Graph Entropy and Related Topics

Yanni Dong

Graph entropy and related topics

Yanni Dong

GRAPH ENTROPY
AND RELATED TOPICS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Wednesday the 5^{th} of July 2023 at 10:45 hours

by

Yanni Dong

born on the 5^{th} of January 1993
in Lanzhou, Gansu, China

This dissertation has been approved by the supervisors:
prof. dr. ir. H.J. Broersma and prof. dr. S. Zhang

The research reported in this thesis has been carried out within the framework of the MEMORANDUM OF AGREEMENT FOR A DOUBLE DOCTORATE DEGREE BETWEEN NORTHWESTERN POLYTECHNICAL UNIVERSITY, PEOPLE'S REPUBLIC OF CHINA AND THE UNIVERSITY OF TWENTE, THE NETHERLANDS.

DSI Ph.D. Thesis Series No. 23-005
Digital Society Institute
P.O. Box 217, 7500 AE Enschede, The Netherlands.

ISBN (print): 978-90-365-5666-8
ISBN (digital): 978-90-365-5667-5
ISSN: 2589-7721 (DSI Ph.D. thesis Series No. 23-005)
DOI: 10.3990/1.9789036556675
URL: https://doi.org/10.3990/1.9789036556675

Typeset with LaTeX
Printed by Ipskamp Printing, Enschede
Cover design by Yanni Dong

Graduation Committee

Chairman/secretary:

Prof. dr. J.N. Kok	University of Twente

Supervisors:

Prof. dr. ir. H.J. Broersma	University of Twente
Prof. dr. S. Zhang	Northwestern Polytechnical University

Prof. dr. J.L. Hurink	University of Twente
Prof. dr. M.I.A. Stoelinga	University of Twente
Prof. dr. X. Li	Nankai University
Prof. dr. ir. D. Paulusma	Durham University
Prof. dr. I. Schiermeyer	Technische Universität Bergakademie Freiberg

Preface

This thesis consists of an introductory chapter (Chapter 1), followed by six technical chapters. These six chapters have been written in the style of journal papers, based on the five joint papers which are listed on the next page.

The presented results of this thesis deal with graph entropy, a concept in chemical graph theory inspired by the well-known Shannon entropy in information theory. In fact, the thesis focuses on different variants of graph entropy, mainly on degree-based and distance-based entropies.

Apart from Chapter 5, the other chapters of the thesis are mainly based on the research results that the author obtained when she was working as a joint Ph.D. candidate at Northwestern Polytechnical University in Xi'an, P.R. China and the University of Twente in Enschede, the Netherlands.

Chapter 2 studies the effect of graph operations on the degree-based entropy. The results of this chapter provide tools for the research on follow-up extremal problems. Chapter 3 to Chapter 5 mainly deal with extremal problems involving the degree-based entropy restricted to specific graph classes. Chapter 5 is based on research that was carried out while the author was visiting the Algorithms and Complexity group in the Computer Science department of Durham University in Durham, UK.

Chapter 6 addresses extremal problems involving two important distance-based entropies. Chapter 7 studies the computational complexity of spanning tree problems for graphical function indices. These indices unify a large number of well-studied topological indices originating from chemical graph theory, and are closely related to degree-based graph entropies.

Papers underlying this thesis

[1] The effect of graph operations on the degree-based entropy, *Appl. Math. Comput.*, **437** (2023) 127533 (with Hajo Broersma, Changwu Song, Pengfei Wan and Shenggui Zhang).
(Chapters 2 and 3)

[2] Extremal values of degree-based entropies of bipartite graphs, submitted (with Stijn Cambie and Matteo Mazzamurro).
(Chapter 4)

[3] Graphs with minimum degree-based entropy, submitted (with Maximillien Gadouleau, Pengfei Wan and Shenggui Zhang).
(Chapter 5)

[4] On the main distance-based entropies: the eccentricity- and Wiener-entropy, submitted (with Stijn Cambie).
(Chapter 6)

[5] The complexity of spanning tree problems involving graphical indices, submitted (with Hajo Broersma, Yuhang Bai and Shenggui Zhang).
(Chapter 7)

Other recent joint papers by the author

[1] Maximum values of degree-based entropies of bipartite graphs, *Appl. Math. Comput.*, **401** (2021) 126094 (with Shengning Qiao, Bing Chen, Pengfei Wan and Shenggui Zhang).

[2] Two-dimensional structural entropy of complete k-partite graphs, in preparation (with Shenggui Zhang, Shengning Qiao, Guihai Yu and Yang Long).

[3] Graphs with extremal values of vertex-degree function index, in preparation (with Hajo Broersma and Shenggui Zhang).

Contents

Chapter 1

Introduction

> *"...everything that is going on in Nature means an increase of the entropy of the part of the world where it is going on."* -Erwin Schrödinger[106]

As the title of the thesis indicates, this thesis deals with contributions to graph theory involving the concept of graph entropy and related topics. However, the notion of entropy originates from thermodynamics, not graph theory. Therefore, before introducing the relevant graph-theoretical definitions and notation, let us first briefly sketch the background on the origin of the term entropy from thermodynamics and its widely studied variants. More details on the history and relevance of entropy can be found in Section 1.2.

Entropy is not an arcane concept in the literature. It has appeared in many scientific disciplines, such as mathematics, chemistry, physics and information theory. As some of the readers might know, this concept originating from thermodynamics was introduced by the German physicist R. Clausius [34, 35]. He gave the first mathematical version of the concept of entropy, and also introduced its name. In fact, he used the notion of entropy in his attempt to give a mathematical interpretation of the second law of thermodynamics.

About ten years after the definition and formula of entropy given by R. Clausius, L. Boltzmann introduced a variant of entropy in his kinetic theory of gases [14]. His formula for entropy expresses that the entropy is proportional to the logarithm of probability, and later became known as Boltzmann entropy.

Boltzmann entropy can be thought of as a measure associated with a state of disorder, randomness or uncertainty. The more chaotic a system is, the more uniform the distribution of the macroscopic state will be. The Boltzmann entropy and the more general Clausius entropy agree for systems with slow changes in particle density, energy density, and momentum density on a microscopic scale.

Figure 1.1: The formula of Boltzmann entropy carved on his gravestone in the Central Cemetery of Vienna.

The equation $S = k \log W$ on L. Boltzmann's gravestone (see Figure 1.1) was not formulated in this form by L. Boltzmann, but by M. Planck [103]. He credited L. Boltzmann for the main idea of the formula. Moreover, he underlined his respect by using the symbol "W", which stands for the German word "Wahrscheinlichkeit" meaning probability and referring to the probability of a microscopic state.

In 1948, C.E. Shannon extended the concept of entropy from statistical physics to information communication, thereby laying one of the key foundations for information theory [108]. Although the mathematical concept of information predates Shannon entropy, C.E Shannon combined this concept with ideas originating from ergodic theory and random coding, to give a measure for the transmission of information from one place to another.

To explain the development from Shannon entropy to graph entropy, the paper *Three great challenges for half-century-old computer science* [26] by F.P. Brooks plays a crucial role. In his paper, which appeared in the Journal of the ACM in 2003, this recipient of the National Medal of Technology (in 1985) and the Turing Award (in 1999) posed the following challenge:

> *"We have no theory, however, that gives us a metric for the information embodied in structure, especially physical structure. We know that an automobile is a more complex structure than a rowboat. We cannot yet say it is x times more complex, where x is some number. Yet we know that the complexity is related to the Shannon information that would be required to specify the structures of the car and the boat."*

This quotation addresses the challenge to develop good measures for the quantification of structural information. F.P. Brooks believed that such a missing metric was the fundamental gap in theory between information science and computer science.

Moreover, further recent developments have made it timely to address this issue. The encoding of genetic information by DNA is apparently simple enough to be handled by existing theories of communication. The way in which the four amino acids can form pairs determines the acid sequence in the DNA protein. Despite this fact that proteins have a relatively simple basic structure, the way they are folded determines their functionality to a large extent. So entropy and energy considerations are probably not powerful enough to explain and predict their functional structure.

Nevertheless, defining and using several different variants of graph entropy seems a good starting point for developing measures for quantifying the complexity of structures that can be modeled by graphs. These variants of graph entropy will be introduced in Section 1.3, after presenting the relevant graph-theoretical terminology and notation in Section 1.1, and more historical annotations in Section 1.2.

This introductory chapter will end with a short description of the main results of this thesis in Section 1.4. Particular questions that will be addressed are: what are the extremal values of the studied variant of graph entropy? Can we characterize the graphs attaining the maximum or minimum value of this graph entropy? Such questions are typical for the field of extremal graph theory.

1.1 Terminology and notation

Before we start with the formal terminology and notation, let us try to make intuitively clear how graphs come into play when considering structures like molecules.

A graph is a mathematical object consisting of a finite set of vertices that can be interpreted as abstractions of atoms in a molecule, and a finite set of edges that can be interpreted as abstractions of bonds between pairs of atoms in a molecule. In this graph model, each edge represents a bond between a pair of atoms, and multiple bonds between the same pair of atoms are represented by multiple edges between the corresponding pair of vertices in the graph.

With the above in mind, in this section we introduce some basic terminology and notation. All graphs considered in this thesis are finite, undirected and without loops. We use standard graph-theoretic terminology and notation, as can be found in the textbook of Bondy and Murty [22].

A *graph* G is an ordered pair $(V(G), E(G))$ consisting of a *vertex set* $V(G)$ and an *edge set* $E(G)$, together with an *incidence function* ψ_G that associates with each edge of G an unordered pair of vertices of G. If e is an edge and u and v are vertices such that $\psi_G(e) = uv$, then e is said to *join* u and v, and u and v are also called *adjacent*. We say that u and v are the *end-vertices* of e, and that e is *incident* with u and v. If G is a *simple* graph, i.e., contains no multiple edges, then we can avoid the use of ψ_G. We use uv or vu to indicate the unique edge e with end-vertices u and v. In this thesis, a graph may contain multiple edges, but we assume it is simple unless otherwise indicated.

For a vertex $v \in V(G)$, we use $N_G(v)$ to denote the *set of neighbors* of v in G, i.e., the set of all vertices adjacent to v. The cardinality of $V(G)$ and $E(G)$ are the *order* and *size* of G, respectively. The *degree* of a vertex v in G, denoted by $\deg_G(v)$, is the number of edges incident with v in G. We also use $\deg(v)$ to denote the degree of v in G if there is no ambiguity. A vertex with degree 0 is called an isolated vertex. We use $\deg_{\max}(G)$ and $\deg_{\min}(G)$ to denote the maximum degree and the minimum degree among the vertices of G, respectively.

A graph is *connected* if, for every partition of its vertex set into two nonempty sets A and B, there is an edge with one end-vertex in A and one end-vertex in B; otherwise the graph is *disconnected*. A *coloring* of a graph G is a partition of its vertex set into independent sets. Here an *independent* set is a set of mutually nonadjacent vertices. The *chromatic number* of G, denoted by $\chi(G)$, is the smallest number of classes in any coloring of G.

Two graphs G and H are said to be *isomorphic* if there exist two bijections $\theta : V(G) \to V(H)$ and $\phi : E(G) \to E(H)$ such that $\psi_G(e) = uv$ if and only if $\psi_H(\phi(e)) = \theta(u)\theta(v)$. An *automorphism* of a graph G is an isomorphism of G to itself. The collection of automorphisms forms a group $Aut(G)$ under composition. Vertices u and v are said to be *similar* if there exists an automorphism mapping u to v. An *orbit* of $Aut(G)$ is the set of all vertices similar to a given vertex. The collection of orbits constitutes a partition of $V(G)$.

We next introduce some special families of graphs. A *path* of length n, denoted by P_{n+1}, is an alternating sequence of distinct vertices and edges $v_0e_1v_1 \cdots v_{n-1}e_nv_n$, such that the vertices v_{i-1} and v_i are the end-vertices of the edge e_i for all $i \in \{1, 2, \ldots, n\}$. A *tree* is a connected graph in which any two distinct vertices are connected by a unique path. A graph is said to be a *complete graph* and denoted by K_n if all its n vertices are pairwise adjacent. A graph G is called *bipartite* if $V(G)$ can be partitioned into two disjoint sets X and Y such that every edge of G has one end-vertex in X and one end-vertex in Y. A bipartite graph with bipartition (X, Y) satisfying $xy \in E$ for any pair in $\{(x, y) : x \in X, y \in Y\}$ is a *complete bipartite graph* and is denoted by $K_{s,t}$ if $|X| = s$ and $|Y| = t$. In particular, the complete bipartite graph $K_{1,n-1}$ is also called a *star*.

Let G and H be two graphs. Then H is called a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A *spanning subgraph* of G is a graph obtained by only deleting edges of G. A *spanning tree* of G is a spanning subgraph that is a tree. An *induced subgraph* of G is a subgraph obtained from G by deleting a number of vertices and all the edges incident with these vertices.

A *perfect graph* is a graph in which the chromatic number of every induced subgraph equals the order of a largest clique of that subgraph. Here a *clique* is a set of mutually adjacent vertices.

The *distance* between two vertices u and v, denoted by $d(u,v)$, is the length of a shortest path from u to v. If there is no path between u and v, then we set $d(u,v) := \infty$. The *eccentricity* of a vertex v, denoted by $\mathrm{ecc}(v)$, is the maximum distance from v to any other vertex (i.e., $\mathrm{ecc}(v) = \max\{d(v,u) : u \in V\}$). The *diameter* $\mathrm{diam}(G)$ of G is the maximum eccentricity among all vertices of G. For a vertex $v \in V$, the *j-sphere* of v is the set of vertices at distance j from v denoted by $S_j(v,G)$ (i.e., $S_j(v,G) = \{u \in V : d(u,v) = j\}$).

We define two graph operations. The *complement* $\overline{G}$ of $G = (V,E)$ is the graph with $E(\overline{G}) = E(K_{|V(G)|}) \setminus E(G)$ and $V(\overline{G}) = V(G)$. The *conormal product* (also known as disjunction) $G \vee G'$ of two disjoint graphs G and G' is the graph with vertex set $V(G) \times V(G')$ in which $uv \in E(G)$ or $u'v' \in E(G')$ produces an edge $(u,u')(v,v') \in E(G \vee G')$. Let $G^{\vee n}$ be the n-th *conormal power* of G with vertex set $V(G^{\vee n}) = \underbrace{V \times V \times \cdots \times V}_{n} = V^n$ and edge set

$$E(G^{\vee n}) = \{(u,v) \in V^n \times V^n : u = (u_1,u_2,\ldots,u_n), v = (v_1,v_2,\ldots,v_n) \text{ and}$$

$$\exists\ i : u_i v_i \in E(G)\}$$

We also need to define several matrices associated with graphs. For this purpose, let G be a graph with vertex set $\{v_1,v_2,\ldots,v_n\}$. The *distance matrix* of G, denoted by $D(G)$, is an $n \times n$ matrix, the (i,j)-entry of which is $[D(G)]_{i,j} = d(v_i,v_j)$. The *adjacency matrix* of G is an $n \times n$ matrix, denoted by $M(G)$, the (i,j)-entry of which is defined as follows

$$[M(G)]_{i,j} = \begin{cases} 1, & \text{if } v_i v_j \in E(G); \\ 0, & \text{if } v_i v_j \notin E(G). \end{cases}$$

The *degree matrix* of G is an $n \times n$ matrix, denoted by $D_e(G)$, the (i, j)-entry of which is defined as follows

$$[D_e(G)]_{i,j} = \begin{cases} \deg(v_i), & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases}$$

The *combinatorial Laplacian matrix* (for short, Laplacian) of G is the matrix

$$L(G) = D_e(G) - M(G).$$

If G has size m, then the *density matrix* of G is the matrix

$$\rho(G) = \frac{1}{2m} L(G).$$

Before continuing with the formal introduction of several variants of graph entropy, in the next section we present more details on the history of entropy. Readers with a major focus on the mathematical aspects or a main interest in the results of this thesis can safely skip the next section.

1.2 History of entropy

From thermodynamics to information theory, entropy plays an important role in their development.

The discovery of the laws of thermodynamics is closely related to the study of improving the efficiency of heat engines. The steam engine was invented in the 18th century, but it took a long time from its inception to its widespread use. The French physicist, mathematician and inventor D. Papin, one of the founders of thermodynamics, worked with R. Boyle from 1676 to 1679. During this time, he invented the steam digester which inspired the later invention of the pressure cooker and the steam engine. It was the English inventor T. Newcomen who developed the atmospheric engine in 1712, as the first practical fuel-burning engine. The steam pump, invented by T. Savery, an English inventor and engineer, is regarded as a model of the Newcomen steam engine. Nearly 70 years later, J. Watt advanced T. Newcomen's steam engine, which laid the foundation for the Industrial Revolution.

In the years that followed, many efforts were made to improve the efficiency of these steam engines. In 1824, a young French engineer named S. Carnot wrote his only published book *Reflections on the Motive Power of Fire* [31]. In this book, he proposed the concept of the Carnot heat engine and the Carnot cycle. He also argued that only heat is not enough to give impetus, but cold must be present. Without cold, heat will be useless. He believed that just as the flow of water could drive a water mill, the flow of a hot substance across a temperature difference (from a heat source to a so-called cold sink) could also lead to the completion of mechanical work. It was this idea that led S. Carnot to equate "the motion of heat" with "the flow of heat" (from high to low) and from this to derive "the unidirectional nature of the motion of heat". S. Carnot's work was unread and unacknowledged during his lifetime. But his work was later restated by R. Clausius and W. Thomson, 1st Baron Kelvin, and was an important basis for establishing the formal definition of entropy as the second law of thermodynamics. *Reflections on the Motive Power of Fire* [31] also became the symbol of thermodynamics as a modern science. In 1832, S. Carnot unfortunately caught cholera and died in Paris on August 24, at the age of 36.

