-$ for all $1 \leq i \leq n$. It readily follows that the probability densities $f^\pm$ and $f^{\mp, \pm}$ have just one atom at d . Moreover, we have $\hat{Z}_k^\pm = d^k := \mu^k$ for all $k \geq 1$, hence $W^\pm = 1$ and

$$\mathbb{P}(\mathcal{H}_n \leq x) = 1 - \exp \left\{ -\frac{d^{\lfloor \log_d n \rfloor + \lfloor x \rfloor}}{(d-1)n} \right\}, \quad x \in \mathbb{R}. \quad (10.36)$$

In Figure 10.1(a) we have plotted the probability mass functions of both the hopcount distribution and that of its theoretical limit (10.36). The plots are indistinguishable in the figure, with a Kolmogorov-Smirnov distance of 1.3×10^{-4} . This shows that for non-random sequences, the approximation provided by Theorem 10.9 is almost exact.

I.I.D. bi-degree sequence with independent in- and out-degrees

Following the same approach as for the numerical experiments in Chapter 6, see Section 6.5, we have generated bi-degree sequences $\text{IID}(\mathcal{D}^+, \mathcal{D}^-)$, where $\mathcal{D}^+$ and $\mathcal{D}^-$ are independent Poisson random variables with mixed with Pareto rates with scale parameter 1 and shape parameter $3/2$. Note that the independence between $\mathcal{D}^+$ and $\mathcal{D}^-$ implies that the size-biased distributions $f^{-, +}$ and $f^{+, -}$ are equal to the unbiased ones, i.e., $f^{\mp, \pm} = f^\pm$. Hence, $\mu = \nu = \mathbb{E}[\mathcal{D}^\pm] = 3$ and the branching processes $\{\hat{Z}_k^\pm : k \geq 1\}$ are not delayed.

We remark that in addition to computing the hopcount distribution, we are now also faced with the challenge of determining $\mathbb{P}(\mathcal{H}_n > k)$, since the distributions of $W^\pm$ are not known. In order to simulate $W^\pm$ we use another advanced numerical approach from [27], which describes a bootstrap algorithm for simulating the endogenous solutions of branching linear recursions. For this we first observe that W^+ and W^- satisfy the following stochastic fixed-point equations:

$$W^+ \stackrel{d}{=} \sum_{i=1}^{\mathcal{D}^+} \frac{W_i^+}{\mu} \quad \text{and} \quad W^- \stackrel{d}{=} \sum_{i=1}^{\mathcal{D}^-} \frac{W_i^-}{\mu},$$

Figure 10.1: Hopcount probability mass function compared to the approximation provided by Theorem 10.9 for: (a) a 3-regular bi-degree sequence; (b) a bi-degree sequence generated by the IID algorithm with independent in- and out-degrees; and (c) a bi-degree sequence generated by the IID algorithm with dependent in- and out-degrees. The Kolmogorov-Smirnov distance in each case is: (a) 1.3×10^{-4} , (b) 0.0583, and (c) 0.0353. In all cases the graphs had $n = 10^6$ nodes.

where $W_i^\pm$ are i.i.d. copies of $W^\pm$, independent of $\mathcal{D}^+$ and $\mathcal{D}^-$. Using the algorithm in [27] for 30 generations of the trees with a sample pool of size 10^6 , we obtained 10^6 observations for each of W^+ and W^- , with the sample for W^+ independent of that for W^- . We then used these samples to estimate

$$\mathbb{E} \left[\exp \left\{ -\frac{\nu \mu^k W^+ W^-}{(\mu - 1)n} \right\} \middle| W^+ W^- > 0 \right] \quad \text{for } k = 0, 1, \dots$$

The results for the hopcount distribution are shown in Figure 10.1(b). The Kolmogorov-Smirnov distance in this case is 0.0583.

I.I.D. bi-degree sequence with dependent in- and out-degrees

Our third and last example is for a bi-degree sequence obtained using the IID algorithm but for the case where $\mathcal{D}^+$ and $\mathcal{D}^-$ are dependent. We take the extreme case where $\mathcal{D}_i^+ = \mathcal{D}_i^-$ for all $1 \leq i \leq n$. To obtain such a sequence we generate the $\mathcal{D}_i^+$ by sampling from a Zeta distribution with exponent $7/2$ and set $\mathcal{D}_i^- = \mathcal{D}_i^+$, that is

$$\mathbb{P}(\mathcal{D}^+ = t) = t^{-7/2} / \zeta(7/2) \quad \text{for all } t = 1, 2, \dots,$$

where $\zeta(s)$ is the Riemann zeta function. Observe that since the exponent is larger than 3, the distribution has finite $2 + \varepsilon$ moment, for $0 < \varepsilon < 1/2$. Therefore, it follows from Theorem 3.4 that this bi-degree sequence satisfies Assumption 10.1 with high probability. We used a Zeta distribution here since then, the sized-bias distribution will again be a Zeta distribution, with exponent $5/2$.

The $W^\pm$ were again simulated using the approach from [27] with the same parameters as for the independent case above, but with the appropriate sized-biased distribution and the corresponding delay for the first generation of the tree. The results for the hopcount are shown in Figure 10.1(c), and the Kolmogorov-Smirnov distance is 0.0353.

Chapter 11

Concluding remarks

In this thesis we have analyzed the asymptotic behavior of several structural properties of random graphs.

We have introduced rank-correlation measures for degree-degree correlations in directed graphs (Chapter 4) and have shown that these converge, under very general assumptions, to a meaningful limit (Chapter 5). In addition we have presented two random graph models which satisfy these conditions and analyzed the limits of the rank-correlation measures on these models.

We proved in a rigorous way (Chapter 6) that the directed configuration model has asymptotic neutral mixing. Since this model also generates graphs with a given limiting degree distribution, it is a suitable null model for the analysis of degree-degree correlations on networks. In particular, we conclude that the erased configuration model should be used for generating simple graphs with asymptotic neutral mixing and specified degree distribution, since it is easy to implement and works for all degree distribution with finite mean. Moreover, if the degree distribution has finite variance the average number of removed edges scales as n^{-1} (Chapter 8), which is negligible for large graphs. Therefore, the erased configuration model should be preferred over the repeated configuration model, whose computation time increases with the variance of the degrees.

We saw (Chapter 9) that when the degrees are sampled from regularly-varying distributions with infinite variance, then the values of the rank-correlation measures for the Out-In degree-degree correlation type in the erased configuration model were always negative for finite size graphs. We have used extensive numerical experiments to establish the scaling, up to some slowly varying function, of the fluctuations of the rank-correlation in this model and have shown that this scaling undergoes a phase-transition. Moreover, we were able to fit a normal distribution to these fluctuations, in the finite-variance regime, which is strong evidence that in this case the rank-correlation measures are normal statistics. However this remains to be proven and would make for an important next step in the analysis of degree-degree correlations in networks.

With respect to degree-degree correlations in the erased configuration model, there are several interesting questions left unanswered: 1) What is the exact

scaling of the negative Out-In degree-degree correlations, in terms of the graph size and the exponents of the degree distributions? 2) Is it possible to prove central limit theorems for the rank-correlation measures that exhibit the observed phase-transition and if so, 3) what are the limit distributions?

We have also analyzed (Chapter 8) the scaling of the number of the number of erased edges in the erased configuration model, when the out- and in-degrees are sampled from independent regularly-varying distributions. Here we were able to give a scaling for the conditional expectation, with respect to the degrees, of the average number of removed edges. The next natural question then is how to extend our analysis to the full expectation, which we also believe to have scaling $n^{2-\gamma^*}$, in the directed case, and $n^{2-\gamma}$ in the undirected case. Furthermore, it would be interesting to prove a central limit theorem for the number of removed edges in the erased configuration model, where we again think that the above scaling terms will apply.

Apart from the configuration model, we have also introduced a model for generating undirected graphs with given degree distributions and for which Spearman's rho is minimal (Chapter 7). For this model we obtained a full characterization of the joint degree distribution, which was completely described by the size-biased degree distribution. We have used this characterization to prove that the minimal value of Spearman's rho is independent of the tail of the degree distribution.

Interestingly, numerical experiments have shown that the minimal value of Spearman's rho on maximally disassortative graphs, with Poisson or Pareto degree distributions, can take any value in $[-1, 0]$, depending on the specific shape of the distributions. For the Pareto distribution the parameter that controls this is the exponent γ , while this was the inverse mean λ^{-1} for the Poisson distribution. The experiments suggested that as $\gamma, \lambda^{-1} \rightarrow \infty$ the minimal value of Spearman's rho on graphs with these degree distributions converges to 0 and to -1 when $\gamma, \lambda^{-1} \rightarrow 0$. This result, however, remains to be proven. In addition, classifying which properties of the distribution control the behavior of Spearman's rho, and hence degree-degree correlations, remains a relevant and open question.

The above observations show a discrepancy between the values of rank-correlations measures and the intuitive interpretation that in assortative graphs large nodes often connect to large nodes, while in disassortative graphs they have a preference to connect to small nodes. The main reason behind this is that although scale-free graphs have some very large nodes, there are only a finite number of them, in terms of the graph size, while the number of small nodes grows with the size of the graph. Therefore, assigning a single value to a graph in order to describe degree-degree correlations might, in general, not reflect the intuitive ideas behind degree assortativity. Hence, more effort should be spent on the mathematical analysis of measures for degree-degree correlations that do take into account the value of the degrees, such as the average nearest neighbor degree [54, 67], angular measure [10, 71] or a weighted version of Kendall's tau [90].

Extensions of our research in Chapter 7 could include: 1) extending the model to the directed setting and 2) developing a model for generating maximally assortative simple graphs with given degree distributions.