The German physicist R. Clausius was the first to rigorously formulate the laws of thermodynamics. In 1850, R. Clausius published a paper [36] discussing the significance of unidirectional restriction of spontaneous thermal motion. He declared that "there is no process whose sole result is a transfer of energy from a cooler to a hotter environment". Since heat can only flow spontaneously from hot to cold, the opposite process (sometimes called refrigeration) requires work. W. Thomson believed that in any thermodynamic process, there is always some degradation of available energy. This belief inspired R. Clausius to develop a concept that he defined as "the direction of spontaneous change". He called this concept "entropy", which he described as "the property of a system that can measure the degradation of energy availability associated with spontaneous changes". This definition is sometimes called the "thermodynamic definition of entropy". R. Clausius explained that entropy in the system does not decrease because of spontaneous changes. The entropy in the system can only increase or stay the same. Therefore, spontaneous changes will continue to occur until entropy reaches a maximum.

J.C. Maxwell, a Scotsman, was a prominent theoretical physicist and a big believer in the atomic theory of matter. J.C. Maxwell believed (as did many before him) that, in any system, the microscopic particles in the system must obey Newton's laws of motion. The information in [123] tells us that, from 1859 to 1866, J.C. Maxwell developed the theory of the distributions of velocities in particles of a gas. He connected thermodynamics with mechanics. He explained that while it is obviously impossible to determine the motion of individual microscopic particles because of their infinitesimal size and large number, the nature of "the system as a whole" can be determined by using statistical averages.

In 1877, the Austrian physicist L. Boltzmann [14], inspired by J.C. Maxwell's theory of dynamics, equated the overall state of a system with the internal activities of atoms. He came up with another thermodynamic definition of entropy, and he used mathematical probability to define entropy: entropy is a measure of the probability of finding any given system configuration. This definition is called the "statistical definition of entropy". According to L. Boltzmann's definition, entropy increases until it reaches a maximum. This maximum is determined as the point at which the atoms of the system adopt the most random arrangement (no additional activity can further increase the degree of randomness of the system). The maximum value of the Boltzmann entropy can be calculated by entering the probability of the Boltzmann formula. This definition of entropy leads to another way of looking at the second law of thermodynamics, which implies that the states of any system will always shift from less likely to more likely states, and will continue to do so until the most likely state is reached. It is this interpretation of the second law of thermodynamics that eventually leads to the common understanding that entropy increases as the system spontaneously moves from order to disorder.

The next quote from N. Wiener in his *Cybernetics: or Control and Communication in the Animal and the Machine* [121] points out the important position of information: "Information is information, not matter or energy. No materialism which does not admit this can survive at the present." The profound meaning of this sentence is that material, energy and information are the three elements that constitute the objective world. To extend L. Boltzmann's interpretation of entropy to the transmission of information, one way to think about it is that when there are multiple possible scenarios for something, the uncertainty about which of those scenarios is specific to someone is called entropy. The thing that removes someone's uncertainty about the matter is called information. The measure used to describe the uncertainty of an event should have the characteristic that when the event is perfectly certain, the value of the measure should be zero. The more possible states or consequences of something, the greater the value of the measure should be. When the possible outcomes are known and the probability of each outcome is equal, the uncertainty reaches a maximum, that is, this kind of event is the most uncertain. As a measure of uncertainty, the Shannon entropy [108] meets the requirements of an uncertainty measure as proposed above.

In the next section, we start with the definition of the Shannon entropy, followed by a description of a large number of graph entropies which have been introduced over the years, based on the Shannon entropy.

1.3 Graph entropies

All variants of graph entropies are based on the same original ideas and concept due to Shannon. In [108], he defined what is now known as the Shannon entropy of a discrete random variable X as

$$-\sum_{i=1}^{n} p_i \log_b p_i,$$

where the possible outcomes x_i of X occur with probability p_i for $i = 1, 2, \ldots, n$, and where b denotes the base of the logarithm. The most commonly used value for b is 2. Throughout the thesis, we use log to denote $\log_2$.

To get from the Shannon entropy to a graph entropy, a simple and natural idea is to replace the probabilities p_i in the above expression by fractions which add up to one. For defining such fractions one might use any of the graph parameters or invariants introduced in Section 1.1. Indeed, many different graph entropies have been introduced this way, as we will see in this section.

With such graph-theoretical interpretations of the Shannon entropy, leading-edge investigations on graph entropy are conducted at the intersection of information science, biology, physics, mathematics and chemistry. A limitation of all the existing measures is that structurally non-equivalent graphs may result in the same value of the chosen graph entropy. Therefore, many different variants of graph entropy have been introduced and studied. According to different research objectives, different measures have emerged. We classify some important graph entropies into:

- global graph entropies, e.g., Körner entropy [79, 111], von Neumann entropy [24], degree-based entropy [29] and distance-based entropies [21, 43, 41];
- local graph entropies, e.g., local vertex entropies[45, 50].

In the following two subsections, we give an overview of some important graph entropies according to the above classification. The section ends with some applications of graph entropy.

1.3.1 Global entropies

N. Rashevsky was an American theoretical physicist, one of the pioneers of mathematical biology, and is considered the father of mathematical biophysics and theoretical biology. He founded the *Bulletin of Mathematical Biophysics*, the first international journal of mathematical biology. In 1955, N. Rashevsky proposed the concept of the structural information content of graphs, considered as the first graph entropy, in biological applications [104]. Such a measure is based on the Shannon entropy defined by C.E. Shannon [108] and related to the probability distribution of the partition induced by the equivalence relation implied by the automorphism group of a graph. A year later, E. Trucco [115] gave the formal definition of N. Rashevsky's structural information content. Let G be a graph of order n, and let A_i for $i = 1, \ldots, k$ denote the orbits of G. These orbits define a finite probability distribution $\mathbb{P} = (p_1, p_2, \ldots, p_k)$, where $p_i = \frac{|A_i|}{n}$. Then the *structural information content* of G is given by

$$\sum_{i=1}^{k} p_i \log(p_i).$$

In 1968, A. Mowshowitz [97, 98, 99, 100] successively published four further papers on this structural information content. In [97], A. Mowshowitz gave some behaviour of undirected graphs under different graph operations such as complement, sum, join, Cartesian product and composition. In [98], he extended these results to directed graphs. In [99], he gave a construction algorithm to investigate the properties of directed graphs with zero information content. In addition, an algorithm to compute the automorphism group of digraphs and to find the condition content to ensure that two digraphs have the same information was presented [99]. In [100], A. Mowshowitz proposed the chromatic information content defined as follows, and compared it with the previous measure. Let G be a graph of order n. Let $\mathscr{P} = \{V_1, V_2, \ldots, V_\chi(G)\}$ be a partition of $V(G)$ such that V_i is an independent set for $i = 1, 2, \ldots, \chi(G)$, $\cup_{i=1}^{k} V_i = V(G)$ and $V_s \cap V_t = \emptyset$ for $s \neq t$. The chromatic information content of G is

$$\min_{\mathscr{P}} \left\{ -\frac{|V_i|}{n} \log\left(\frac{|V_i|}{n}\right) \right\},$$

where $\mathscr{P} = \{V_1, V_2, \ldots, V_\chi(G)\}$ ranges over all possible partitions of $V(G)$ satisfying the corresponding conditions of $\mathscr{P}$.

D. Bonchev and N. Trinajstić [21] discussed entropy measures based on graph distance. Before introducing these measures, we first recall the definition of the information content of a given partition of a set of n elements, as defined in [25]. Let $\mathscr{S} = \{S_1, S_2, \ldots, S_k\}$ be a partition of an n-element set S such that $\cup_{t=1}^{k} S_t = S$ and $S_i \cap S_j = \emptyset$ for $i \neq j$. The definition of the information content of the partition $\mathscr{S} = \{S_1, S_2, \ldots, S_k\}$ of n-element set S is

$$n \log(n) - \sum_{i=1}^{k} |S_i| \log(|S_i|).$$

According to the above information content and the Shannon entropy, the authors in [21] defined two measures of information on distances of a given graph. Let G be a connected graph with vertex set $\{v_1, v_2, \ldots, v_n\}$. Let $2n_i$

be the number of times that i appears in the distance matrix $D(G)$ with $i = 1, 2, \ldots, k$. Let $p_0 = \frac{n}{n^2} = \frac{1}{n}$ and $p_i = \frac{2n_i}{n^2}$ for $i = 1, 2, \ldots, k$. Two measures of information on distances of G in [21] are

$$n^2 \log(n^2) - n \log(n) - \sum_{i=1}^{\mathrm{diam(G)}} 2n_i \log(2n_i) \tag{1.1}$$

and

$$-\frac{1}{n} \log\left(\frac{1}{n}\right) - \sum_{i=1}^{\mathrm{diam(G)}} \frac{2n_i}{n^2} \log\left(\frac{2n_i}{n^2}\right). \tag{1.2}$$

It is not hard to find that equation (1.1) is n^2 times equation (1.2). Among all graphs of order n, graphs minimize (resp., maximize) equation (1.1) also minimize (resp., maximize) equation (1.2).

As a generalized distance entropy, M. Dehmer [43, 41] proposed a definition of graph entropy based on the j-spheres of the graph. Let G be a graph with vertex set $\{v_1, v_2, \ldots, v_n\}$, and let $S_j(v_i, G)$ denote the j-sphere of v_i in G. He defined a graph entropy based on the j-spheres of G by

$$I_{s_1}(G) = -\sum_{i=1}^{n} \left(\frac{\sum_{j=1}^{\mathrm{diam}(G)} \alpha^{c_j |S_j(v_i,G)|}}{\sum_{t=1}^{n} \sum_{j=1}^{\mathrm{diam}(G)} \alpha^{c_j |S_j(v_t,G)|}} \right) \log \left(\frac{\sum_{j=1}^{\mathrm{diam}(G)} \alpha^{c_j |S_j(v_i,G)|}}{\sum_{t=1}^{n} \sum_{j=1}^{\mathrm{diam}(G)} \alpha^{c_j |S_j(v_t,G)|}} \right),$$

for some fixed real numbers $\alpha > 0$ and $c_j > 0$ with $j = 1, 2, \ldots, \mathrm{diam}(G)$.

Another graph entropy of G taking into account all j-spheres was defined in [47, 84] by

$$I_{s_2}(G) = -\sum_{i=1}^{n} \left(\frac{\sum_{j=1}^{\mathrm{diam}(G)} c_j |S_j(v_i, G)|}{\sum_{t=1}^{n} \sum_{j=1}^{\mathrm{diam}(G)} c_j |S_j(v_t, G)|} \right) \log \left(\frac{\sum_{j=1}^{\mathrm{diam}(G)} c_j |S_j(v_i, G)|}{\sum_{t=1}^{n} \sum_{j=1}^{\mathrm{diam}(G)} c_j |S_j(v_t, G)|} \right),$$

for some fixed real numbers $c_j > 0$ with $j = 1, 2, \ldots, \mathrm{diam}(G)$.

If we relax the condition of c_j from greater than zero to greater than or equal to zero, then for $c_1 = 1$ and $c_j = 0$ for $j = 2, 3, \ldots, \mathrm{diam}(G)$, this leads to the degree-entropy $I_d(G)$ defined in [29]. Let G be a graph with vertex set

$V = \{v_1, v_2, \dots, v_n\}$ and size m. A simpler expression for $I_d(G)$ is

$$I_d(G) = -\sum_{i=1}^{n} \frac{\deg(v_i)}{2m} \log\left(\frac{\deg(v_i)}{2m}\right).$$

In the literature [47, 84], the following entropy regarding the eccentricity of G has been defined by

$$I_e(G) = -\sum_{i=1}^{n} \frac{c_i \mathrm{ecc}(v_i)}{\sum_{j=1}^{n} c_j \mathrm{ecc}(v_j)} \log\left(\frac{c_i \mathrm{ecc}(v_i)}{\sum_{j=1}^{n} c_j \mathrm{ecc}(v_j)}\right),$$

for some fixed real numbers $c_i > 0$ and $i = 1, 2, \dots, n$. For $c_i = 1$, the entropy is called eccentricity-entropy. In [47, 84], the authors proved some extremal results on this entropy and proposed some related conjectures.

Related to graph invariants involving vertices, M. Dehmer unified some graph entropies by using a so-called information functional [42]. Let G be a graph with vertex set $V = \{v_1, v_2, \dots, v_n\}$. Let $f : V \to \mathbb{R}_{\geq 0}$ be an arbitrary information functional. Then the entropy of G related to f is defined by

$$I_f(G) = -\sum_{i=1}^{n} p_i \log(p_i),$$

where

$$p_i = \frac{f(v_i)}{\sum_{j=1}^{n} f(v_j)}$$

for $i = 1, 2, \dots, n$. In [45], the authors gave some special examples of the above defined entropy associated with different information functionals. In [29, 43, 44, 119, 118, 120], the interested reader can find some of the more recent examples. In [49], M. Dehmer and A. Mowshowitz derived a number of inequalities involving graph entropies based on an information functional.

Related to spectral properties of graphs, S.L. Braunstein, S. Ghosh and S. Severini [24] proposed the von Neumann entropy of a graph G, which is based on the eigenvalues of the density matrix $\rho(G)$. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of $\rho(G)$. Then the von Neumann entropy of G is defined by the formula

$$I_{von}(G) = \sum_{i=1}^{n} \lambda_i \log(\lambda_i).$$

K. Anand and G. Bianconi [8] showed that the Shannon entropy is linearly related to the von Neumann entropy for so-called scale-free networks.

We end this subsection by presenting three versions of the Körner entropy. These versions look somewhat different but are in fact equivalent. The following version, which was put forward in [39], is not the first one but probably the easiest one to understand. Let $G = (V, E)$ be a graph. The *vertex packing polytope* of G, denoted by $VP(G)$, is the convex hull of the characteristic vectors of the independent sets of G. Let $\mathbb{P} = (p_1, p_2, \ldots, p_n)$ be a probability distribution on $V(G)$, and let $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in VP(G)$ be a vector. The Körner entropy of G with respect to $\mathbb{P}$ is

$$I(G, \mathbb{P}) = \min_{\mathbf{a} \in VP(G)} \sum_{i=1}^{n} p_i \log\left(\frac{1}{a_i}\right).$$

The other two versions of the Körner entropy have appeared in [79]. The next one is related to the concept of mutual information. For two random variables X and Y, their mutual information is defined as

$$I(X;Y) = \sum_{x \in X, y \in Y} p_{(X,Y)}(x,y) \log\left(\frac{p_{(X,Y)}(x,y)}{p_X(x) p_Y(y)}\right),$$

where $p_{(X,Y)}$ is the joint probability mass function of X and Y, and p_X and p_Y are the marginal probability mass functions of X and Y, respectively. Let variable X take its values on $V(G)$ and Y on the independent sets of G, and let their joint distribution satisfy $X \in Y$ having probability 1. The marginal distribution of X is given by a probability distribution $\mathbb{P}$ on $V(G)$. The Körner entropy of G regarding $\mathbb{P}$ is defined by

$$I(G, \mathbb{P}) = \min I(X;Y).$$

J. Körner [79] defined the following version as the original Körner entropy. Let $\mathbb{P}$ be a probability distribution on $V(G)$. Define the probability distribution

p^n on V^n by $p^n(x) = \prod_{i=1}^{n} p(x_i)$. Let $0 < \epsilon < 1$ be a real number and

$$T_\epsilon^{\vee n} = \{U \subseteq V^n : \sum_{\mathbf{x} \in U} p^n(\mathbf{x}) > 1 - \epsilon\}.$$

For $U \subseteq V^n$, let $G^{\vee n}[U]$ denote the subgraph induced by U in $G^{\vee n}$, where $G^{\vee n}$ is the n-th conormal power of G. Then the Körner entropy $I(G, \mathbb{P})$ is defined by

$$I(G, \mathbb{P}) = \lim_{n \to \infty} \min_{U \in T_\epsilon^{\vee n}} \frac{1}{n} \log\left(\chi(G^{\vee n}[U])\right).$$

For more details and for proofs of the equivalence of these three versions of the Körner entropy, we refer the reader to [39, 79, 110].

The above exposition gives just a small impression of the many variants of global graph entropies which have been introduced and studied over the last seventy years, since the first concepts of so-called information content or entropy were introduced in the 1950s by N. Rashevsky [104] and E. Trucco [115]. We refer the interested reader to the two survey papers [48, 110] and the two books [19, 46] for more information.

In the next subsection, we give a few examples of local graph entropies.

1.3.2 Local entropies

Let G be a connected graph with vertex set $V = \{v_1, v_2, \ldots, v_n\}$. In [45], M. Dehmer and F. Emmert-Streib defined a local entropy based on the distances from one vertex to all the other vertices, as follows:

$$I_\ell(v_i) = \sum_{j=1}^{n} \frac{c_j d(v_i, v_j)}{\sum_{t=1}^{n} c_t d(v_i, v_t)} \log\left(\frac{c_j d(v_i, v_j)}{\sum_{t=1}^{n} c_t d(v_i, v_t)}\right),$$

where $c_j > 0$ for $j = 1, 2, \ldots, n$. By setting $c_j = 1$ for $j = 1, 2, \ldots, n$, a special case of this local entropy was defined and studied in [78]. Two global distance-based entropies can be obtained by taking the average or the sum of the above expression over all vertices. Taking the average, one graph entropy of G is defined by the formula

$$\bar{I}_\ell(G) = \frac{\sum_{i=1}^{n} I_\ell(v_i)}{n}.$$

Taking the sum, the other graph entropy of G is defined by the formula

$$\hat{I}_\ell(G) = \sum_{i=1}^{n} I_\ell(v_i).$$

For the statistical analysis of certain chemical structures, M. Dehmer, K. Varmuza, S. Borgert and F. Emmert-Streib [50] explored several special cases of the above local entropy. For this purpose, they applied several different choices for the coefficients, including

- $c_1 = n, c_2 = n-1, \ldots, c_n = 1$ (linearly decreasing coefficients), and
- $c_1 = n, c_2 = ne^{-1}, \ldots, c_n = ne^{1-n}$ (exponentially decreasing coefficients).

These coefficients are depending on the number of vertices n.

Based on the j-spheres of a connected graph G of order n, E.V. Konstantinova and A.A. Paleev [78] defined a local entropy by the formula

$$I_s(v_i) = -\sum_{j=1}^{\mathrm{ecc}(v_i)} \frac{|S_j(v_i,G)|}{n} \log\left(\frac{|S_j(v_i,G)|}{n}\right).$$

The authors studied the sensitivity of these local entropies when applied to so-called polycyclic graphs. They also defined and studied the associated global graph entropy of G, obtained by taking the sum

$$\hat{I}_s = \sum_{i=1}^{n} I_s(v_i).$$

Since this thesis is focused on global graph entropies, we close this subsection on local entropies here, and turn to some applications. For more information on local graph entropies, we refer the interested reader to the book [113].