Lastly (Chapter 10), we have analyzed the distribution of the hopcount in the directed configuration model and showed that it scales logarithmically with the size of the graph, as long as the out- and in-degree distribution have finite covariance. This condition emphasizes the subtle difference between the directed and undirected model, where in the latter distances scale logarithmically only when the degree-distribution has finite variance. Hence, it would be interesting to investigate how the joint distribution between the out- and in-degree of a node influences the distances. Moreover, since the configuration model has neutral mixing of degrees, we could ask to what extend degree-degree correlations influence the behavior of distances. The first step in this direction would then be to analyze the behavior of the hopcount on models with extreme joint degree structure. For this, the model we have introduced for generating maximally disassortative graphs could be used.

For the analysis of the hopcount distribution we used a direct coupling between a breadth-first exploration of a graph generated by the configuration model and a branching process. This coupling result is of independent interest since it holds until an equivalent of $n^{1-\delta}$ nodes have been discovered by the exploration process, for arbitrary small $\delta > 0$, and hence improves on currently known coupling techniques. Furthermore, an important ingredient for proving the coupling result was that the out-degree of a node to which an edge points was independent of the in-degree of the node to which it points. In other words, we have used that the directed configuration model has neutral mixing of degrees for the Out-In degree-degree correlation type. Hence it might be possible to extend this coupling to a general class of random graphs that have this type of neutral mixing, asymptotically.

Bibliography

- [1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [2] D. L. Alderson and L. Li. Diversity of graphs with highly variable connectivity. *Physical Review E*, 75(4):046102, 2007.
- [3] O. Angel, R. van der Hofstad, and C. Holmgren. Limit laws for self-loops and multiple edges in the configuration model. *arXiv preprint arXiv:1603.07172*, 2016.
- [4] K. B. Athreya and P. E. Ney. *Branching processes*, volume 196. Springer Science & Business Media, 2012.
- [5] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42. ACM, 2012.
- [6] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [7] A. Barbour and G. Reinert. Small worlds. *Random Structures & Algorithms*, 19(1):54–74, 2001.
- [8] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of theoretical biology*, 235(2):275–288, 2005.
- [9] K. E. Bassler, C. I. Del Genio, P. L. Erdős, I. Miklós, and Z. Toroczkai. Exact sampling of graphs with prescribed degree correlations. *New Journal of Physics*, 17(8):083052, 2015.
- [10] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.
- [11] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. *preparation. Version August*, 19, 2007.
- [12] J. Biggins and N. Bingham. Large deviations in the supercritical branching process. *Advances in Applied Probability*, pages 757–772, 1993.

- [13] N. Bingham and R. Doney. Asymptotic properties of supercritical branching processes I: The Galton-Watson process. *Advances in Applied Probability*, pages 711–731, 1974.
- [14] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [15] J. Blitzstein and P. Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4):489–522, 2011.
- [16] M. Boguná, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Physical review letters*, 90(2):028701, 2003.
- [17] M. Boguná, R. Pastor-Satorras, and A. Vespignani. Cut-offs and finite size effects in scale-free networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):205–209, 2004.
- [18] M. Boguñá, M.á. R. Pastor-Satorras, and A. Vespignani. Epidemic spreading in complex networks with degree correlations. In *Statistical mechanics of complex networks*, pages 127–147. Springer, 2003.
- [19] M. Boguñá and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66(4):047104, 2002.
- [20] P. Boldi, M. Rosa, and S. Vigna. HyperANF: Approximating the neighbourhood function of very large graphs on a budget. In *Proceedings of the 20th international conference on World wide web*, pages 625–634. ACM, 2011.
- [21] P. Boldi and S. Vigna. The webgraph framework I: compression techniques. In *Proceedings of the 13th international conference on World Wide Web*, pages 595–602. ACM, 2004.
- [22] P. Boldi and S. Vigna. In-core computation of geometric centralities with hyperball: A hundred billion nodes and beyond. In *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*, pages 621–628. IEEE, 2013.
- [23] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.
- [24] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.
- [25] D. L. Burkholder. Distribution function inequalities for martingales. *the Annals of Probability*, pages 19–42, 1973.
- [26] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Physical Review E*, 71(2):027103, 2005.

- [27] N. Chen, N. Litvak, and M. Olvera-Cravioto. Ranking algorithms on directed configuration networks. *Preprint arXiv:1409.7443*, 2014.
- [28] N. Chen and M. Olvera-Cravioto. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013.
- [29] N. Chen and M. Olvera-Cravioto. Efficient simulation for branching linear recursions. In *Proceedings of the 2015 Winter Simulation Conference*, pages 1–12, Huntington Beach, CA, 2015.
- [30] A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [31] C. Cooper and A. Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combinatorics, Probability and Computing*, 13(03):319–337, 2004.
- [32] G. D’Agostino, A. Scala, V. Zlatić, and G. Caldarelli. Robustness and assortativity for diffusion-like processes in scale-free networks. *EPL (Europhysics Letters)*, 97(6):68006, 2012.
- [33] S. de Franciscis, S. Johnson, and J. J. Torres. Enhancing neural-network performance via assortativity. *Physical Review E*, 83(3):036114, 2011.
- [34] E. del Barrio, E. Giné, and C. Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Annals of Probability*, pages 1009–1071, 1999.
- [35] C. I. Del Genio, H. Kim, Z. Toroczkai, and K. E. Bassler. Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PloS one*, 5(4):e10012, 2010.
- [36] M. Denuit and P. Lambert. Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57, 2005.
- [37] P. Deprez and M. V. Wüthrich. Construction of directed assortative configuration graphs. *arXiv preprint arXiv:1510.00575*, 2015.
- [38] S. Dorogovtsev, A. Ferreira, A. Goltsev, and J. Mendes. Zero Pearson coefficient for strongly correlated growing trees. *Physical Review E*, 81(3):031135, 2010.
- [39] S. N. Dorogovtsev and J. F. Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.
- [40] R. Durrett. *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1 edition, 2010.
- [41] P. Embrechts and C. M. Goldie. On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society (Series A)*, 29(02):243–256, 1980.
- [42] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.

- [43] P. Flajolet, É. Fusy, O. Gandouet, and F. Meunier. Hyperloglog: the analysis of a near-optimal cardinality estimation algorithm. *DMTCS Proceedings*, (1), 2008.
- [44] J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, 107(24):10815–10820, 2010.
- [45] P. Hagmann, M. Kuran, X. Gigandet, P. Thiran, V. J. Wedeen, R. Meuli, and J.-P. Thiran. Mapping human whole-brain structural networks with diffusion MRI. *PloS one*, 2(7):e597, 2007.
- [46] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The annals of mathematical statistics*, pages 293–325, 1948.
- [47] T. Hurd. The Construction and Properties of Assortative Configuration Graphs. *arXiv preprint arXiv:1512.03084*, 2015.
- [48] R. Kannan, P. Tetali, and S. Vempala. Simple Markov-chain algorithms for generating bipartite graphs and tournaments. *Random Structures and Algorithms*, 14(4):293–308, 1999.
- [49] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
- [50] H. Kim, C. I. Del Genio, K. E. Bassler, and Z. Toroczkai. Constructing and sampling directed graphs with given degree sequences. *New Journal of Physics*, 14(2):023012, 2012.
- [51] W. R. Knight. A computer method for calculating Kendall’s tau with ungrouped data. *Journal of the American Statistical Association*, 61(314):436–439, 1966.
- [52] D. Li, Y. Qi, and A. Rosalsky. A refinement of the Kolmogorov–Marcinkiewicz–Zygmund strong law of large numbers. *Journal of Theoretical Probability*, 24(4):1130–1156, 2011.
- [53] N. Litvak and R. van der Hofstad. Uncovering disassortativity in large scale-free networks. *Physical Review E*, 87(2):022801, 2013.
- [54] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.
- [55] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: correlation profile of the internet. *Physica A: Statistical Mechanics and its Applications*, 333:529–540, 2004.
- [56] J. Menche, A. Valleriani, and R. Lipowsky. Asymptotic properties of degree-correlated scale-free networks. *Physical Review E*, 81(4):046103, 2010.
- [57] M. Mesfioui and A. Tajar. On the properties of some nonparametric concordance measures in the discrete case. *Nonparametric Statistics*, 17(5):541–554, 2005.

- [58] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.
- [59] M. E. Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.
- [60] M. E. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.
- [61] M. E. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
- [62] M. E. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, 2001.
- [63] M. E. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Physical Review E*, 60(6):7332, 1999.
- [64] R. Noldus and P. Van Mieghem. Assortativity in complex networks. *Journal of Complex Networks*, page cnv005, 2015.
- [65] I. Norros, H. Reittu, et al. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75, 2006.
- [66] J. Park and M. E. Newman. Origin of degree correlations in the Internet and other networks. *Physical Review E*, 68(2):026112, 2003.
- [67] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the Internet. *Physical review letters*, 87(25):258701, 2001.
- [68] M. D. Penrose. The strong giant in a random digraph. *arXiv preprint arXiv:1409.4371*, 2014.
- [69] M. Piraveenan, M. Prokopenko, and A. Zomaya. Assortative mixing in directed biological networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 9(1):66–78, 2012.
- [70] M. Raschke, M. Schläpfer, and R. Nibali. Measuring degree-degree association in networks. *Physical Review E*, 82(3):037102, 2010.
- [71] S. I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer, 2007.
- [72] W. E. Schlauch, E. Á. Horvát, and K. A. Zweig. Different flavors of randomness: comparing random graph models with fixed degree sequences. *Social Network Analysis and Mining*, 5(1):1–14, 2015.
- [73] C. Schmeltzer, A. H. Kihara, I. M. Sokolov, and S. Rüdiger. Degree correlations optimize neuronal network sensitivity to sub-threshold stimuli. *PloS one*, 10(6):e0121794, 2015.