1.3.3 Applications

Since graphs can be used to model any set of objects and their pairwise relations, it will not come as a surprise that graph entropies have been applied in many different applications areas. In these applications, the entropy of a graph is usually interpreted as its structural information content and as a measure of its complexity. Most of such applications can be found in chemical graph theory, where the entropy is used in attempts to characterize the atomic structure and properties of molecules. A good source for more information on applications in chemistry is the book by D. Bonchev [16]. Here we start our exposition by highlighting some applications in other areas.

From 1973 to 1992, the Körner entropy which we have defined earlier has been widely studied. J. Körner [79] discovered that the subadditivity of graph entropy in connection with a problem in coding theory can be used to prove good bounds for graph covering problems. In his follow up paper [80], he showed that a result which is known as the Fredman-Komlós lemma is a consequence of a simple inequality between entropies of graphs. This in turn enabled him to handle more problems on separating partition systems, including problems related to hashing. A few years later, in a joint paper by J. Körner and K. Marton [82], the authors derived new bounds for perfect hashing based on an extension of graph entropy to hypergraphs. Their bounds improved the bounds obtained by M. Fredman and J. Komlós [60] for a complete family of hash functions. In another paper by J. Körner and K. Marton [81], it was proved that for a bipartite graph G and an arbitrary probability distribution $\mathbb{P}$ on its vertex set, the entropies of G and its complement add up to the entropy of $\mathbb{P}$. Their results have interesting connections with the well-known Ford–Fulkerson method in the theory of network flows.

Another equivalent version of the Körner entropy was defined in terms of the vertex packing polytope of the graph, and first appeared in [39]. Using this version, I. Csiszár, J. Körner, L. Lovász, K. Marton and G. Simony in [39] gave an information theoretic characterization of so-called perfect graphs. In [83], J. Körner, G. Simonyi and Z. Tuza gave the following alternative characterization of perfect graphs, based on the additivity of graph entropy: G is a perfect graph if and only if the entropy of the complete graph is equal to the sum of those of G and its complement. For more information on the relationship between graph entropy and perfect graphs, we refer the reader to the survey [111].

In another interesting application area related to computational complexity, I. Newman and A. Wigderson [101] used hypergraph entropy to derive lower bounds on the formula size of Boolean functions.

Our next applications of graph entropy fall within the area of information and communication theory, and start with another classic work due to C.E. Shannon [109]. In this paper, he formulated and studied the zero-error capability of a single-input single-output channel. For this, he represented a memoryless channel as a graph whose vertices are the input letters of an alphabet, and in which any two vertices are joined by an edge if they are not distinguishable (can be confused) at the output of the channel. In [109], C.E. Shannon dealt with all graphs on at most five edges, except the cycle on five vertices. For this remaining case, he only established a lower bound. This problem was not solved until 1979, when L. Lovász [93] proved that C.E. Shannon's lower bound was tight. In [93], other special graph classes are treated as well.

In the general setting, a sender wants to pass a piece of information accurately to a receiver who has some (possibly related) data. In [4], N. Alon and A. Orlitsky examined the expected number of bits that the sender must transmit in order to pass the information correctly. They consider single and multiple instances of two related communication scenarios. They show that the expected number of bits is related to the chromatic number of the graph which represents the data. Interestingly, they show that the Körner entropy of the graph gives a lower bound for the single instance case, and that it is precisely the asymptotic per-instance number of bits for the unrestricted-inputs scenario.

One of the reasons why there are so many different variants of graph entropy around is that it is impossible to capture the complexity or information content of graphs in one measure. Different applications usually require different types of structural information and naturally lead to different measures. In [11], it is stated that describing complexity quantitatively is a large and rapidly developing subject. Although various interpretations of the term have been proposed in different disciplines, no comprehensive discussion has been attempted. The intuitive notion of complexity is well expressed by the Collins English dictionary definition: "Complexity is the state of having many different parts connected or related to each other in a complicated way".

In the work of E.B. Allen [3], an information-theoretic approach has been used to measure the complexity of graph abstractions of software systems and modules in computer-related disciplines. In this approach, graphs are used for representing many abstractions of software at the system and module level, and extensions to hypergraphs are suggested as well. Graph entropy measures are proposed for measuring the size, length, complexity, coupling and cohesion of software engineering abstractions.

Graph entropy measures have also been applied to gene networks. In [7], G. Altay and F. Emmert-Streib used information theory technology to statistically analyze network inference algorithms for gene networks. They employ local network-based entropy measures to assess the performance of these algorithms. This initiated a graph-theoretic perspective on the problem, and enabled studying arbitrary network components instead of the entire network. Using this approach they compared four different network inference algorithms. They demonstrated that these measures allow an exploratory analysis of inference algorithms on the level of network components, e.g., edges, motifs or subnetworks. In addition, F. Emmert-Streib and M. Dehmer [57] explored information dissemination in gene networks by performing single gene knockouts. For information theoretical analysis of networks, K. Anand and G. Bianconi [8] discussed physics-based network entropies including the von Neumann entropy, and defined network integration.

Also within the field of robotics and multi-agent systems, graph entropy measures have proved to be useful and effective. In [13], T. Balch introduced the concept of hierarchic social entropy in order to quantify the diversity of robot teams. He also examined the correlation between these measures by using a collection of business processes that represent the network. In [87], A. Li and Y. Pan proposed that structural entropy is the first measure of the dynamical complexity of networks, measuring the complexity of interactions, communications, operations, and even the evolution of networks. In [88], by using structural entropy, a generalized degree-graph entropy, the authors proved that the modular structure of the genome spatial organization may be fundamental to even a small cohort of single cells. To classify some of the most important complexity measures, we recommend the reader to consider the following outline due to M. Dehmer [43]:

- classical information measures [15, 16, 20, 18, 21, 19, 61, 108, 112];
- entropic measures for characterizing graph classes [79, 111];
- information-theoretic measures to determine the structural information content of a network [16, 41, 97, 98, 99, 100, 104, 115];
- complexity measures for networks based on the principle of Kolmogorov-complexity [17, 20, 89];
- information-theoretic robustness measures for complex networks [55, 56];
- statistical correlation measures for structurally characterizing complex networks [112];
- simulated annealing methods to investigate network structures [112, 107].

It is clear from the previous sections that many different variants of graph entropy have been introduced and studied in many different application areas. We close this introductory chapter by presenting a short outline of our main contributions in the next section.

1.4 Outline of the main results of this thesis

Apart from this introductory chapter, this thesis consists of six technical chapters that are based on earlier submitted papers. In these six chapters, we mainly concentrate on determining extremal values of degree-based and distance-based entropies restricted to certain graph classes. We also consider some related problems, such as the complexity of spanning tree problems involving graphical indices, including the degree-based graph entropy. The remainder of this thesis is organized as follows.

In Chapter 2, we consider the effect of graph operations on the value of the degree-entropy. We derive several new results, based on graph operations including concepts like the complement, the weak product, the blow-up and the identification of vertices.

In Chapter 3, we determine the minimum and maximum values of the degree-entropy among trees and unicyclic graphs with a given bipartition, respectively. We identify the corresponding extremal graphs. We also determine the minimum value of the degree-entropy among trees with a given diameter and characterize the extremal graphs.

In Chapter 4, we characterize the bipartite graphs that minimize the degree-entropy, among all bipartite graphs of a given size, or a given size and (upper bound on the) order. The extremal graphs turn out to be complete bipartite graphs, or nearly complete bipartite. Here we make use of an equivalent representation of bipartite graphs by means of Young diagrams, which makes it easier to compare the entropy of related graphs. We conclude that the general characterization of the extremal graphs is a difficult problem, due to its connections with number theory, but they are easy to find for specific values of the order and size. We also give a direct argument to characterize the graphs of a given order and size maximizing the entropy. We indicate how our ideas extend to other graphical function-indices as well. This implies the known result due to S. Cao, M. Dehmer and Y. Shi [29] that the path and star attain the maximum and minimum degree-entropy among trees of order n, respectively.

In Chapter 5, we study the extremal problems of finding the graphs attaining the minimum degree-entropy among graphs of a given order and size. We characterize the unique extremal graph achieving the minimum value. The extremal graphs turn out to be so-called threshold graphs.

In Chapter 6, we consider the Wiener-entropy, which is together with the eccentricity-entropy one of the most natural distance-based graph entropies. By deriving the (asymptotic) extremal behavior, we conclude that the Wiener-entropy of graphs of a given order is more spread than the eccentricity-entropy.

We resolve three known conjectures on the eccentricity-entropy and propose two new conjectures on the Wiener-entropy. These conjectures are reflecting some surprising behavior of the graphs minimizing it.

In Chapter 7, we consider the computational complexity of spanning tree problems involving the graphical function-index. This index was recently introduced by X. Li and D. Peng [90] as a unification of a long list of chemical and topological indices. We present a number of unified approaches to determine the $\mathscr{NP}$-completeness and $\mathscr{APX}$-completeness of maximum and minimum spanning tree problems involving this index. We give many examples of well-studied topological indices for which the associated complexity questions are covered by our results.

Chapter 2

Graph operations on degree-entropy

The results in this chapter deal with the effect of certain graph operations on the degree-entropy, which was introduced in [29]. Recall that this degree-entropy $I_d(G)$ of a graph G of size $m > 0$ is obtained from the Shannon entropy $-\sum_{i=1}^{n} p(x_i) \log p(x_i)$ by replacing the probabilities $p(x_i)$ by the fractions $\frac{\deg(v_i)}{2m}$, where $\{v_1, v_2, \ldots, v_n\}$ is the vertex set of G, and $\deg(v_i)$ is the degree of v_i in G. The effect of graph operations on some topological indices could be used to study the extremal problems on the corresponding indices. This motivates us to do some research of the effect of graph operations on the degree-entropy. In Chapter 3, we study extremal problems involving the degree-entropy restricted to trees and unicyclic graphs under some given parameters. This is mainly done by analyzing the effect of graph operations on the degree-entropy.

2.1 Introduction

We start this section with some background and preliminary results underpinning our research.

2.1.1 Background

Recalling that the degree of a vertex is the number of edges having this vertex as an end-vertex, it is obvious that each edge contributes 2 to the sum of the degrees taken over all vertices of G. Hence, the sum of the vertex degrees is equal to twice the number of edges of G, a folklore result that goes back to L. Euler [59], who proved this result in 1736. This result implies that in a graph G of size $m > 0$, the fraction $\frac{\deg(v)}{2m}$ is between 0 and 1 for every vertex v of G, and that the sum of these fractions taken over all vertices of G is equal to 1. As we discussed in the previous chapter, it is natural to replace the probabilities in the formula of the Shannon entropy by these fractions. This is the basic idea behind the following definition of the degree-entropy which we adopted from [29].

Let G be a graph with vertex set $V = \{v_1, v_2, \ldots, v_n\}$ and size $m > 0$. We recall that the *degree-entropy* of G, denoted by $I_d(G)$, is defined as

$$I_d(G) = -\sum_{i=1}^{n} \frac{\deg(v_i)}{2m} \log\left(\frac{\deg(v_i)}{2m}\right). \tag{2.1}$$

For later reference, we also define a function

$$h_d(G) = \sum_{i=1}^{n} \deg(v_i) \log(\deg(v_i)). \tag{2.2}$$

Straightforward calculations show that $I_d(G) = \log(2m) - \frac{1}{2m} h_d(G)$. This function $h_d(G)$ comes in handy if we want to compare $I_d(G)$ and $I_d(G')$ for two graphs G and G' of the same size, or if we want to obtain the minimum or maximum of $I_d(G)$, where G ranges over all members of a class of graphs of size m.

S. Cao, M. Dehmer and Y. Shi [29] were the first to study the extremal values of entropies based on degree powers for certain families of graphs, and in more detail for the entropy defined in (2.1). More recently, A. Ghalavand, M. Eliasi and A.R. Ashrafi [63] established the first maximum and minimum values of the entropy in (2.1) for families of trees and unicyclic graphs, by applying majorization techniques. We come back to this in Subsection 2.1.2, where we extend and apply one of their fundamental lemmas. In a very recent paper, J. Yan [125] investigated the extremal properties of this entropy for general graphs.

Our main results in the next section deal with the effect of certain graph operations on the value of the degree-entropy from (2.1). It has been demonstrated in many papers that graph operations can form an effective and valuable tool in determining extremal values of several topological indices. Examples of their benefit in obtaining these values have been illustrated with respect to the eccentric connectivity coindex [10], the first and second Zagreb indices [76], and the hyper-Wiener index [77]. For the degree-entropy, results in this direction are generally lacking. Our contributions are motivated by the above observations.

In the next subsection, we introduce some terminology and notation which will be used later.

2.1.2 Preliminaries

In this subsection, we give some additional terminology and notation, and we state and prove a number of lemmas which will be used in the proofs of our results.

Whenever we use the term graph in this chapter, we allow multiple edges. Let $G=(V,E)$ be a graph. If the number of vertices in G with degree d_i is a_i for $i=0,1,\dots,k$, then we denote by $D(G)=(\underbrace{d_k,\dots,d_k}_{a_k},\underbrace{d_{k-1},\dots,d_{k-1}}_{a_{k-1}},\dots,$ $\underbrace{d_1,\dots,d_1}_{a_1})$ the degree sequence of G in which $0=d_0<d_1<\cdots<d_k$ and $a_0+a_1+\cdots+a_k=|V(G)|$.

In the context of our research, majorization is a useful relationship between two non-increasing integer (degree) sequences $A=(a_1,a_2,\dots,a_n)$ and $B=(b_1,b_2,\dots,b_n)$. We say that *A majorizes B*, denoted by $A\succeq B$, if for all $k\in\{1,2,\dots,n-1\}$:

$$\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} b_i, \quad \text{and} \quad \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i.$$

If at least one of the above inequalities is strict, then we say the majorization is *strict*. We use $A \succ B$ to express that A strictly majorizes B.

A. Ghalavand, M. Eliasi and A.R. Ashrafi [63] proved the following result for simple graphs, but it is straightforward to extend their proof to graphs. We omit the proof.

Lemma 2.1. *Let G and G' be two graphs of the same order and size. If $D(G) \succeq D(G')$, then $I_d(G) \leq I_d(G')$, with equality holding in the inequality if and only if $D(G) = D(G')$.*

2.2 The effect of graph operations

We consider some unary operations, an operation with only one operand, in the following subsection.

2.2.1 Unary operations

Let $G = (V, E)$ be a graph. For a subset $S \subseteq E$, we use $G-S$ to denote the graph $(V, E \setminus S)$ (for which we restrict the incidence function ψ_G to $E \setminus S$ if necessary). Similarly, we use $G + F$ to denote the graph obtained from G by adding a set F of new edges incident with pairs of distinct vertices of G (possibly creating multiple edges and defining or extending the incidence function ψ_G in the obvious way). If $S = \{e\}$ or $F = \{e\}$, we use $G - e$ and $G + e$ as shorthand for $G - \{e\}$ and $G + \{e\}$, respectively. Similarly, we use $G - e + f$ as shorthand for $(G - e) + f$.

The following four results and their consequences deal with the effect of edge additions and edge deletions on the value of the degree-entropy $I_d(G)$ of (2.1).

Theorem 2.1. *Let u, v, w and x be four vertices of a graph G. Set $G' = G+e$ and $G'' = G + f$, in which $\psi_{G'}(e) = uv$ and $\psi_{G''}(f) = wx$. If $\deg_G(u) \geq \deg_G(w)$ and $\deg_G(v) \geq \deg_G(x)$, then $I_d(G') \leq I_d(G'')$, with equality holding in the latter inequality if and only if $\deg_G(u) = \deg_G(w)$ and $\deg_G(v) = \deg_G(x)$.*

The following known result is an easy consequence of Theorem 2.1.

Corollary 2.1 ([29]). *Let u, v and w be three vertices of a simple graph G. Suppose that u and v are adjacent, and w and v are not adjacent, and set $G' = G - uv + wv$. If $\deg_G(u) - \deg_G(w) \geq 2$, then $I_d(G) < I_d(G')$.*

In fact, since Theorem 2.1 holds for graphs with multiple edges, we can deduce the slightly stronger statement in which we assume e (with $\psi_G(e) = uv$) is one of possibly more than one edges joining u and v, and f (with $\psi_{G'}(f) = wv$) is a new edge of G' joining the possibly already adjacent vertices w and v of G. We also immediately obtain the following result involving the deletion and addition of pendant edges.

Corollary 2.2. *Let u, v and w be three vertices of a graph G. Suppose that $\deg_G(v) = 1$, u and v are adjacent, and w and v are not adjacent. Set $G' = G - e + f$, in which $\psi_G(e) = uv$ and $\psi_{G'}(f) = wv$. Then $I_d(G) \leq I_d(G')$ if and only if $\deg_G(u) > \deg_G(w)$.*

In our next result, we compare the effect of adding one edge between two different pairs of vertices with the same degree sum.

Theorem 2.2. *Let s be a positive integer, and let u, v, w and x be four vertices of a graph G. Suppose that $\deg_G(u) \geq \deg_G(v) \geq 1$, $\deg_G(w) \geq \deg_G(x) \geq 1$, and $\deg_G(u) + \deg_G(v) = \deg_G(w) + \deg_G(x) = s$. Set $G' = G + e$ and $G'' = G + f$, in which $\psi_{G'}(e) = uv$ and $\psi_{G''}(f) = wx$. If $\deg_G(u) - \deg_G(v) \leq \deg_G(w) - \deg_G(x)$, then $I_d(G') \leq I_d(G'')$, with equality holding in the latter inequality if and only if $\deg_G(u) = \deg_G(w)$.*

In our next subsection, we consider a number of more global operations on a graph.

2.2.2 Binary operations

The firstoperationis the so-called *k-blow up* of G, denoted by $G^{(k)}$. This is the graph obtained by replacing every vertex v of G with $k > 0$ distinct copies, and joining every copy of u to every copy of v in $G^{(k)}$ with ℓ_{uv} edges if and only if there are ℓ_{uv} edges joining u and v in G. We deduce the following expression for the degree-entropy of $G^{(k)}$.

Theorem 2.3. *Let G be a graph with at least one edge, and let $k \geq 1$ be an integer. Then $I_d(G^{(k)}) = I_d(G) + \log(k)$.*

In the next result, we consider the graph $G/\{x, y\}$ obtained from a graph G by *identifying* two distinct nonadjacent vertices x and y, i.e., replacing x and y by a single new vertex z and making z incident to all edges that were incident to x or y (possibly creating multiple edges). The following result shows that the degree-entropy decreases if two distinct nonadjacent vertices are identified.

Theorem 2.4. *Let G be a graph with at least one edge. Suppose that x and y are two distinct nonadjacent vertices of G. Set $G' = G/\{x, y\}$. Then $I_d(G) > I_d(G')$ if and only if $\deg_G(x) > 0$ and $\deg_G(y) > 0$.*

We next consider the identification of two vertices $x \in V(G)$ and $z \in V(H)$ from disjoint graphs G and H, resulting in the graph denoted as $GxHz$. We observe the following effect of the degree of x on the degree-entropy of $GxHz$.

Theorem 2.5. *Let G and H be two disjoint graphs. Suppose that x and y are two vertices of G, and z is a non-isolated vertex of H. If $\deg_G(x) \geq \deg_G(y)$, then $I_d(GxHz) \leq I_d(GyHz)$, with equality holding in the latter inequality if and only if $\deg_G(x) = \deg_G(y)$.*

Our final result of this subsection deals with the *weak product* (also known as *tensor product* or *Kronecker product*) $G \times G'$ of two disjoint graphs G and G'. This is the graph with vertex set $V(G) \times V(G')$, in which every pair of edges $f \in E(G)$ with $\psi_G(f) = uv$ and $f' \in E(G')$ with $\psi_{G'}(f') = u'v'$ produces two edges $e_1, e_2 \in E(G \times G')$ with $\psi_{G\times G'}(e_1) = (u, u')(v, v')$ and $\psi_{G\times G'}(e_2) = (v, u')(u, v')$ (in other words, either $(u, u')(v, v')$ or $(v, u')(u, v')$ has multiplicity ab if uv has multiplicity a and $u'v'$ has multiplicity b).

We deduce the following nice relationship between the degree-entropy of $G \times G'$, G and G'.

Theorem 2.6. *Let G and G' be two disjoint graphs with at least one edge. Then $I_d(G \times G') = I_d(G) + I_d(G')$.*

2.3 Proofs

In this final section, we gathered all the missing proofs of the statements in earlier sections.

Proof of Theorem 2.1. Without loss of generality, we assume that $\deg_G(u) \geq \deg_G(v)$. If $\deg_G(u) = \deg_G(w)$ or $\deg_G(v) = \deg_G(x)$, then we have $D(G') \succeq D(G'')$. Using Lemma 2.1, we get $I_d(G') \leq I_d(G'')$, with equality holding in this inequality if and only if $\deg_G(u) = \deg_G(w)$ and $\deg_G(v) = \deg_G(x)$. We consider $\deg_G(u) > \deg_G(w)$ and $\deg_G(v) > \deg_G(x)$ in the following. It follows that $D(G') \succ D(G'')$ if $\deg_G(u) = \deg_G(v)$. Then, using Lemma 2.1 again, we have $I_d(G') < I_d(G'')$. We only prove the case $\deg_G(u) > \deg_G(w) > \deg_G(v) > \deg_G(x)$, since the other cases $\deg_G(u) > \deg_G(v) = \deg_G(w) > \deg_G(x)$, $\deg_G(u) > \deg_G(v) > \deg_G(w) > \deg_G(x)$, $\deg_G(u) > \deg_G(v) > \deg_G(w) = \deg_G(x)$, and $\deg_G(u) > \deg_G(v) > \deg_G(x) > \deg_G(w)$ can be proved similarly. Let us relabel the vertices of the graph G as $v_1, v_2, \ldots, v_n$ such that $\deg_G(v_1) \geq \deg_G(v_2) \geq \cdots \geq \deg_G(v_n)$. Suppose that $v_i = u$, $v_j = w$, $v_s = v$ and $v_r = x$. We may assume $i < j < s < r$ in which

$$i = \min\{l \,|\, \deg_G(v_l) = \deg_G(u)\},$$

$$j = \min\{l \,|\, \deg_G(v_l) = \deg_G(w)\},$$

$$s = \min\{l \,|\, \deg_G(v_l) = \deg_G(v)\},$$

and

$$r = \min\{l \,|\, \deg_G(v_l) = \deg_G(x)\}.$$

For each $k \in \{1, 2, \ldots, i-1\}$, we have

$$\sum_{t=1}^{k} \deg_{G'}(v_t) = \sum_{t=1}^{k} \deg_{G''}(v_t);$$

for each $k \in \{i, i+1, \ldots, j-1\}$, we have

$$\sum_{t=1}^{k} \deg_{G'}(v_t) > \sum_{t=1}^{k} \deg_{G''}(v_t);$$

for each $k \in \{j, j+1, \ldots, s-1\}$, we have

$$\sum_{t=1}^{k} \deg_{G'}(v_t) = \sum_{t=1}^{k} \deg_{G''}(v_t);$$

for each $k \in \{s, s+1, \ldots, r-1\}$, we have

$$\sum_{t=1}^{k} \deg_{G'}(v_t) > \sum_{t=1}^{k} \deg_{G''}(v_t);$$

for each $k \in \{r, r+1, \ldots, n\}$, we have

$$\sum_{t=1}^{k} \deg_{G'}(v_t) = \sum_{t=1}^{k} \deg_{G''}(v_t).$$

Therefore, $D(G') \succ D(G'')$. Using Lemma 2.1, we conclude that $I_d(G') < I_d(G'')$. $\square$

Proof of Corollary 2.1. Set $H = G - uv$. We have $G = H + uv$ and $G' = H + wv$. Since $\deg_G(u) - \deg_G(w) \geq 2$, $\deg_H(u) = \deg_G(u) - 1 > \deg_G(w) = \deg_H(w)$. By Theorem 2.1, we have $I_d(G) < I_d(G')$. $\square$

Proof of Corollary 2.2. Set $H = G - e$. We have $G = H + e$ and $G' = H + f$.

Suppose that $\deg_G(u) \leq \deg_G(w)$. We have $\deg_H(u) = \deg_G(u) - 1 < \deg_G(w) = \deg_H(w)$. By Theorem 2.1, we have $I_d(G) > I_d(G')$, a contradiction.

Hence, $\deg_G(u) > \deg_G(w)$ and $\deg_H(u) = \deg_G(u) - 1 \geq \deg_G(w) = \deg_H(w)$. By Theorem 2.1, we have $I_d(G) \leq I_d(G')$. $\square$

Proof of Theorem 2.2. Since $\deg_G(u) + \deg_G(v) = \deg_G(w) + \deg_G(x) = s$ and $\deg_G(u) - \deg_G(v) \leq \deg_G(w) - \deg_G(x)$, we have $\deg_G(u) \geq \frac{s}{2}$, $\deg_G(w) \geq \frac{s}{2}$ and $\deg_G(w) \geq \deg_G(u)$. If $\deg_G(u) = \deg_G(w)$, then $I_d(G') = I_d(G'')$. We consider the case $\deg_G(w) > \deg_G(u)$ in the following.

Let $g(t) = t\log\frac{t+1}{t} + \log(t+1) + (s-t)\log\frac{s-t+1}{s-t} + \log(s-t+1)$ for $\frac{s}{2} \le t \le s-1$. By calculating the first-order derivative, we obtain

$$g'(\frac{s}{2}) = 0$$

and

$$g'(t) = \log\frac{t+1}{t} - \log\frac{s-t+1}{s-t} < 0$$

for $\frac{s}{2} < t \le s-1$. This implies $g(t)$ strictly decreases as t increases for $\frac{s}{2} \le t \le s-1$. Let m be the size of G. Because $\deg_G(u) + \deg_G(v) = \deg_G(w) + \deg_G(x) = s$, we have

$$\begin{aligned}
I_d(G') = {} & \log(2m+2) - \frac{1}{2m+2}h_d(G) \\
& + \frac{1}{2m+2}(\deg_G(u)\log(\deg_G(u)) + \deg_G(v)\log(\deg_G(v)) \\
& - (\deg_G(u)+1)\log(\deg_G(u)+1) - (\deg_G(v)+1)\log(\deg_G(v)+1)) \\
= {} & \log(2m+2) - \frac{1}{2m+2}h_d(G) - \frac{1}{2m+2}\left(\deg_G(u)\log\left(\frac{\deg_G(u)+1}{\deg_G(u)}\right)\right. \\
& \left. + \log(\deg_G(u)+1) + \deg_G(v)\log\left(\frac{\deg_G(v)+1}{\deg_G(v)}\right) + \log(\deg_G(v)+1)\right) \\
= {} & \log(2m+2) - \frac{1}{2m+2}h_d(G) - \frac{1}{2m+2}\left(\deg_G(u)\log\left(\frac{\deg_G(u)+1}{\deg_G(u)}\right)\right. \\
& + \log(\deg_G(u)+1) + (s-\deg_G(u))\log\left(\frac{s-\deg_G(u)+1}{s-\deg_G(u)}\right) \\
& \left. + \log(s-\deg_G(u)+1)\right) \\
= {} & \log(2m+2) - \frac{1}{2m+2}h_d(G) - \frac{1}{2m+2}g(\deg_G(u)).
\end{aligned}$$

By similar calculations, $I_d(G'') = \log(2m+2) - \frac{1}{2m+2}h_d(G) - \frac{1}{2m+2}g(\deg_G(w))$.

Since $\deg_G(x) \ge 1$, $s-1 \ge \deg_G(w)$. This implies $\frac{s}{2} \le \deg_G(u) < \deg_G(w) \le s-1$. So we have $g(\deg_G(u)) > g(\deg_G(w))$. Thus $I_d(G') < I_d(G'')$. □

Proof of Theorem 2.3. Let $\{v_1, v_2, \ldots, v_n\}$ denote the vertex set of G. We use $v_{i1}, v_{i2}, \ldots, v_{ik}$ to denote k copies of v_i in the blow-up graph $G^{(k)}$. Let m be the size of G. By definition, we have $\deg_{G^{(k)}}(v_{ij}) = k\deg_G(v_i)$ for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,k$. Thus $\sum_{j=1}^{k}\sum_{i=1}^{n}\deg_{G^{(k)}}(v_{ij}) = \sum_{j=1}^{k}\sum_{i=1}^{n}k\deg_G(v_i) = \sum_{j=1}^{k}2km = 2k^2m$. So we have

$$\begin{aligned}
I_d(G^{(k)}) &= \log(2k^2m) - \frac{1}{2k^2m}\sum_{i=1}^{n}\sum_{j=1}^{k}\deg_{G^{(k)}}(v_{i,j})\log\left(\deg_{G^{(k)}}(v_{i,j})\right)\\
&= \log(2k^2m) - \frac{1}{2k^2m}\sum_{i=1}^{n}(k^2\deg_G(v_i))\log(k\deg_G(v_i))\\
&= \log(2k^2m) - \frac{k^2}{2k^2m}\left(\log(k)\sum_{i=1}^{n}\deg_G(v_i) + \sum_{i=1}^{n}\deg_G(v_i)\log\left(\deg_G(v_i)\right)\right)\\
&= \log(2m) - \frac{1}{2m}h_d(G) + \log(k)\\
&= I_d(G) + \log(k).
\end{aligned}$$

□

Proof of Theorem 2.4. Let m be the size of G. Identifying x and y of G, we use a vertex z to replace these vertices. This implies $\deg_{G'}(z) = \deg_G(x) + \deg_G(y)$.

Suppose that $\deg_G(x) = 0$ or $\deg_G(y) = 0$. It follows from $\deg_G(x) = 0$ (resp., $\deg_G(y) = 0$) that $\deg_{G'}(z) = \deg_G(y)$ (resp., $\deg_{G'}(z) = \deg_G(x)$). Because the case with $\deg_G(y) = 0$ can be proved similarly, we only consider the case that $\deg_G(x) = 0$. We have

$$\begin{aligned}
I_d(G') - I_d(G) &= \frac{1}{2m}(\deg_G(x)\log(\deg_G(x)) + \deg_G(y)\log(\deg_G(y))\\
&\quad - \deg_{G'}(z)\log(\deg_{G'}(z)))\\
&= \frac{1}{2m}\left(0\log(0) + \deg_G(y)\log(\deg_G(y)) - \deg_G(y)\log(\deg_G(y))\right)\\
&= 0,
\end{aligned}$$

a contradiction.

This contradiction implies $\deg_G(x) > 0$ and $\deg_G(y) > 0$, and

$$\frac{\deg_G(x)^{\deg_G(x)} + \deg_G(y)^{\deg_G(y)}}{(\deg_G(x) + \deg_G(y))^{(\deg_G(x)+\deg_G(y))}} < 1.$$

So we have

$$\begin{aligned}
I_d(G')-I_d(G) &= \frac{1}{2m}(\deg_G(x)\log\deg_G(x)+\deg_G(y)\log\deg_G(y)\\
&\quad -\deg_{G'}(z)\log\deg_{G'}(z))\\
&= \frac{1}{2m}(\deg_G(x)\log\deg_G(x)+\deg_G(y)\log\deg_G(y)\\
&\quad -(\deg_G(x)+\deg_G(y))\log(\deg_G(x)+\deg_G(y)))\\
&= \frac{1}{2m}\log\left(\frac{\deg_G(x)^{\deg_G(x)}+\deg_G(y)^{\deg_G(y)}}{(\deg_G(x)+\deg_G(y))^{(\deg_G(x)+\deg_G(y))}}\right)\\
&< 0. \qquad \square
\end{aligned}$$

Proof of Theorem 2.5. Let G and H be two graphs satisfying the hypothesis of the theorem. Since z is a non-isolated vertex of H, we have $\deg_H(z)\geq 1$. If $\deg_G(x)>\deg_G(y)$, then

$$\frac{\deg_G(y)^{\deg_G(y)}(\deg_G(x)+\deg_H(z))^{\deg_G(x)}}{\deg_G(x)^{\deg_G(x)}(\deg_G(y)+\deg_H(z))^{\deg_G(y)}}>1$$

and

$$\frac{\deg_G(x)+\deg_H(z)}{\deg_G(y)+\deg_H(z)}>1.$$

This implies

$$\begin{aligned}
h_d(GxHz)-h_d(GyHz) &= (\deg_G(x)+\deg_H(z))\log(\deg_G(x)+\deg_H(z))\\
&\quad +\deg_G(y)\log\deg_G(y)\\
&\quad -(\deg_G(y)+\deg_H(z))\log(\deg_G(y)+\deg_H(z))\\
&\quad -\deg_G(x)\log\deg_G(x)\\
&= \deg_G(x)\log\left(\frac{\deg_G(x)+\deg_H(z)}{\deg_G(x)}\right)\\
&\quad +\deg_G(y)\log\left(\frac{\deg_G(y)}{\deg_G(y)+\deg_H(z)}\right)\\
&\quad +\deg_G(z)\log\left(\frac{\deg_G(x)+\deg_H(z)}{\deg_G(y)+\deg_H(z)}\right)\\
&= \log\left(\frac{\deg_G(y)^{\deg_G(y)}(\deg_G(x)+\deg_H(z))^{\deg_G(x)}}{\deg_G(x)^{\deg_G(x)}(\deg_G(y)+\deg_H(z))^{\deg_G(y)}}\right)
\end{aligned}$$

$$+\deg_G(z)\log\left(\frac{\deg_G(x)+\deg_H(z)}{\deg_G(y)+\deg_H(z)}\right)$$
$$> 0.$$

So we have $I_d(GxHz) < I_d(GyHz)$. If $\deg_G(x) = \deg_G(y)$, then $D(GxHz) = D(GyHz)$. Therefore, $I_d(GxHz) = I_d(GyHz)$. $\square$

Proof of Theorem 2.6. Let m (resp., m') be the size of G (resp., G'). By definition, $\deg_{G\times G'}((u,u')) = \deg_G(u)\deg_{G'}(u')$ for $u \in V(G)$ and $u' \in V(G')$. So we have

$$\begin{aligned}
&\sum_{\substack{u\in V(G)\\ u'\in V(G')}} \deg_{G\times G'}((u,u'))\\
&= \sum_{u\in V(G)}\sum_{u'\in V(G')} \deg_G(u)\deg_{G'}(u')\\
&= 2m' \sum_{u\in V(G)} \deg_G(u)\\
&= 4mm'.
\end{aligned}$$

This implies

$$\begin{aligned}
I_d(G\times G') &= \log(4mm') - \frac{1}{4mm'}\sum_{\substack{u\in V(G)\\ u'\in V(G')}} \deg_{G\times G'}((u,u'))\log(\deg_{G\times G'}((u,u')))\\
&= \log(4mm') - \frac{1}{4mm'}\sum_{\substack{u\in V(G)\\ u'\in V(G')}} \deg_G(u)\deg_{G'}(u')\log(\deg_G(u)\deg_{G'}(u'))\\
&= \log(4mm') - \frac{1}{4mm'}\sum_{\substack{u\in V(G)\\ u'\in V(G')}} (\deg_G(u)\deg_{G'}(u')\log(\deg_G(u))\\
&\quad + \deg_G(u)\deg_{G'}(u')\log(\deg_{G'}(u'))\\
&= \log(4mm') - \frac{2m'\sum_{u\in V(G)}\deg_G(u)\log(\deg_G(u))}{4mm'}\\
&\quad - \frac{2m\sum_{u'\in V(G')}\deg_{G'}(u')\log(\deg_{G'}(u'))}{4mm'}\\
&= \log(2m) - \frac{1}{2m}h_d(G) + \log(2m') - \frac{1}{2m'}h_d(G')
\end{aligned}$$

$$= I_d(G) + I_d(G'). \qquad \square$$

Chapter 3

Degree-entropy of trees and unicyclic graphs

In this chapter, we characterize the graphs that minimize or maximize the degree-entropy among trees and unicyclic graphs with some given parameters relying on some results from Chapter 2. Firstly, we determine the minimum and maximum values of the degree-entropy among trees and unicyclic graphs with given bipartitions, respectively. Secondly, we characterize all the corresponding extremal graphs. Finally, we determine the minimum value of the degree-entropy among trees and unicyclic graphs with a given diameter.