- [74] M. Serge Dubuc. La densité de la loi-limite d'un processus en cascade expansif. *Probability Theory and Related Fields*, 19(4):281–290, 1971.
- [75] C. Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904.
- [76] I. Stanton and A. Pinar. Constructing and sampling graphs with a prescribed joint degree distribution. *Journal of Experimental Algorithmics (JEA)*, 17:3–5, 2012.
- [77] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs. *Journal of Statistical Physics*, 133(1):169–202, 2008.
- [78] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, 8(3):111–141, 2005.
- [79] R. van der Hofstad. Random Graphs and Complex Networks. *Lecture notes in prep*, 2014.
- [80] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures & Algorithms*, 27(1):76–123, 2005.
- [81] R. van der Hofstad, G. Hooghiemstra, D. Znamenski, et al. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, 12(25):703–766, 2007.
- [82] R. van der Hofstad and N. Litvak. Degree-degree dependencies in random graphs with heavy-tailed degrees. *Internet mathematics*, 10(3-4):287–334, 2014.
- [83] P. van der Hoorn and N. Litvak. Convergence of rank based degree-degree correlations in random directed networks. *Moscow Journal of Combinatorics and Number Theory*, 4(4):45–83, 2014.
- [84] P. van der Hoorn and N. Litvak. Degree-Degree Dependencies in Directed Networks with Heavy-Tailed Degrees. *Internet Mathematics*, 11(2):155–179, 2015.
- [85] P. van der Hoorn and N. Litvak. Phase transitions for scaling of structural correlations in directed networks. *Physical Review E*, 92(2):022803, 2015.
- [86] P. van der Hoorn and N. Litvak. Upper bounds for number of removed edges in the Erased Configuration Model. In *Algorithms and Models for the Web Graph*, pages 54–65. Springer, 2015.
- [87] P. van der Hoorn and M. Olvera-Cravioto. Distances in the directed configuration model. *arXiv preprint arXiv:1511.04553*, 2015.
- [88] P. van der Hoorn, L. Ostroumova Prokhorenkova, and E. Samosvat. Generating maximally disassortative graphs with given degree distribution. *arXiv preprint arXiv:1607.01742*, 2016.

- [89] A. Vázquez and Y. Moreno. Resilience to damage of graphs with degree correlations. *Phys. Rev. E*, 67:015101, Jan 2003.
- [90] S. Vigna. A weighted correlation index for rankings with ties. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1166–1176. International World Wide Web Conferences Steering Committee, 2015.
- [91] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [92] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.
- [93] W. Whitt. *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. Springer, 2002.
- [94] N. C. Wormald. Some problems in the enumeration of labelled graphs. *Bulletin of the Australian Mathematical Society*, 21(01):159–160, 1980.
- [95] R. Xulvi-Brunet and I. Sokolov. Reshuffling scale-free networks: From random to assortative. *Physical Review E*, 70(6):066102, 2004.
- [96] R. Xulvi-Brunet and I. Sokolov. Changing correlations in networks: assortativity and dissortativity. *Acta Physica Polonica B*, 36(5):1431, 2005.
- [97] D. Yang, L. Pan, and T. Zhou. Lower bound of assortativity coefficient in scale-free networks. *arXiv preprint arXiv:1602.04350*, 2016.
- [98] D. Zhou, H. E. Stanley, G. D’Agostino, and A. Scala. Assortativity decreases the robustness of interdependent networks. *Physical Review E*, 86(6):066103, 2012.

Glossary

Graphs and degrees

Directed graphs:

<i>Number of edges from i to j:</i>	G_{ij}
<i>Out-degree of node i:</i>	$D_i^+ = \sum_{j=1}^n G_{ij}$
<i>In-degree of node i:</i>	$D_i^- = \sum_{j=1}^n G_{ji}$
<i>Number of edges:</i>	$L_n = \sum_{i=1}^n D_i^+ = \sum_{i=1}^n D_i^-$
<i>Summation over edges:</i>	$\sum_{i \rightarrow j} \phi(i, j) = \sum_{i, j=1}^n G_{ij} \phi(i, j)$

Undirected graphs:

<i>Number of edges between i and j:</i>	$G_{ij} = G_{ji}$
<i>Degree of node i:</i>	$D_i = \sum_{j=1}^n G_{ij}$
<i>Number of stubs:</i>	$L_n = \sum_{i=1}^n D_i$
<i>Summation over edges:</i>	$\sum_{i-j} \phi(i, j) = \sum_{1 \leq i \leq j \leq n} G_{ij} \phi(i, j)$

Empirical probability mass functions

Directed graphs:

Degree distribution:

$$f_n^\pm(k) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i^\pm=k\}} \qquad F_n^\pm(k) = \sum_{t=0}^k f_n^\pm(t)$$

Joint degree distribution of nodes:

$$h_n(k, \ell) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i^+=k\}} \mathbb{1}_{\{D_i^-=\ell\}} \qquad H_n(k, \ell) = \sum_{t=0}^k \sum_{s=0}^{\ell} h_n(t, s)$$

Size-biased degree distribution:

$$\begin{aligned} f_n^{\alpha, \beta}(k) &= \frac{1}{L_n} \sum_{i=1}^n D_i^\alpha \mathbb{1}_{\{D_i^\beta=k\}} & F_n^{\alpha, \beta}(k) &= \sum_{t=0}^k f_n^{\alpha, \beta}(t) \\ \mathcal{F}_n^{\alpha, \beta}(k) &= F_n^{\alpha, \beta}(k) + F_n^{\alpha, \beta}(k-1) \end{aligned}$$