3.1 Introduction

Before we present our results , we start with some background and recall some terminology and notation.

In this chapter, we call a graph of order n an *n-vertex graph*. Let G be an n-vertex graph with $V = \{v_1, v_2, \dots, v_n\}$. We refer to equations (2.1) and (2.2) in Chapter 2 for the definition of the degree-entropy $I_d(G)$ of G, and to the function $h_d(G) = \sum_{i=1}^{n} \deg_G(v_i) \log \deg_G(v_i)$, respectively. From the discussion following equation (2.2), we see that, for certain families of graphs of order n and size m, the extremal values of $I_d(G)$ can be obtained directly from the extremal values of $h_d(G)$. We will frequently use this in our proofs.

One of the fundamental and first natural problems in studying any (new) graph invariant is determining its minimum and maximum values, and characterizing the extremal graphs attaining these values. For such problems, restrictions to special graph classes are also often considered. This holds in particular for graph entropies and more generally for topological indices of graphs. Especially in application areas like chemistry, the class of trees and more sophisticated special graph classes are motivated by the atomic structure of hydrocarbon molecules or their carbon atom skeleton.

The problems of determining the minimum and maximum values of the degree-entropy of trees and unicyclic graphs have been studied in [29, 63]. S. Cao, M. Dehmer and Y. Shi [29] determined the minimum value and maximum value of degree-entropy among trees of a given order, and characterized the corresponding extremal graph. Before showing their results, we introduce some notation. As usual, P_n denotes the n-vertex path. We also recall that by $K_{1,n-1}$ we denote the n-vertex star, and we call v the *center* of $K_{1,n-1}$ if $\deg_{K_{1,n-1}}(v) = n-1$.

Theorem 3.1 ([29]). *Let T be an n-vertex tree. We have*

(a) $I_d(T) \le I_d(P_n)$*, with equality holding if and only if* $T \cong P_n$*;*

(b) $I_d(T) \ge I_d(K_{1,n-1})$*, with equality holding if and only if* $T \cong K_{1,n-1}$.

Related to the work presented here, H. Zhang and S. Li [127] established sharp lower bounds on the so-called cover cost among trees with given diameters and bipartition. Similarly, Z. Du [54] determined the minimum and maximum Wiener indices of trees with a given bipartition, as well as the minimum Wiener index of unicyclic graphs with a given bipartition. Motivated by the above results, in this chapter we study the minimum and maximum values of the degree-entropy of trees and unicyclic graphs with a given bipartition, and the minimum value of the degree-entropy of trees and unicyclic graphs with a given diameter.

3.2 Results

In this section, we present our results on the extremal values of the degree-entropy among trees and unicyclic graphs for which we fix some graph invariants. In addition, we give the corresponding extremal graphs.

Before we can present our first result in this section, we need some additional notation.

For an integer $k \geq 1$, let $T_n(n_0, n_1, \ldots, n_k)$ be the n-vertex tree which is obtained from the path $P_{k+1} = v_0 v_1 \cdots v_k$ by attaching n_i pendant vertices to the vertex v_i for $i = 1, 2, \ldots, k-1$, so with $n-2 = \sum_{i=1}^{k-1}(n_i+1)$. Now we let $\mathscr{T}^*_{n,k}$ (see Figure 3.1) denote the set of all n-vertex trees $T_n(n_0, n_1, \ldots, n_k)$ with $n_1 = \cdots = n_{i-1} = n_{i+1} = \cdots = n_{k-1} = 0$ and $n_i = n-1-k$ for $i = 1, 2, \ldots, k-1$.

All the trees of Figure 3.1 have the same degree-entropy $\log(2n-2) - \frac{(n-k+1)\log(n-k+1)}{2n-2} - \frac{k-1}{n-1}$, and they appear naturally in the following extremal result.

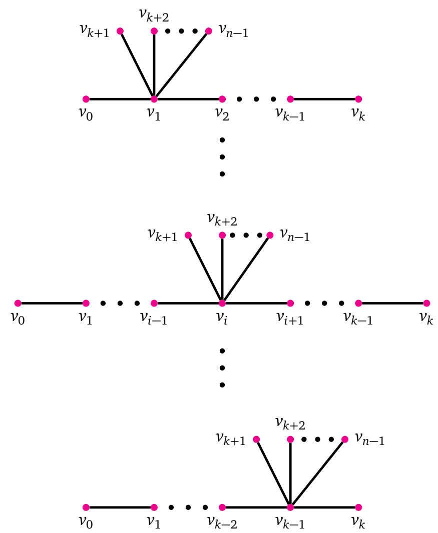


Figure 3.1: The trees in $\mathscr{T}^*_{n,k}$.

Theorem 3.2. *Let T be an n-vertex tree with diameter $k \geq 1$. If $I_d(T)$ attains the minimum value among all n-vertex trees with diameter k, then $T \in \mathscr{T}^*_{n,k}$.*

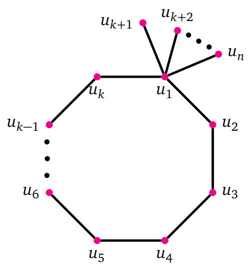


Figure 3.2: The unicyclic graph $C_n(n_1,n_2,\dots,n_k)$ for $n_1 = n-k > 0$ and $n_2 = \cdots = n_k = 0$.

We can prove an analogous result for unicyclic graphs (possibly a tree with one double edge). Let C_n denote the n-vertex cycle with $n \geq 2$ (where C_2 corresponds to a double edge). Let $C_n(n_1,n_2,\dots,n_k)$ be the n-vertex unicyclic graph obtained from the cycle $C_k = u_1u_2\cdots u_ku_1$ by attaching n_i pendant neighbors to the vertex u_i for $i = 1,2,\dots,k$. Figure 3.2 shows the unicyclic graph $C_n(n_1,n_2,\dots,n_k)$ for $n_1 = n-k > 0$ and $n_2 = \cdots = n_k = 0$ (i.e., $C_n(n-k,\underbrace{0,\dots,0}_{k-1})$).

Theorem 3.3. *Let C be an n-vertex unicyclic graph containing a cycle of order $k \geq 2$. If $I_d(C)$ attains the minimum value among all n-vertex unicyclic graphs containing a k-vertex cycle, then $C \cong C_n(n-k,\underbrace{0,\dots,0}_{k-1})$.*

We continue with the extremal results for specific subclasses of trees and unicyclic graphs, but first need some additional terminology and notation.

We say that a graph G admits a (p,q)-*bipartition* if $V(G) = V_1 \cup V_2$ for disjoint sets V_1 and V_2 with $|V_1| = p > 0$ and $|V_2| = q > 0$, and each edge of G has end-vertices in V_1 and V_2.

Let $\mathscr{T}(p,q)$ denote the set of all trees admitting a (p,q)-bipartition. Let $S^*(p,q)$ be the member of $\mathscr{T}(p,q)$ obtained by attaching $p-1$ and $q-1$ pendant vertices to the two vertices of a P_2, respectively (as indicated in Figure 3.3).

Let G be a graph, and recall that $\deg_{\max}(G)$ and $\deg_{\min}(G)$ denote the maximum degree and minimum degree among the vertices in G, respectively.

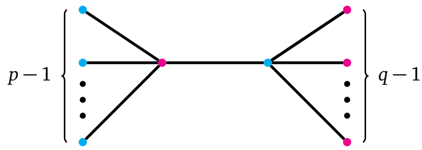


Figure 3.3: The tree $S^*(p,q)$.

We consider another specific subclass of $\mathscr{T}(p,q)$. Let $\mathscr{T}^*(p,q)$ denote the set of all trees that can be obtained from any q-vertex tree T with $\deg_{\max}(T) \leq \left\lceil \frac{p+q-1}{q} \right\rceil$ in the following way. First subdivide every edge of T, i.e., replace each edge $e = uv$ by a path $ux_{uv}v$ for a newly added vertex x_{uv}. The new tree clearly admits a $(q-1,q)$-partition with $V_2 = V(T)$ and V_1 consisting of the newly added vertices. Next attach $p-q+1$ pendant vertices to the vertices of V_2 in such a way that the maximum degree of the vertices in V_2 exceeds their minimum degree by at most 1. The construction of one member of $\mathscr{T}^*(15,6)$ is illustrated in Figure 3.4. Clearly, by construction every tree in $\mathscr{T}^*(p,q)$ has a (p,q)-bipartition.

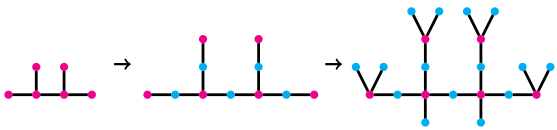


Figure 3.4: The construction of a tree in $\mathscr{T}^*(15,6)$.

The next result determines the minimum value of $I_d(T)$ among all trees $T \in \mathscr{T}(p,q)$ and characterizes the unique extremal tree.

Theorem 3.4. *Let p and q be integers with $p \geq q \geq 1$. Then $I_d(T)$ attains the minimum value among all trees in $\mathscr{T}(p,q)$ if and only if $T \cong S^*(p,q)$.*

The following result determines the maximum value of $I_d(T)$ among all trees $T \in \mathscr{T}(p,q)$ and characterizes all the extremal trees.

Theorem 3.5. *Let p and q be integers with $p \geq q \geq 1$. Then $I_d(T)$ attains the maximum value among all trees in $\mathscr{T}(p,q)$ if and only if $T \in \mathscr{T}^*(p,q)$.*

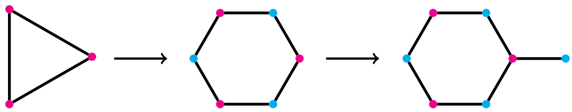


Figure 3.5: The construction of one member of $\mathscr{C}^*(4,3)$.

We finish this section with the counterparts of the above tree results for unicyclic graphs. For this, we let $\mathscr{C}(p,q)$ denote the set of all unicyclic graphs (possibly a tree with one double edge) admitting a (p,q)-bipartition. Obviously, every member of $\mathscr{C}(p,q)$ has a unique cycle of an even order. We define a subclass $\mathscr{C}^*(p,q)$ of $\mathscr{C}(p,q)$ in a similar way as we did for trees. Let $\mathscr{C}^*(p,q)$ consist of all unicyclic graphs that can be obtained from a q-vertex unicyclic graph C with $\deg_{\max}(C) \leq \left\lceil \frac{p+q}{q} \right\rceil$ in the following way. First subdivide every edge of C to obtain a unicyclic graph which admits a (q,q)-bipartition with $V_2 = V(C)$ and V_1 consisting of the newly added vertices. Next attach $p-q$ pendant vertices to the vertices of V_2 in such a way that the maximum degree of the vertices in V_2 exceeds their minimum degree by at most 1. The construction of one member of $\mathscr{C}^*(4,3)$ is illustrated in Figure 3.5. Clearly, by construction every unicyclic graph in $\mathscr{C}^*(p,q)$ has a (p,q)-bipartition.

The next result determines the minimum value of $I_d(C)$ among all unicyclic graphs $C \in \mathscr{C}(p,q)$ and identifies the unique extremal graph.

Theorem 3.6. *Let p and q be integers with $p \geq q \geq 1$. Then $I_d(C)$ attains the minimum value among all unicyclic graphs in $\mathscr{C}(p,q)$ if and only if $C \cong C_n(p-1,q-1)$.*

Note that $C_n(p-1,q-1)$ in the above statement is a special case of the previously defined class $C_n(n_1,n_2,\ldots,n_k)$, and that $C_n(p-1,q-1)$ can be obtained from the tree in Figure 3.3 by replacing the middle edge by a double edge.

Our final result determines the maximum value of $I_d(C)$ among all unicyclic graphs $C \in \mathscr{C}(p,q)$ and characterizes all the extremal graphs.

Theorem 3.7. *Let p and q be integers with $p \geq q \geq 1$. Then $I_d(C)$ attains the maximum value among all unicyclic graphs in $\mathscr{C}(p,q)$ if and only if $C \in \mathscr{C}^*(p,q)$.*

3.3 Preliminaries

In this section, we give some additional terminology and notation, and we state a number of lemmas which will be used in the proofs of our results.

Let G be a graph, and let S be a nonempty subset of $V(G)$. If the number of vertices in S with degree d_i is a_i for $i = 0, 1, \ldots, k$, then we denote by $D(S) = (\underbrace{d_k, \ldots, d_k}_{a_k}, \underbrace{d_{k-1}, \ldots, d_{k-1}}_{a_{k-1}}, \ldots, \underbrace{d_1, \ldots, d_1}_{a_1}) = [d_k^{a_k}, d_{k-1}^{a_{k-1}}, \ldots, d_1^{a_1}]$ the degree sequence of S in which $0 = d_0 < d_1 < \cdots < d_k$ and $a_0 + a_1 + \cdots + a_k = |S|$. We use $D(G)$ to represent the degree sequence $D(V(G))$. If there exists a graph G with degree sequence $D = D(G)$, then D is called *graphic*, and G is called a *realization* of D. Let A and B be two non-increasing sequences. Adopting the notation of Subsection 2.1.2, we use $A \succeq B$ and $A \succ B$ to denote that A majorizes B and that A strictly majorizes B, respectively.

Let G be a fixed graph with at least one edge, and let T be a randomly chosen n-vertex tree. Denote by $GuTw$ the graph obtained from G and T by identifying a fixed non-isolated vertex $u \in V(G)$ and a randomly chosen vertex $w \in V(T)$. Let v be the center of the star $K_{1,n-1}$. For our proofs of Theorems 3.2 and 3.3, the following two lemmas are key ingredients.

The first result shows that among all n-vertex trees, $I_d(GuK_{1,n-1}v)$ attains the minimum value.

Lemma 3.1. *Let T be an n-vertex tree with $w \in V(T)$, and let v be the center of $K_{1,n-1}$. Suppose that G is a fixed graph with a fixed non-isolated vertex $u \in V(G)$. Then $I_d(GuTw) \geq I_d(GuK_{1,n-1}v)$, with equality holding in the inequality if and only if $GuTw \cong GuK_{1,n-1}v$.*

Proof. Suppose that $I_d(GuTw) \leq I_d(GuK_{1,n-1}v)$ and $GuTw$ is not isomorphic to $GuK_{1,n-1}v$. So we have $\deg_T(w) < n-1$. This implies that $\deg_{GuTw}(u) = \deg_G(u) + \deg_T(w) < \deg_G(u) + n - 1 = \deg_{GuK_{1,n-1}v}(u)$. By Theorem 3.1 (b), and recalling (2.2) and the remarks we made there, we have $h_d(T) \leq h_d(K_{1,n-1})$. Since $\deg_G(u) \geq 1$, the function $g(t) = (t + \deg_G(u))\log(t + \deg_G(u)) - t\log t$ strictly increases as t increases for $t > 0$. This implies

$$\begin{aligned} h_d(GuK_{1,n-1}v) - h_d(GuTw) =& h_d(K_{1,n-1}) - (n-1)\log(n-1) \\ &+ \deg_{GuK_{1,n-1}v}(u)\log\deg_{GuK_{1,n-1}v}(u) - h_d(T) \end{aligned}$$

$$\begin{aligned}
&+\deg_T(w)\log\deg_T(w)-\deg_{GuTw}(u)\log\deg_{GuTw}(u)\\
=&h_d(K_{1,n-1})-h_d(T)+(\deg_G(u)+n-1)\log(\deg_G(u)+n-1)\\
&-(\deg_G(u)+\deg_T(w))\log(\deg_G(u)+\deg_T(w))\\
&+\deg_T(w)\log\deg_T(w)-(n-1)\log(n-1)\\
=&h_d(K_{1,n-1})-h_d(T)+g(n-1)-g(\deg_T(w))\\
>&0,
\end{aligned}$$

a contradiction. □

For our next lemma, let $\mathscr{T}_{n,k}$ (resp., $\mathscr{C}_{n,k}$) denote the set of all trees $T_n(n_0,n_1,\dots,n_k)$ (resp., all unicyclic graphs $C_n(n_1,n_2,\dots,n_k)$). We consider the extremal results for $\mathscr{T}_{n,k}$ and $\mathscr{C}_{n,k}$.

Lemma 3.2. *Let $\mathscr{T}_{n,k}$ and $\mathscr{C}_{n,k}$ be defined as above. Then*

(a) *$I_d(T)$ attains the minimum value among all trees in $\mathscr{T}_{n,k}$ if and only if $T\in\mathscr{T}^*_{n,k}$ for $n>k\geq 1$;*

(b) *$I_d(C)$ attains the minimum value among all unicyclic graphs in $\mathscr{C}_{n,k}$ if and only if $C\cong C_n(n-k,\underbrace{0,\dots,0}_{k-1})$ for $n\geq k\geq 2$.*

Proof. We only prove (a) because (b) can be proved similarly.

Suppose that $T\notin\mathscr{T}^*_{n,k}$ and $I_d(T)$ attains the minimum value among all trees in $\mathscr{T}_{n,k}$. Let $P_{k+1}=v_0v_1\cdots v_k$ be the diametrical path of T. This implies that there exist two distinct vertices v_i and v_j with $\deg_T(v_i)\geq 3$ and $\deg_T(v_j)\geq 3$. Without loss of generality, we may assume that $\deg_T(v_i)\leq\deg_T(v_j)$. Let $v\notin V(P_{k+1})$ be a neighbor of v_i. Set $T'=T-v_iv+v_jv$. By Corollary 2.1, we have $I_d(T')<I_d(T)$, a contradiction. □

We need the next two lemmas for our proof of Theorem 3.6. Let n, p, q and k be four integers with $p+q=n$, $p\geq q\geq 1$ and $k\geq 2$. The following lemma shows that, among all possible values for $n_1,n_2,\dots,n_k$ with $C_n(n_1,n_2,\dots,n_k)\in\mathscr{C}(p,q)$, the minimum value is attained by $I_d(C_n(p-\frac{k}{2},q-\frac{k}{2},\underbrace{0,\dots,0}_{k-2}))$.

Lemma 3.3. *Let n, p, q and k be four integers with $p+q=n$, $p\geq q\geq 1$ and $k\geq 2$. If $C_n(n_1,n_2,\ldots,n_k)\in\mathscr{C}(p,q)$, then $I_d(C_n(n_1,n_2,\ldots,n_k))\geq I_d(C_n(p-\frac{k}{2},q-\frac{k}{2},\underbrace{0,\ldots,0}_{k-2}))$.*

Proof. Let $C_k=u_1u_2\cdots u_ku_1$ be the cycle of $C_n(n_1,n_2,\ldots,n_k)$ and let (V_1,V_2) correspond to a (p,q)-bipartition of $C_n(n_1,n_2,\ldots,n_k)$. Suppose that $u_{2j}\in V_1$ and $u_{2j-1}\in V_2$ for $j=1,2,\ldots,\frac{k}{2}$. This implies that

$$\sum_{j=1}^{\frac{k}{2}}\big(\deg_{C_n(n_1,n_2,\ldots,n_k)}(u_{2j-1})-2\big)=p-\frac{k}{2}$$

and

$$\sum_{j=1}^{\frac{k}{2}}\big(\deg_{C_n(n_1,n_2,\ldots,n_k)}(u_{2j})-2\big)=q-\frac{k}{2}.$$

It follows that

$$2\leq\deg_{C_n(n_1,n_2,\ldots,n_k)}(u_{2j})\leq p-\frac{k}{2}+2$$

and

$$2\leq\deg_{C_n(n_1,n_2,\ldots,n_k)}(u_{2j-1})\leq q-\frac{k}{2}+2$$

for $j=1,2,\ldots,\frac{k}{2}$, and

$$\deg_{C_n(n_1,n_2,\ldots,n_k)}(u)=1$$

for $u\in V(C_n(n_1,n_2,\ldots,n_k))\setminus V(C_k)$. We have $D(C_n(p-\frac{k}{2},q-\frac{k}{2},\underbrace{0,\ldots,0}_{k-2}))=[p-\frac{k}{2}+2,q-\frac{k}{2}+2,2^{k-2},1^{n-k}]$. This implies $D(C_n(p-\frac{k}{2},q-\frac{k}{2},\underbrace{0,\ldots,0}_{k-2}))\succeq D(C_n(n_1,n_2,\ldots,n_k))$. By Lemma 2.1, we have

$$I_d(C_n(p-\frac{k}{2},q-\frac{k}{2},\underbrace{0,\ldots,0}_{k-2}))\leq I_d(C_n(n_1,n_2,\ldots,n_k)).\qquad\square$$

We state and prove one more lemma to complete this section. This lemma shows the effect of shortening the cycle of $C_n(s,t,\underbrace{0,\ldots,0}_{k-2})$ on the value of the degree-entropy.

Lemma 3.4. *Let n, s, t and k be four integers with $s+t+k=n$, $n \geq k \geq 4$ and $s, t \geq 0$. Then*

$$I_d(C_n(s+1,t+1,\underbrace{0,\ldots,0}_{k-4})) < I_d(C_n(s,t,\underbrace{0,\ldots,0}_{k-2})).$$

Proof. Without loss of generality, we assume that $s \geq t$. It follows from

$$D(C_n(s,t,\underbrace{0,\ldots,0}_{k-2})) = [s+2,t+2,2^{k-2},1^{n-k}]$$

and

$$D(C_n(s+1,t+1,\underbrace{0,\ldots,0}_{k-4})) = [s+3,t+3,2^{k-4},1^{n-k+2}]$$

that $D(C_n(s+1,t+1,\underbrace{0,\ldots,0}_{k-4})) \succ D(C_n(s,t,\underbrace{0,\ldots,0}_{k-2}))$. By Lemma 2.1, we have $I_d(C_n(s+1,t+1,\underbrace{0,\ldots,0}_{k-4})) < I_d(C_n(s,t,\underbrace{0,\ldots,0}_{k-2}))$. □

3.4 Proofs

We present all proofs of our main results in this section.

Proof of Theorem 3.2. Let $P_{k+1} = v_0 v_1 \cdots v_k$ be a diametrical path of T, and let T^i be the component of $T - E(P_{k+1})$ containing v_i for $i = 0, 1, \ldots, k$. Let H be the component of $T - E(T^i)$ containing v_i. By Lemma 3.1, we have $I_d(Hv_iT^iv_i) \geq I_d(Hv_iK_{1,|V(T^i)|-1}v)$ in which v is the center of $K_{1,|V(T^i)|-1}$. This implies that $T \in \mathscr{T}_{n,k}$. And by Lemma 3.2 (a), we have $T \in \mathscr{T}^*_{n,k}$. □

Proof of Theorem 3.3. Let $C_k = u_1 u_2 \cdots u_k u_1$ be the cycle of C, and let T^i be the component of $C - E(C_k)$ containing u_i for $i = 1, 2, \ldots, k$. Let H be the component of $C - E(T^i)$ containing u_i. By Lemma 3.1, we have $I_d(Hu_iT^iu_i) \geq I_d(Hu_iK_{1,|V(T^i)|-1}v)$ in which v is the center of $K_{1,|V(T^i)|-1}$. This implies that $C \in \mathscr{C}_{n,k}$. And by Lemma 3.2 (b), we have $C \cong C_n(n-k,\underbrace{0,\ldots,0}_{k-1})$. □

Proof of Theorem 3.4. Let $P_{t+1} = v_0 v_1 \cdots v_t$ be a diametrical path of T. Suppose that the diameter of T is at least 4 (i.e, $t \geq 4$). Let A be the set of neighbors of v_3 excluding v_2 of T. Let H_1 (resp., H_2) be the component of $T - A$ (resp., $T - v_2 v_3$) containing v_3. Then T can be obtained from H_1 and H_2 by identifying $v_3 \in V(H_1)$ and $v_3 \in V(H_2)$. Let T' be the tree obtained from H_1 and H_2 by identifying $v_1 \in V(H_1)$ and $v_3 \in V(H_2)$. We may partition the vertex set of T' and T in the same way. Thus T' has a (p,q)-bipartition. Clearly, $\deg_{H_1}(v_1) > \deg_{H_1}(v_3)$. By Theorem 2.5, we have $I_d(T') < I_d(T)$, a contradiction. Therefore, the diameter of T is at most 3, that is, $T \cong S^*(p,q)$. □

Proof of Theorem 3.5. Let V_1 and V_2 be two subsets of $V(T)$ satisfying $|V_1| = p$ and $|V_2| = q$, and such that each edge of T joins a vertex in V_1 and a vertex in V_2. Let $r = (p+q-1) - q\lfloor \frac{p+q-1}{q} \rfloor$. We state a claim.

Claim 1. *$D(V_1) = [2^{q-1}, 1^{p-q+1}]$ and $D(V_2) = [(\lceil \frac{p+q-1}{q} \rceil)^r, (\lfloor \frac{p+q-1}{q} \rfloor)^{q-r}]$.*

Proof. We only prove $D(V_1) = [2^{q-1}, 1^{p-q+1}]$, since $D(V_2) = [(\lceil \frac{p+q-1}{q} \rceil)^r, (\lfloor \frac{p+q-1}{q} \rfloor)^{q-r}]$ can be proved similarly. Suppose that $D(V_1) = [2^{q-1}, 1^{p-q+1}]$ does not hold. This implies there exist two vertices $u \in V_1$ and $v \in V_1$ satisfying $\deg_T(u) - \deg_T(v) \geq 2$. Let P be the path from u to v, and let $w \notin V(P)$ be a neighbor of u. Let A be the set of neighbors of u excluding w. Let H_1 (resp., H_2) be the component of $T - A$ (resp., $T - uw$) containing u. Then T can be obtained from H_1 and H_2 by identifying $u \in V(H_1)$ and $u \in V(H_2)$. Let T' be the tree obtained from H_1 and H_2 by identifying $u \in V(H_1)$ and $v \in V(H_2)$. We may partition the vertex set of T and T' in the same way. This implies that T' has a (p,q)-bipartition. Clearly, $\deg_{H_2}(u) > \deg_{H_2}(v)$. By Theorem 2.5, we have $I_d(T) < I_d(T')$, a contradiction. Since all graphs in $\mathscr{T}^*(p,q)$ are realizations of $D(V_1) = [2^{q-1}, 1^{p-q+1}]$ and $D(V_2) = [(\lceil \frac{p+q-1}{q} \rceil)^r, (\lfloor \frac{p+q-1}{q} \rfloor)^{q-r}]$, this pair of degree sequences is graphic. ♢

Using Claim 1, to prove $T \in \mathscr{T}^*(p,q)$, it suffices to show that realizations of $D(V_1) = [2^{q-1}, 1^{p-q+1}]$ and $D(V_2) = [(\lceil \frac{p+q-1}{q} \rceil)^r, (\lfloor \frac{p+q-1}{q} \rfloor)^{q-r}]$ are in $\mathscr{T}^*(p,q)$. Let T'' be a tree obtained from T by deleting pendant vertices in V_1, and identifying vertices with degree 2 in V_1 with one of their neighbors (avoiding loops). It follows that T'' is a q-vertex tree with maximum degree at most

$\lceil \frac{p+q-1}{q} \rceil$, and T can be obtained by subdividing every edge of T'' and attaching the pendant vertices to the original vertices. Thus $T \in \mathscr{T}^*(p,q)$. $\square$

Proof of Theorem 3.6. Let V_1 and V_2 be two subsets of $V(C)$ satisfying $|V_1| = p$ and $|V_2| = q$, and such that each edge of C joins a vertex in V_1 and a vertex in V_2. Let $C_k = u_1u_2\cdots u_ku_1$ be the cycle of C. If $k = n = 2$, then $C \cong C_2$. Suppose that $C \cong C_n$ (i.e., $k = n$) for $n \geq 4$. If $C_n \in \mathscr{C}(p,q)$ and $n \geq 4$, then $C_n(1,1,\underbrace{0,\dots,0}_{n-4}) \in \mathscr{C}(p,q)$. Then, by Lemma 3.4, $I_d(C_n) > I_d(C_n(1,1,\underbrace{0,\dots,0}_{n-4}))$ for $n \geq 4$, which leads to a contradiction. So we have $k < n$ for $n \geq 4$. This implies there is a vertex u_i of degree at least 3.

Let A be the set of neighbors of u_i of C excluding u_{i-1} and u_{i+1}, where the addition is taken modulo k. Let H_1 (resp., T^1) be the component of $C-A$ (resp., $C-\{e_1,e_2\}$) containing u_i in which $\psi_C(e_1) = u_{i-1}u_i$ and $\psi_C(e_2) = u_iu_{i+1}$. Then C can be obtained from H_1 and T^1 by identifying $u_i \in V(H_1)$ and $u_i \in V(T^1)$. We next prove that T^1 is a star.

Suppose that T^1 is not a star. This implies that T^1 has a (p',q')-bipartition with $p',q' \geq 2$. Let T' be the tree obtained by attaching $p'-1$ and $q'-1$ pendant vertices to the two vertices u and w of a P_2, respectively. By Theorem 3.4, we have $I_d(T') \leq I_d(T^1)$ (i.e., $h_d(T') \geq h_d(T^1)$). Let C' be the unicyclic graph obtained from H_1 and T' by identifying either $u_i \in V(H_1)$ and $u \in V(T')$, or $u_i \in V(H_1)$ and $w \in V(T')$, such that C' has a (p,q)-bipartition. Without loss of generality, we assume that C' is obtained from H_1 and T' by identifying $u_i \in V(H_1)$ and $u \in V(T')$. It is easy to check that $\deg_{T^1}(u_i) \leq \deg_{T'}(u)$. Thus $(\deg_{T^1}(u_i)+2)\log(\deg_{T^1}(u_i)+2) - \deg_{T^1}(u_i)\log\deg_{T^1}(u_i) - (\deg_{T'}(u)+2)\log(\deg_{T'}(u)+2) + \deg_{T'}(u)\log\deg_{T'}(u) \leq 0$. So we have

$$\begin{aligned}
h_d(C') - h_d(C) &= h_d(H_1) + h_d(T') - 2\log 2 - \deg_{T'}(u)\log\deg_{T'}(u) \\
&\quad + (\deg_{T'}(u)+2)\log(\deg_{T'}(u)+2) \\
&\quad - h_d(H_1) - h_d(T^1) + 2\log 2 + \deg_{T^1}(u_i)\log\deg_{T^1}(u_i) \\
&\quad - (\deg_{T^1}(u_i)+2)\log(\deg_{T^1}(u_i)+2) \\
&= h_d(T') - h_d(T^1) + (\deg_{T^1}(u_i)\log\deg_{T^1}(u_i) \\
&\quad - (\deg_{T^1}(u_i)+2)\log(\deg_{T^1}(u_i)+2)) \\
&\quad - (\deg_{T'}(u)\log\deg_{T'}(u) - (\deg_{T'}(u)+2)\log(\deg_{T'}(u)+2)) \\
&\geq 0.
\end{aligned}$$

This implies $I_d(C') \le I_d(C)$. Let B be the set of neighbors of w in T' excluding u of T'. Since $p', q' \ge 2$, we have $B \ne \emptyset$. Let H_2 (resp., T^2) be the component of $C' - B$ (resp., $C' - f$) containing w in which $\psi_{C'}(f) = uw$. Then C' can be obtained from H_2 and T^2 by identifying $w \in V(H_2)$ and $w \in V(T^2)$. Let C'' be the unicyclic graph obtained from H_2 and T^2 by identifying $u_{i+1} \in V(H_2)$ and $w \in V(T^2)$. Clearly, we may partition the vertices of C' and C'' in the same way, that is, C'' also has a (p,q)-bipartition. It is easy to check that $\deg_{H_2}(u_{i+1}) > \deg_{H_2}(w)$, where the addition is taken modulo k. By Theorem 2.5, we have $I_d(C'') < I_d(C') \le I_d(C)$, a contradiction. There exist some integers $n_1, n_2, \ldots, n_k$ and k such that $C \cong C_n(n_1, n_2, \ldots, n_k)$. By Lemmas 3.3 and 3.4, we have $C \cong C_n(p-1, q-1)$. □

Proof of Theorem 3.7. Let V_1 and V_2 be two subsets of $V(C)$ satisfying $|V_1| = p$ and $|V_2| = q$, and such that each edge of C joins a vertex in V_1 and a vertex in V_2. The statement is trivial for $q = 1$. So we only consider $q \ge 2$ in the following. Let $r = (p+q) - q\lfloor \frac{p+q}{q} \rfloor$. We state a claim.

Claim 2. *$D(V_1) = [2^q, 1^{p-q}]$ and $D(V_2) = [(\lceil \frac{p+q}{q} \rceil)^r, (\lfloor \frac{p+q}{q} \rfloor)^{q-r}]$.*

Proof. We only prove $D(V_1) = [2^q, 1^{p-q}]$, since $D(V_2) = [(\lceil \frac{p+q}{q} \rceil)^r, (\lfloor \frac{p+q}{q} \rfloor)^{q-r}]$ can be proved similarly. Suppose that $D(V_1) = [2^q, 1^{p-q}]$ does not hold. This implies there exist two vertices $u \in V_1$ and $v \in V_1$ satisfying $\deg_C(u) - \deg_C(v) \ge 2$. Let $C_k = u_1 u_2 \cdots u_k u_1$ be the cycle of C. Let T^i be the component of $C - E(C_k)$ containing u_i for $i = 1, 2, \ldots, k$. We distinguish three cases.

Case 1. *$u \in V(C_k)$ and $v \in V(C_k)$.*

We have $\deg_C(u) \ge 4$. Let $w \notin V(C_k)$ be a neighbor of u. Let A be the set of neighbors of u excluding w. Let H_1 (resp., H_2) be the component of $C - A$ (resp., $C - e$) containing u in which $\psi_C(e) = uw$. Then C can be obtained from H_1 and H_2 by identifying $u \in V(H_1)$ and $u \in V(H_2)$. Let C' be the unicyclic graph obtained from H_1 and H_2 by identifying $u \in V(H_1)$ and $v \in V(H_2)$. We may partition the vertex set of C' and C in the same way. This implies that C' has a (p,q)-bipartition. Clearly, $\deg_{H_2}(u) > \deg_{H_2}(v)$. By Theorem 2.5, we have $I_d(C) < I_d(C')$, a contradiction.

Case 2. *$u \in V(C_k)$ and $v \notin V(C_k)$.*

Let $w \in V(C_k)$ be the neighbor of u different from v. Let $C' = C - e + f$ in which $\psi_C(e) = uw$ and $\psi_C(f) = wv$. We may partition the vertex set of C' and C in the same way. Thus C' has a (p,q)-bipartition. Since $\deg_C(u) - \deg_C(v) \geq 2$,

$$\begin{aligned}
h_d(C) - h_d(C') = \; & \deg_C(u)\log\deg_C(u) + \deg_C(v)\log\deg_C(v) \\
& - (\deg_C(u) - 1)\log(\deg_C(u) - 1) \\
& - (\deg_C(v) + 1)\log(\deg_C(v) + 1) \\
= \; & (\deg_C(u)\log\deg_C(u) - (\deg_C(u) - 1)\log(\deg_C(u) - 1)) \\
& - ((\deg_C(v) + 1)\log(\deg_C(v) + 1) - \deg_C(v)\log\deg_C(v)) \\
= \; & (\log\xi_1 + \frac{1}{\ln 2}) - (\log\xi_2 + \frac{1}{\ln 2}) \\
> \; & 0,
\end{aligned}$$

where $\xi_1 \in (\deg_C(u) - 1, \deg_C(u))$ and $\xi_2 \in (\deg_C(v), \deg_C(v) + 1)$. Thus $I_d(C) < I_d(C')$, a contradiction.

Case 3. $u \notin V(C_k)$.

Suppose that $u \in V(T^i)$. If $v \in V(T^i)$, we denote the path from u to v by P; otherwise, we denote the path from u to u_i by P. Let $w \notin V(P)$ be a neighbor of u. Let A be the set of neighbors of u excluding w. Let H_1 (resp., H_2) be the component of $C - A$ (resp., $C - e$) containing u in which $\psi_C(e) = uw$. Then C can be obtained from H_1 and H_2 by identifying $u \in V(H_1)$ and $u \in V(H_2)$. Let C' be the unicyclic graph obtained from H_1 and H_2 by identifying $u \in V(H_1)$ and $v \in V(H_2)$. We may partition the vertex set of C and C' in the same way. Thus C' has a (p,q)-bipartition. Clearly, $\deg_{H_2}(u) > \deg_{H_2}(v)$. By Theorem 2.5, we have $I_d(C) < I_d(C')$, a contradiction.