Joint degree distribution of edges:

$$h_n^{\alpha,\beta}(k, \ell) = \frac{1}{L_n} \sum_{i \rightarrow j} \mathbb{1}_{\{D_i^\alpha=k\}} \mathbb{1}_{\{D_j^\beta=\ell\}} \quad H_n^{\alpha,\beta}(k, \ell) = \sum_{t=0}^k \sum_{s=0}^{\ell} h_n^{\alpha,\beta}(t, s)$$

$$\mathcal{H}_n^{\alpha,\beta}(k, \ell) = H_n^{\alpha,\beta}(k, \ell) + H_n^{\alpha,\beta}(k, \ell - 1) + H_n^{\alpha,\beta}(k - 1, \ell) + H_n^{\alpha,\beta}(k - 1, \ell - 1)$$

Undirected graphs:

Degree distribution:

$$f_n(k) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i=k\}} \quad F_n(k) = \sum_{t=0}^k f_n^\pm(t)$$

Size-biased degree distribution:

$$f_n^*(k) = \frac{1}{L_n} \sum_{i=1}^n D_i \mathbb{1}_{\{D_i=k\}} \quad F_n^*(k) = \sum_{t=0}^k f_n^*(t)$$

$$\mathcal{F}_n^*(k) = F_n^*(k) + F_n^*(k - 1)$$

Joint degree distribution of edges:

$$h_n^*(k, \ell) = \frac{1}{L_n} \sum_{i \rightarrow j} \mathbb{1}_{\{D_i=k\}} \mathbb{1}_{\{D_j=\ell\}} \quad H_n^*(k, \ell) = \sum_{t=0}^k \sum_{s=0}^{\ell} h_n^*(t, s)$$

Summary

Our world is filled with complex systems, ranging from technological systems such as the Internet and the World Wide Web, to the human brain and social interactions between individuals or even organizations. Many of such systems can be modeled as a network, consisting of nodes and relations between them, which are called edges. These edges can be undirected, representing symmetric relations such as a friendship in a social network, or have a direction, in which case the relations are asymmetric. For instance, in the World Wide Web, nodes are web-pages and the edges are the hyperlinks between them, which are directed since they point from one page to another.

Understanding the structure of these networks is of vital importance for understanding the behavior of the systems they represent and the processes that run on those systems. In this thesis we analyze structural properties of networks by combining the mathematical theory of probability and graphs with large scale numerical experiments. Due to the fact that most real-world networks we want to analyze are large, we are interested in the asymptotic behavior of the structure of networks, as the number of nodes goes to infinity. The structural properties we consider in this thesis are degree-degree correlations and directed paths.

Given a network, each node has a number of edges which we call the degree of the node. When the network is directed we distinguish between out-degree, the number of edges pointing out of the node, and in-degree, the number of edges pointing towards the node. Degree-degree correlations describe the correlation between the (out/in) degrees of connected nodes in a network. A network is said to have assortative mixing (of degrees) if this correlation is positive, for instance, if nodes of large degree connect to nodes of large degree. When the correlation is negative, for example when nodes of large degree are connected to nodes of small degree, the network is said to have disassortative mixing. We say that a network has neutral mixing if the correlation is zero. In this case nodes have no preference, in terms of degrees, to which nodes they connect.

A measure for degree-degree correlations assigns to a network a number in between -1 and 1 , that characterizes the correlation between the degrees of connected nodes. In this work, we analyze the asymptotic behavior of different measures for degree-degree correlations in directed networks. We conclude that the most commonly used measure, related to Pearson's correlation coefficient, has significant short comings. We propose several alternative measures, related to rank-correlations, and show that these are statistically consistent estimators of degree-degree correlations in directed networks, under very general assumptions.

We also consider two models for generating networks with a specific distribution of the degrees and with specific degree-degree correlations. The first is the well-known directed configuration model. This model assigns to each node an out- and in-degree, which we see as a number of half-edges pointing out of, and into the node. Then we create the network by randomly pairing half-edges pointing outwards to those pointing inwards, to form edges between nodes. The general version of this model will generate networks where nodes can be connected to themselves (self-loops), or where there can be multiple edges from one node to another. Networks that do not have these features are called simple, and this is the case for most real-world networks. Therefore, there are two versions of the configuration model that generate simple networks. The first is the repeated version, where we simply repeat the connection steps, until the resulting network is simple. The second one is the erased version in which we remove all self-loops and replace all multiple edges between two nodes by just one edge.

We analyze degree-degree correlations in all three versions of the configuration model and prove that they all have neutral mixing, asymptotically. In addition we analyze the number of edges that are removed in the erased configuration model. Here we establish a new scaling result in terms of the number of nodes in the network. We use the insights from this analysis to analyze the finite size effects on the measures for degree-degree correlations in the directed configuration model. When the degree distributions follows a power-law, we find an interesting phase transition in the fluctuations of the measures, in terms of the exponent of the power-law.