Since all graphs in $\mathscr{C}^*(p,q)$ are realizations of $D(V_1) = [2^q, 1^{p-q}]$ and $D(V_2) = [(\lceil \frac{p+q}{q} \rceil)^r, (\lfloor \frac{p+q}{q} \rfloor)^{q-r}]$, this pair of degree sequences is graphic. $\diamondsuit$

Using Claim 2, to prove $C \in \mathscr{C}^*(p,q)$, it suffices to show that realizations of $D(V_1) = [2^q, 1^{p-q}]$ and $D(V_2) = [(\lceil \frac{p+q}{q} \rceil)^r, (\lfloor \frac{p+q}{q} \rfloor)^{q-r}]$ are in $\mathscr{C}^*(p,q)$. Let C'' be a unicyclic graph obtained by deleting all pendant vertices in V_1, and identifying the vertices of degree 2 in V_1 with one of their neighbors (avoiding loops). It is easy to check that C'' is a q-vertex unicyclic graph with maximum degree at most $\lceil \frac{p+q}{q} \rceil$, and C can be obtained by subdividing every edge of C'' and attaching the pendant vertices to the original vertices. Thus $C \in \mathscr{C}^*(p,q)$. $\square$

Chapter 4

Extremalities of degree-entropy of bipartite graphs

In this chapter, we characterize the bipartite graphs that minimize the degree-entropy, among all bipartite graphs of a given size, or a given size and (upper bound on the) order. The extremal graphs turn out to be complete bipartite graphs, or nearly complete bipartite. Here we make use of an equivalent representation of bipartite graphs by means of Young diagrams, which make it easier to compare the degree-entropy of related graphs. We conclude that the general characterization of the extremal graphs is a difficult problem, due to its connections with an unsolved problem in number theory, but it is easy for specific values of the order n and size m. We also give a direct argument to characterize the graphs maximizing the degree-ntropy. We indicate how some of our ideas extend to other graphical function-indices as well.

4.1 Introduction

Recall that the degree-entropy $I_d(G)$ of a graph G is the Shannon entropy of its degree sequence normalized by the degree sum. In this chapter, we continue our work on determining the extremal graphs (and thus extremal values) for the degree-entropy among all bipartite graphs satisfying some natural restrictions. For convenience, we will study the extremal problems of $I_d(G)$ by making use of the function $h_d(G) = \sum_{i=1}^{n} \deg_G(v_i) \log(\deg_G(v_i))$ for some graph classes of a give size. For the justification of this, we refer to the definitions in equations (2.1) and (2.2), and the subsequent discussion in Chapter 2.

4.1.1 Motivation and contributions

Here we determine the graphs with the extremal degree-entropy among all bipartite graphs of a given size m, or a given size m and (upper bound on the) order n. The maximum value for the degree-entropy is obtained by the graphs for which the degree sequence is as balanced as possible. This is a corollary of Karamata's inequality. Note that the random graphs $G_{n,m}$ and $G_{n,p}$ are close to balanced, i.e., the maximum degree and the minimum degree are nearly the same for $m = \Theta(n^2)$, and thus attain a degree-entropy which is almost maximal. This is in line with the intuition that entropy is a measure for randomness.

In graph theory, bipartite graphs are one of the main special classes to investigate, because they find several applications in pure and applied mathematics. Hall's Matching theorem is a famous theorem on bipartite graphs with several applications in scheduling and matching problems, e.g., the kidney matching process. When edges represent bonds between positive and negative charges, the resulting graph is bipartite. Furthermore, results on bipartite graphs often lead to results on more general classes of graphs. A famous example is given by the Kahn-Zhao theorem, where V. Kahn [74] considered a problem on the maximum number of independent sets for bipartite graphs and Y. Zhao [129] reduced the general case to the bipartite case.

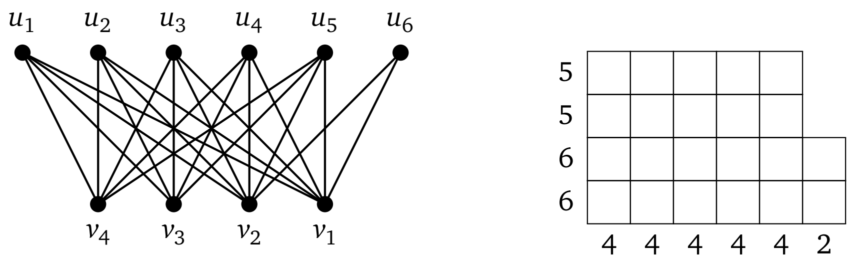
We will show that, given the size m, the bipartite graphs attaining the minimum degree-entropy are precisely the complete bipartite graphs of size m. So if m has $\sigma(m)$ divisors, there are $\left\lceil \frac{\sigma(m)}{2} \right\rceil$ non-isomorphic extremal graphs, all of which are of the form $K_{q,y}$ with $yq = m$.

Theorem 4.1. *If $G = (U \cup V, E)$ is a bipartite graph of size m, then $I_d(G) \geq 1 + \log(\sqrt{m})$, with equality holding if and only if G is a complete bipartite graph.*

Let n, m and y be three integers with $y \nmid m$ and $n > y + \lfloor \frac{m}{y} \rfloor$. When there is an upper bound on the order (one can extend with isolated vertices since $0\log(0) = 0$), such that m cannot be written as a product yq with $n > y + \lfloor \frac{m}{y} \rfloor$, the problem is harder. Let $K_{q,y} = (U \cup V, E)$ be a complete bipartite graph with $|U| = q$, $|V| = y$ and $|E| = qy$ in which $q = \lfloor \frac{m}{y} \rfloor$. Let $B(n, m, y)$ be the bipartite graph of order n and size m obtained from $K_{q,y}$ by adding one vertex adjacent to $m - yq$ vertices in V and adding $n - y - q - 1$ isolated vertices. We will prove that the extremal bipartite graphs of a given size m and order no more than n, are of the form $B(n, m, y)$ for some $1 \leq y \leq \sqrt{m}$.

A Young diagram is a finite collection of cells arranged in left-justified rows with the row lengths in non-decreasing order. We will consider a different representation of bipartite graphs by means of Young diagrams, as it simplifies the description of the graph operations needed in our proofs. For a bipartite graph $G = (U \cup V, E)$, let us write $U = \{u_1, u_2, \ldots, u_x\}$ and $V = \{v_1, v_2, \ldots, v_y\}$ such that $\deg_G(u_i) \geq \deg_G(u_j)$ and $\deg_G(v_i) \geq \deg_G(v_j)$ whenever $i \leq j$. Let $y_i = \deg_G(u_i)$ and $x_j = \deg_G(v_j)$ for every $1 \leq i \leq x$ and $1 \leq j \leq y$. We associate the diagram T which contains a cell (i, j) if and only if $u_i v_j \in E$. Note that this gives a one-to-one correspondence between diagrams and bipartite graphs.

An example of an extremal bipartite graph with $n = 10$ and $m = 22$ has been represented in these two ways in Figure 4.1. Remark that the number of cells in column i corresponds to the degree of u_i and the number of cells in row j equals the degree of v_j.



Graph representation   Associated Young diagram

Figure 4.1: Two representations of the extremal bipartite $(10, 22)$-*graph* $B(10, 22, 4)$.

In Section 4.2, we prove that the bipartite graphs minimizing the degree-entropy given the size are exactly the complete bipartite graphs. We give two different approaches to prove this. One of them is a proof by induction, and the other one consists in proving that the associated diagram has to be a rectangle. In Section 4.3, we study the possible extremal bipartite graphs of given order n and size m, by proving that under certain restrictions they are of the form $B(n,m,y)$, and we estimate the value of h_d for these graphs. Next, in Section 4.4, we prove that the extremal graphs are indeed of the form $B(n,m,y)$. For this, we use the equivalent representation with Young diagrams, and we give local operations that decrease the degree-entropy (or equivalently, increase h_d). Furthermore, in Section 4.5, we remark that some of the ideas can be applied to other graphical function-indices, such as the Second Zagreb index and the Reciprocal Randić index. Just as has been done in [28], we notice that certain results hold more generally, and we give the essence of the proofs of some known results. Here, we give short proofs for the main results in [71] and [53] about the graphs and bipartite graphs maximizing the degree-entropy in the more general context of a class of graphical function-indices. Finally, in Section 4.6, we give some conclusions on the precise extremal graphs $B(n,m,y)$. We do this based on computational results and the estimates in Section 4.3.

4.1.2 Preliminaries

In this subsection, we define the notions and help functions we will frequently use, as well as give some basic results.

We define the function $f(x) = x \cdot \log(x)$ for $x \geq 0$. Here we use the convention that $f(0) = 0$. Let G be a graph with vertex set $\{v_1, v_2, \dots, v_n\}$. We have $h_d(G) = \sum_{i=1}^{n} f(\deg_G(v_i))$.

For formulating the following inequality, which became known as Karamata's inequality, we only consider non-increasing sequences. We also recall the definition of (strict) majorization, as introduced in Subsection 2.1.2.

Theorem 4.2 ([75]). *Let $(x_1, x_2, \dots, x_n)$ be majorizing $(y_1, y_2, \dots, y_n)$. Then for every convex function g, we have $\sum_{1 \leq i \leq n} g(x_i) \geq \sum_{1 \leq i \leq n} g(y_i)$. Furthermore* *this inequality is strict if the sequences are not equal and g is a strictly convex function. For concave functions, the same holds with the opposite sign.*

It is not hard to see that Lemma 2.1 is a corollary of Karamata's inequality, which was also observed by A. Ghalavand, M. Eliasi and A.R. Ashrafi [63].

We will also make use of so-called difference graphs, which were introduced by P.L. Hammer, U.N. Peled, and X. Sun [67]. The following equivalent characterization is due to N.V.R. Mahadev and U.N. Peled [94].

Theorem 4.3 ([94]). *Let $G = (U \cup V, E)$ be a bipartite graph. The graph G is a difference graph if and only if one of the following equivalent conditions holds:*

(a) there are no $u_1, u_2 \in U$ and $v_1, v_2 \in V$ such that $u_1v_1, u_2v_2 \in E(G)$ and $u_1v_2, u_2v_1 \notin E(G)$;

(b) every induced subgraph without isolated vertices has on each side of the bipartition a domination vertex, that is, a vertex which is adjacent to all the vertices on the other side of the bipartition.

Lemma 4.1. *If G is a graph maximizing $h_d(G)$ among all bipartite graphs of size m, then G is a difference graph.*

Proof. Suppose that G is not a difference graph. By Theorem 4.3 (a), there are four vertices $u_1, u_2 \in U$ and $v_1, v_2 \in V$ such that $u_1v_1, u_2v_2 \in E(G)$ and $u_1v_2, u_2v_1 \notin E(G)$. Assume without loss of generality that $\deg_G(u_1) \geq \deg_G(u_2)$. In that case the degree sequence of the graph $G' = G - u_2v_2 + u_1v_2$ strictly majorizes the degree sequence of G. So we get the desired contradiction by Lemma 2.1. □

Lemma 4.2. *Let x, t, ℓ be fixed positive integers. Under the condition that $\sum_{j=1}^{x} z_i = t$ and all $z_i \geq 0$ are integers, $\sum_{j=1}^{x} f(z_j+\ell) - \sum_{j=1}^{x} f(z_j)$ is maximized when $z_1 = z_2 = \cdots = z_x = \left\lfloor \frac{t}{x} \right\rceil$.*

Proof. Note that the function $\Delta^{\ell}(z) = f(z+\ell) - f(z)$ is a strictly concave function for every $\ell > 0$, and that every sequence of x integers with sum t majorizes the sequence with $z_1 = z_2 = \cdots = z_x = \left\lfloor \frac{t}{x} \right\rceil$. Now the result follows immediately from Karamata's inequality. □

For the function Δ^1, we write Δ for ease of notation. We define it here separately, as it will be used frequently. Recall that $f(x) = x \cdot \log(x)$ for $x \geq 0$.

Definition 4.1. *The function Δ is defined by $\Delta(x) = f(x) - f(x-1) = 1 + \int_{x-1}^{x} \log(t)\,\mathrm{d}t$. This is a strictly increasing and concave function.*

We also use Landau notation, such as $o(\cdot)$ and $\omega(\cdot)$. A function $q(x,y)$ or expression is $o_y(p(x,y))$ or $\omega_y(p(x,y))$ if $\frac{q(x,y)}{p(x,y)} \to 0$ respectively $\frac{|q(x,y)|}{|p(x,y)|} \to \infty$ when $y \to \infty$. Sometimes the dependency on y is removed to keep the notation light. We also use the notation $[k] = \{1,2,\dots,k\}$ and $[k..\ell] = \{k, k+1, \dots, \ell-1, \ell\}$.

4.2 Minimum degree-entropy of bipartite graphs of given size

In this section, we prove Theorem 4.1. In both approaches, we assume that our bipartite graph is $G = (U \cup V, E)$, $U = \{u_1, u_2, \dots, u_x\}$ and $V = \{v_1, v_2, \dots, v_y\}$ are the vertices with degree at least 1. Here we assume these are ordered, i.e., $\deg_G(u_i) \geq \deg_G(u_j)$ and $\deg_G(v_i) \geq \deg_G(v_j)$ whenever $i \leq j$. Let $y_i = \deg_G(u_i)$ and $x_j = \deg_G(v_j)$ for every $1 \leq i \leq x$ and $1 \leq j \leq y$. So with this notation, we have $y = y_1$ and $x = x_1$.

4.2.1 Approach 1

Proof of Theorem 4.1. We prove the statement by induction on m. The base case $m = 1$ is trivial since we only have one edge. So assume the statement is true when the size is at most $m-1$ and let G be a graph maximizing $h_d(G)$ among all graphs of size m. By Lemma 4.1, we can assume that G is a difference graph. This implies that u_1 is a dominating vertex (dominates V). Let $G' = G - u_1$. By the induction hypothesis applied to G' and Lemma 4.2 applied with $\ell = 1$ and $z_i = x_i - 1$, we know

$$h_d(G) = h_d(G') + f(y) + \sum_{j=1}^{y} f(x_j) - \sum_{j=1}^{y} f(x_j - 1)$$

$$\leq h_d(K_{1,m-y}) + f(y) + y\left(f\left(\frac{m}{y}\right) - f\left(\frac{m}{y} - 1\right)\right).$$

Equality here is only attained if $x_j = \frac{m}{y}$ for every $1 \leq j \leq y$. Now the conclusion is direct as

$$\begin{aligned} h_d(G) \leq &(m-y)\log(m-y) + y\log(y) + m(\log(m) - \log(y)) \\ &-(m-y)(\log(m-y) - \log(y)) \\ &= m\log(m). \end{aligned}$$

By induction, we see that equality is attained precisely for the complete bipartite graphs of size m. Using the observation after equations (2.1) and (2.2), we get $I_d(G) \geq 1 + \log(\sqrt{m})$. □

4.2.2 Approach 2

In this subsection, we give an alternative proof for the fact that $K_{1,m}$ minimizes the degree-entropy among all bipartite graphs of size m, which does not use any prerequisites. Note that the proof can be formulated without the notion of a diagram and that we give a short proof of Lemma 4.1 as a claim in this notation.

In every cell (i, j) of the associated diagram T of a bipartite graph G, we put $\log(y_i x_j) = \log(x_j) + \log(y_i)$. The sum over all cells, $h_d(T) = \sum_{(i,j)\in T} \log(x_j y_i)$ is now exactly equal to $\sum_j f(x_j) + \sum_i f(y_i) = h_d(G)$.

Proof of Theorem 4.1. We first prove that the associated diagram T is a Young diagram, i.e., if $(i, j) \in T$ and $0 < i' \leq i$ and $0 < j' \leq j$, then $(i', j') \in T$.

Claim 1. *If $G = (U \cup V, E)$ is a bipartite graph maximizing $h_d(G)$ among all bipartite graphs of size m, then its associated diagram is a Young diagram.*

Proof. Assume $(i, j) \in T$ and $(i', j') \notin T$. If $i' < i$ and $j' < j$, then $G' = G - u_i v_j + u_{i'} v_{j'}$ satisfies $h_d(G') = h_d(G) - \Delta(y_i) - \Delta(x_j) + \Delta(y_{i'} + 1) + \Delta(x_{j'} + 1) > h_d(G)$. The latter holds due to Δ being strictly increasing and $y_i \leq y_{i'}$ and $x_j \leq x_{j'}$. If $i' = i$ or $j' = j$, it is analogous as $h_d(G') - h_d(G)$ equals

$\Delta(y_{i'}+1)-\Delta(y_i)>0$ or $\Delta(x_{j'}+1)-\Delta(x_j)>0$. This implies that G is not extremal and so we reach a contradiction. $\diamondsuit$

Next, we note that

$$\sum_{(i,j)\in T} x_j y_i \le \sum_{i,j} x_j y_i = \left(\sum_j x_j\right)\left(\sum_i y_i\right) = m^2.$$

So by Jensen's inequality, we have

$$\sum_{(i,j)\in T} \log(x_j y_i) \le m\log\left(\frac{m^2}{m}\right) = m\log(m)$$

with equality if and only if $x_j y_i = m$ for every choice of (i,j), i.e., G is complete bipartite. $\square$

4.3 Minimum degree-entropy among dense bipartite graphs

We state two propositions in terms of the Young diagrams. Note that by taking $x = q+1$, we conclude that $B(n,m,y)$ is extremal in both cases.

Proposition 4.1. *Let $x > y > r > 0$ and $m = xy - r$ be integers. Among all Young diagrams with m cells in $[x]\times[y]$, $h_d(T)$ is maximized by*

$$T' = ([x]\times[y]) \setminus (x \times [y-r+1..y]).$$

If $x > y = \omega(r^2)$, then the latter diagram has $h_d(T') \sim m\log m - r + o(1)$.