We also introduce a new model, that generates graphs that have maximal disassortative mixing. Our model first assigns, similar to the configuration model, degrees to each node. Then we employ a greedy algorithm for connecting the half-edges, which minimizes the value of measures for degree-degree correlations and makes sure that the resulting network is simple. We show that the model gives a lower bound on the value of measures for degree-degree correlations in general networks and prove several asymptotic properties of this model. Using large scale experiments we also show that the minimal achievable value of a measure for degree-degree correlations depends on the full distribution of the degrees and this value can be much larger than the absolute minimum -1 .

In the last part of this thesis we investigate the hopcount in the directed configuration model, where the degree distribution has finite mean and infinite variance. Given a directed network, the hopcount of the network is defined as the shortest directed path between two nodes, selected at random. We show that the hopcount scales logarithmically with the number of nodes and determine the fluctuations around this value. Interestingly, we find that for the directed model it is the correlation between the out- and in-degrees of individual nodes that plays an important role in the scaling of the hopcount and not the moments of the degree distribution. This is different from the undirected model, where, for infinite variance degrees, the hopcount was found to scale double logarithmically in the number of nodes. To complement our results, we use recently developed, state-of-the-art, algorithms to compute the hopcount distribution on large networks generated by the configuration model and find that already for graphs of reasonable size, one million nodes, our asymptotic results are extremely close to the true fluctuations.

Samenvatting

Onze wereld is gevuld met complexe systemen. Van technologische systemen zoals het Internet en het World Wide Web, tot het menselijk brein en sociale interacties tussen individuen of organisaties. Veel van deze systemen kunnen gemodelleerd worden als een netwerk, bestaande uit een verzameling punten en relaties hiertussen, die we verbindingen noemen. Deze verbindingen kunnen ongericht zijn, bijvoorbeeld wanneer ze een symmetrische relaties zoals vriendschap in een sociaal netwerk representeren, of een richting hebben in welk geval de relaties asymmetrisch zijn. In het World Wide Web bijvoorbeeld, zijn de punten de webpagina's en de verbindingen de hyperlinks tussen de pagina's. Deze zijn gericht omdat een hyperlink van de ene naar de andere pagina wijst.

Inzicht in, en begrip van de structuur van netwerken is van essentieel belang om het gedrag te begrijpen van de systemen die ze representeren en de processen die erop plaats vinden. In dit proefschrift analyseren wij structurele eigenschappen van netwerken door het combineren van de wiskundige theorie van kansrekening en grafen met grootschalige numerieke experimenten. Omdat de meeste netwerken in de echte wereld, die we willen analyseren, erg groot zijn, zijn wij vooral geïnteresseerd in het asymptotisch gedrag van de structuur, als het aantal knopen in het netwerk naar oneindig gaat. In dit proefschrift onderzoeken wij twee structurele eigenschappen, te weten, degree-degree correlaties en gerichte paden tussen knopen.

Elk punt in een netwerk heeft een aantal verbindingen. In een ongericht netwerk noemen wij dit aantal de graad van het punt, of degree in het Engels. Als het netwerk gericht is dan maken wij onderscheid tussen de uit-graad van een punt, het aantal verbindingen die van het punt af wijzen, en de in-graad, het aantal verbindingen die naar het punt wijzen. Degree-degree correlaties beschrijven de correlatie tussen de graden van verbonden punten in een netwerk. Een netwerk wordt assortatief genoemd als deze correlatie positief is, bijvoorbeeld als punten met een hoge graad verbonden zijn met andere punten met een hoge graad. Als de correlatie negatief is, in welk geval punten met een hoge graad verbonden zijn met punten met een lage graad, dan wordt het netwerk disassortatief genoemd. We zeggen dat een netwerk neutraal is als de correlatie nul is. In dit geval hebben punten geen voorkeur, in termen van graden, met welke andere punten ze verbonden zijn.

Een maat voor degree-degree correlaties kent een getal tussen de -1 (disassortatief) en 1 (assortatief) toe aan een netwerk, die de correlatie tussen de graden van verbonden knopen karakteriseert. In dit werk analyseren wij het asymptotische

tisch gedrag van verschillende maten voor degree-degree correlaties in gerichte netwerken. Onze conclusie is dat de meest gebruikte maat, die gerelateerd is aan Pearsons correlatie coëfficiënt, belangrijke tekortkomingen heeft. Wij introduceren verschillende alternatieve maten, die gebaseerd zijn op correlatie maten uit de order statistiek, en laten zien dat deze wel statistisch consistente schatters zijn voor degree-degree correlaties in gerichte netwerken.

Naast verschillende maten, bekijken wij ook twee modellen voor het genereren van netwerken met een specifieke verdeling van de graden van knopen en specifieke degree-degree correlaties. Het eerste model is het bekende gerichte configuratie model. Dit model wijst eerst aan elk punt een uit- en in-graad toe, die wij representeren als half-verbindingen die van het punt en naar het punt toe wijzen. Het netwerk wordt vervolgens gemaakt door elke half-verbinding die van een punt afwijkt te verbinden met een willekeurig geselecteerde half-verbinding die naar een punt wijst, om zo een gerichte verbinding tussen de punten te maken. Deze algemene versie van het model kan netwerken creëren waar punten een verbinding met zichzelf kunnen hebben (self-loop) of waar er meerdere verbindingen van een punt naar ander punt kunnen zijn. Netwerken die deze eigenschappen niet hebben worden simpel genoemd. Dit is het geval voor de meeste netwerken in de echte wereld. Daarom zijn er twee versies van het configuratie model die simpele netwerken kunnen maken. De eerste is het repeated model, waar we simpelweg de verbindingsstap blijven herhalen totdat het resulterende netwerk simpel is. De tweede versie is het erased model, waarin alle self-loops worden verwijderd en waar we meerdere verbindingen van een punt naar een andere vervangen door slechts een verbinding.

Wij analyseren degree-degree correlaties in alle drie de versies van het configuratie model en bewijzen dat deze allemaal netwerken genereren die, asymptotisch, neutraal zijn. Daarnaast analyseren we ook het aantal verbindingen dat wordt verwijderd in het erased model. Hier verkrijgen we een nieuw resultaat die de relatie laat zien tussen het aantal verwijderde verbindingen en het aantal punten in het netwerk. We gebruiken de inzichten van deze analyse om te kijken naar de effecten die optreden in de maten voor degree-degree correlaties voor het gerichte configuratie model, als de netwerken een eindig aantal punten hebben. Als de verdeling van de graden van punten een inverse macht van het aantal punten is (power-law), vinden wij een interessante faseovergang in de fluctuaties van de maten.

Naast het configuratie model introduceren wij ook een nieuw model dat netwerken genereert die maximaal disassortatief zijn. Ons model wijst eerst, net als het configuratie model, graden toe aan elk punt. Hierna gebruiken wij een speciaal ontwikkeld gretig algoritme dat de half-verbindingen met elkaar verbindt die de waarde van maten voor degree-degree correlaties minimaliseert en zodat het netwerk simpel is. Wij laten zien dat dit model een ondergrens geeft voor de waarde van maten voor degree-degree correlaties voor netwerken in het algemeen en bewijzen verschillende asymptotische eigenschappen van dit model. Met behulp van grootschalige numerieke experimenten laten we zien dat de minimaal haalbare waarde voor een maat voor degree-degree correlaties sterk afhangt van de volledige verdeling van de graden van de punten. Bovendien kan deze minimale waarde veel groter zijn dan het absolute minimum -1 .

In het laatste deel van dit proefschrift onderzoeken wij de hopcount in het gerichte configuratie model, waar de verdeling van de graden van de punten eindige verwachting en oneindige variantie heeft. Voor een gericht netwerk is de hopcount gedefinieerd als het kortste gerichte pad tussen twee willekeurig geselecteerde punten. We laten zien dat de hopcount logaritmisch schaal met het aantal punten en bepalen de asymptotische fluctuaties. Het interessante dat wij vinden, is dat het de correlatie tussen het aantal uitgaande en inkomende verbindingen van een punt is die een belangrijke rol speelt voor de hopcount in het gerichte model, en niet de momenten van de verdeling van de graden. Dit is anders dan in het ongerichte configuratie model, waar de hopcount dubbel logaritmisch schaal met het aantal punten, als de variantie van de verdeling van de graden oneindig is. Om onze resultaten aan te vullen gebruiken wij recent ontwikkelde, state-of-the-art, algoritmes om de hopcount verdeling op grote configuratie netwerken te berekenen en concluderen we dat al vanaf netwerken met een miljoen punten, onze asymptotische resultaten de werkelijkheid erg goed benaderen.

About the author

Pim van der Hoorn was born on August 18, 1985, in Leiderdorp, the Netherlands. After obtaining his high school diploma in 2005 Pim started studying Physics and Astronomy at the University of Utrecht. Given his interest in theoretical physics and cosmology Pim followed a great amount of mathematics courses and in this way obtained a bachelor's degree in both physics and mathematics, with the thesis *Inverse limits in the A -dèle ring*. By then, his interest had shifted to pure theoretical mathematics and Pim started with the master Mathematical Sciences at the same institution, with a focus on mathematical logic and geometry. He obtained his master's degree cum laude in 2011, with a thesis on categorical quantum mechanics titled *Orthogonal bases in inverse semigroups, a categorical approach*.

After obtaining his master's degree, Pim went into industry as a software engineer and architect, designing and writing software solutions for banks and government organizations. In October 2012 he returned to academia to start his Ph.D. research on complex networks within the European funded project New tools and Algorithms for DIrected NEtwork analysis (NADINE), at the University of Twente, the Netherlands, under the supervision of Prof. dr. Richard Boucherie and Dr. Nelly Litvak. During the four years of his research, Pim made a short visit to Columbia University, New York, United States, and did a two months joint internship at Moscow Institute of Physics and Technology and Yandex in Moscow, Russia. In addition, Pim was accepted as a participant in the Heidelberg Laureate Forum 2015 and participated as a board member in the student organization W.S.G. Abacus, affiliated with the mathematics department of the University of Twente.