Proof. Let T be a Young diagram which equals $[x]\times[y]$, except from the r cells which have coordinates $(i_1,j_1),(i_2,j_2),\dots,(i_r,j_r)$, where these are ordered in reverse lexicographic order. In particular $(i_1,j_1)=(x,y)$ and for every $1\le k\le r$ the pair (i_k,j_k) satisfies $i_k+j_k \ge x+y+1-k$, $j_k\le y$ and $i_k\le x$. Since Δ is increasing and concave, by Karamata's inequality, we have

$$\Delta(i_k)+\Delta(j_k) \ge \Delta(x)+\Delta(y+1-k).$$

Now

$$
\begin{aligned}
h_d(T) &= xy\log(xy) - \sum_{k=1}^{r} (\Delta(i_k) + \Delta(j_k)) \\
&\le xy\log(xy) - \sum_{k=1}^{r} (\Delta(x) + \Delta(y+1-k)) \\
&= h_d(T'),
\end{aligned}
$$

and furthermore equality occurs if and only if $(i_k, j_k) = (x, y+1-k)$ for every k, i.e., $T = T'$. Now since $1 + \log(x) - \frac{1}{x-1} \le \Delta(x) < 1 + \log(x)$ and $1 + \log(y) - \frac{k}{y-k} \le \Delta(y+1-k) < 1 + \log(y)$, we note that when $y = \omega(r^2)$, we have

$$
\begin{aligned}
h_d(T') &= xy\log(xy) - r\Delta(x) - \sum_{k=1}^{r} \Delta(y+1-k) \\
&= m\log(m+r) - 2r - o(1) \\
&= m\log(m) - r - o(1),
\end{aligned}
$$

where the $o(1)$ term tends to zero as $x, y \to \infty$ for fixed r. □

Proposition 4.2. *Let $q \ge y > r > 0$ and $m = qy + r$ be integers. Among all Young diagrams containing $[q] \times [y]$ with m cells, $h_d(T)$ is maximized by*

$$
T' = ([q] \times [y]) \cup ((q+1) \times [r]) .
$$

If $q = \omega(r)$, the latter diagram satisfies $h_d(T') \sim m\log(m) - r\log\left(\frac{y}{r}\right)$.

Proof. Let T be a Young diagram which equals $[q] \times [y]$, with r additional cells which have coordinates $(i_1, j_1), (i_2, j_2), \ldots, (i_r, j_r)$, where these are ordered in lexicographic order. We note that for every $1 \le k \le r$ the pair (i_k, j_k) satisfies $i_k + j_k \le q + k + 1$ and $\min\{i_k, j_k\} \le k$. Since Δ is increasing and concave, by Karamata's inequality, we have

$$
\Delta(i_k) + \Delta(j_k) \le \Delta(\min\{i_k, j_k\}) + \Delta(q+1+k-\min\{i_k, j_k\}) \le \Delta(q+1) + \Delta(k).
$$

Now

$$\begin{aligned}
h_d(T) &= yq\log(yq) + \sum_{k=1}^{r}(\Delta(i_k)+\Delta(j_k)) \\
&\le yq\log(yq) + \sum_{k=1}^{r}(\Delta(q+1)+\Delta(k)) \\
&= h_d(T'),
\end{aligned}$$

and furthermore equality occurs if and only if $(i_k, j_k) = (q+1, k)$ for every k, i.e., $T = T'$.

Now since $1+\log(q) \le \Delta(q+1) < 1+\log(q)+\frac{1}{q}$ and $\sum_{k=1}^{r}\Delta(k) = f(r) = r\log(r)$, we note that when $q = \omega(r)$, we have

$$\begin{aligned}
h_d(T') &= yq\log(yq) + r\Delta(q+1) + f(r) \\
&= qy\log(q) + qy\log(y) + r\log(q) + r + f(r) + o(1) \\
&= m\log(m-r) - r\log(y) + r + r\log(r) + o(1) \\
&= m\log(m) - r\log\left(\frac{y}{r}\right) + o(1),
\end{aligned}$$

where the $o(1)$ term tends to zero when $q \to \infty$ for fixed r. $\square$

4.4 Minimum is attained by dense Young diagrams

In this section, we list and prove a number of lemmas showing that the extremal Young diagrams are dense, in the sense that they are of the form presented in Section 4.3.

We start with proving that the extremal Young diagram cannot have the following specific form.

Lemma 4.3. *Let $m = xy - r - s + 1$ with $1 \le r \le x - y$ and $1 \le s < y$, such that $r + s - 1 = wy + r'$ for some integers $0 < w$ and $0 \le r' < y$. If*

$$T = ([x]\times[y])\setminus([x-r+1..x]\times y)\setminus(x\times[y-s+1..y-1]) \text{ and}$$
$$T' = ([x-w]\times[y])\setminus\big((x-w)\times[y-r'+1..y]\big),$$

then $h_d(T) < h_d(T')$. See Figure 4.2.

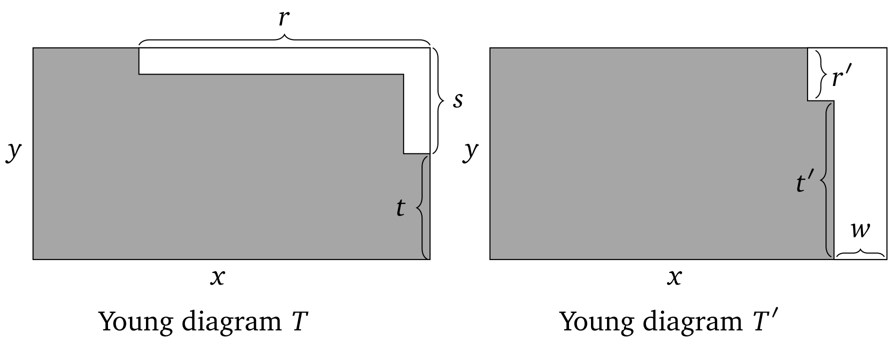


Figure 4.2: The Young diagrams from Lemma 4.3.

Proof. We prove this in a number of steps. First we show that it is sufficient to prove it for $s = 1$.

Claim 2. *If Lemma 4.3 does hold whenever $s = 1$, then so it does for all other cases.*

Proof. Assume there are choices of x, y, r, s (and thus Young diagrams T and T') for which Lemma 4.3 is false. Let $t = y - s$ and $t' = y - r'$, and assume $0 < t < y - 1$. We now consider two cases.

Case 1. $t' \leq t$.

In this case, we note that $\Delta(t+1) + \Delta(x) > \Delta(t'+1) + \Delta(x-w)$ and thus we also have a counterexample with $s-1$ instead of s. We can repeat this until $s = 1$.

Case 2. $t' > t$.

Note that $y \cdot (x - w) > m > x \cdot (y - 1)$, implying $\frac{x-w}{x} > \frac{y-1}{y} > \frac{t}{t+1} \geq \frac{t}{t'}$ and hence $(x - w)t' > xt$. Now we have $\Delta(t) + \Delta(x) < \Delta(t') + \Delta(x - w)$. When $t + x \leq (x - w) + t'$, this is by Karamata's inequality and the fact that Δ is increasing and concave. When $t+x > (x-w)+t'$, we note that $(x-u)(t-u) = xt - u(x + t) + u^2 < (x - w)t' - u((x - w) + t') + u^2 = (x - w - u)(t' - u)$ for every $u \geq 0$ and thus

$$\int_{u=0}^{1} \log\left((x-u)(t-u)\right) \mathrm{d}u < \int_{u=0}^{1} \log\big((x-w-u)(t'-u)\big) \mathrm{d}u.$$

The latter is equivalent to $\Delta(t)+\Delta(x)<\Delta(t')+\Delta(x-w)$. This implies that we can construct a counterexample with $s+1$ instead of s as well and we can repeat this until $s=y$, which corresponds to an example with $s=1$. $\diamondsuit$

Next, we show that we can assume that $x-r=y$.

Claim 3. *If Lemma 4.3 does hold whenever $s=1$ and $r=x-y$, then so it does for all other cases.*

Proof. Assume there are Young diagrams T and T' for which Lemma 4.3 is false, where $s=1$ and $y+r<x$. Now deleting the first column of both diagrams implies that $h_d(T)$ decreases by $\Delta(x-r)+(y-1)\Delta(x)+f(y)$ and $h_d(T')$ has been decreased by $(y-r')\Delta(x-w)+r'\Delta(x-w-1)+f(y)$. Since $\{\underbrace{x,x,\dots,x}_{y-1},x-r\}$ majorizes $\{\underbrace{x-w,x-w,\dots,x-w}_{y-r'},\underbrace{x-w-1,\dots,x-w-1}_{r'}\}$, as Δ is strictly concave, the value $h_d(T')$ has decreased by a larger amount than $h_d(T)$. So Lemma 4.3 is false for a construction with parameters $(x-1,y,r)$. We can repeat this, until $x=y+r$. $\diamondsuit$

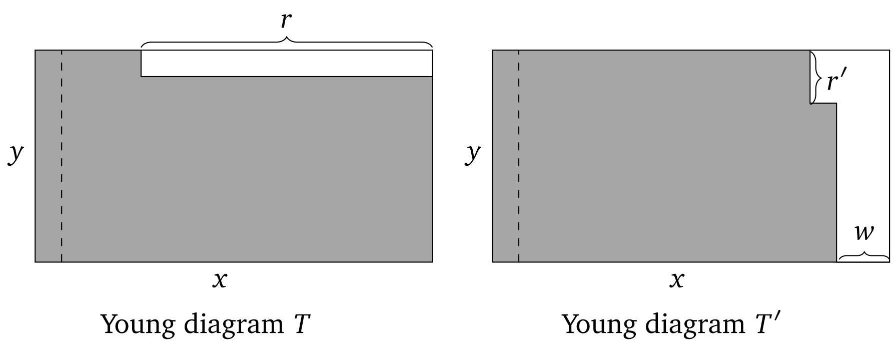
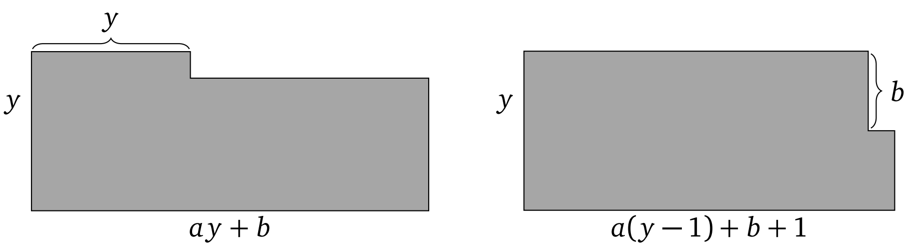


Figure 4.3: The two Young diagrams from Claim 3.

Now we finish the proof by proving that Lemma 4.3 is true when $s=1$ and $r=x-y$. In the latter case, we can write $x=ay+b$ where $a>0$ is an integer and $0\le b<y$, in which case we know the precise shapes of T and T'. These have been represented in Figure 4.4.



Young diagram T Young diagram T'

Figure 4.4: The two Young diagrams in the remaining case.

When $a = 1$, we have $h_d(T) \leq h_d(T')$ by Proposition 4.1, and the inequality is even strict when $b > 1$. We now finish the proof by induction on a. So assume it is proven for the values a, b, y. Going from a to $a + 1$, the value $h_d(T)$ increases by

$$I_1 = y \cdot f(y-1) + (y-1)(f((a+1)y+b) - f(ay+b)) \tag{4.1}$$

while $h_d(T')$ increases by at least

$$I_2 = (y-1) \cdot f(y) + y\left(f\left(\left(a+1+\frac{b}{y}\right)(y-1)\right) - f\left(\left(a+\frac{b}{y}\right)(y-1)\right)\right). \tag{4.2}$$

Now one can note that $I_1 = I_2$ by implementing the following two equalities in the equation (4.1) and equation (4.2).

$$f((a+1)y+b) - f(ay+b) = y + y\int_{u=a+\frac{b}{y}}^{a+1+\frac{b}{y}} (\log(u) + \log(y))\, \mathrm{d}u$$

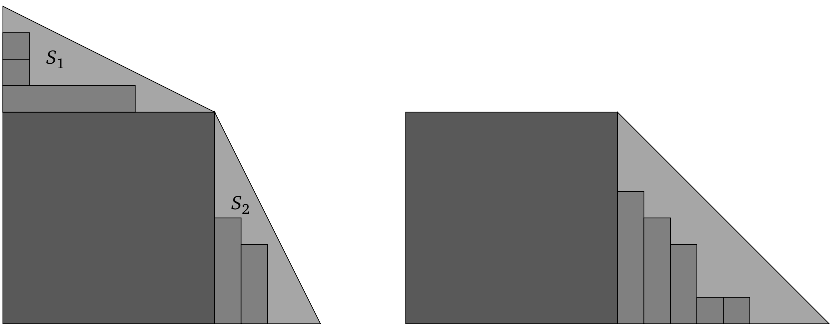
$$\begin{aligned} &f\left(\left(a+1+\frac{b}{y}\right)(y-1)\right) - f\left(\left(a+\frac{b}{y}\right)(y-1)\right) \\ =&y - 1 + (y-1)\int_{u=a+\frac{b}{y}}^{a+1+\frac{b}{y}} (\log(u) + \log(y-1))\, \mathrm{d}u. \end{aligned}$$

□

Having proven this particular case in Lemma 4.3, we continue with some observations from which we can conclude that the extremal diagrams will be as given in Section 4.3.

Lemma 4.4. *If $G = (U \cup V, E)$ is an extremal bipartite of order n and size m, then the minimum degree of the smaller partition class is at least the maximum degree of the other partition class.*

Proof. Consider the associated Young diagram T of G. Let i be the value for which $(i,i) \in T$ and $(i+1,i+1) \notin T$. Let $S_1 = \{(i',j') \in T \mid j' > i\}$ and $S_2 = \{(i',j') \in T \mid i' > i\}$. Now if both S_1 and S_2 are non-empty, we can construct a diagram T' for which the rows in S_1 are deleted and are added as columns, in such a way that another Young diagram T' has been formed. It is not hard to see that the degree sequence of the graph G' associated with T' majorizes the degree sequence of G. For this, note that it is sufficient to compare the degrees of $\{u_1, u_2, \ldots, u_i, v_1, \ldots, v_i\}$ (and thus the number of cells in their rows/columns). Thus, since f is strictly convex, by Karamata's inequality we know $h_d(T') > h_d(T)$. □



Young diagram T Young diagram T'

Figure 4.5: Sketch of the rearrangement in Lemma 4.4.

Theorem 4.4. *For fixed integers m and n, let T_* be a diagram with m cells for which the sum of its length x and width y is at most n. If $h_d(T_*)$ is maximal among all such diagrams, then x, y satisfy $xy - \min\{x, y\} < m \leq xy$.*

Proof. Assume m cannot be written as a product xy with $x+y \le n$, and T_* is an extremal diagram with $m \le (x-1)y$, where $y < x$. Let $x_1 = x, x_2, \dots, x_y$ be the number of cells in every row. By Lemma 4.4, we know that $[y] \times [y] \subset T_*$, and thus $x_i \ge y$ for every $1 \le i \le y$. Furthermore, since $m \le (x-1)y$, there is a smallest index i such that $x_i < x-1$. Now, the first i rows of T_*, i.e., $T_i = T_* \cap ([x] \times [i])$, form a diagram that has the form of the diagram T in Lemma 4.3. Replacing T_i by the corresponding T_i' from Lemma 4.3 (which does not increase the length of the diagram) will imply that we also form a T_*' for which $h_d(T_*') > h_d(T_*)$, since $h_d(T_*') - h_d(T_*) = h_d(T_i') - h_d(T_i) > 0$. For the latter, note that the only rows and columns that have possibly changed, are rows 1 until i and columns $x_i + 1$ until x, all of which do not contain any cell outside T_i or T_i'. This is the desired contradiction. Figure 4.6 presents this final comparison. □

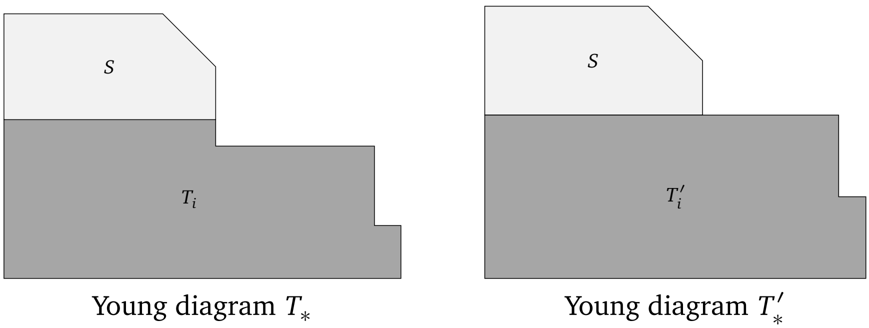


Figure 4.6: The local move increasing $h_d(T_*)$ in Theorem 4.4.

4.5 Extremal values of other graphical function-indices

As we have argued before, the degree-entropy is just one example of a graphical function-index. Let $g(x,y)$ be a real symmetric function. Then the associated graphical function-index is

$$GFI_g(G) = \sum_{uv \in E(G)} g(\deg_G(u), \deg_G(v)).$$

For some special examples of g, this unifying approach has been considered in [65]. For an overview of more examples that are covered by the unifying expression, we refer the reader to [90, Table 1].

There also exist graphical indices as defined in [114, 126] by a real function $g(x)$ and the expression $\sum_{v\in V(G)} g(\deg_G(u))$. This is a special case of GFI_g since $\sum_{uv\in E}(\frac{g(\deg_G(u))}{\deg_G(u)} + \frac{g(\deg_G(v))}{\deg_G(v)}) = \sum_{u\in V} g(\deg_G(u))$ (i.e., this function-index is obtained by choosing $g(x,y) = \frac{g(x)}{x} + \frac{g(y)}{y}$ in the above expression for GFI_g). Therefore, the function $h_d(G) = \sum_{i=1}^n \deg_G(v_i)\log(\deg_G(v_i))$ can be written as $h_d(G) = \sum_{i=1}^n g(\deg_G(v_i))$ by choosing $g(x) = x\log(x)$ (i.e., by choosing $g(x,y) = \frac{x\log(x)}{x} + \frac{y\log(y)}{y} = \log(xy)$ in the above expression).

Our idea of approach 2 in Subsection 4.2.2 works for any GFI_g whenever $g(x,y)$ is an increasing strictly concave function in xy. For example, when $g(x,y) = xy$ or $g(x,y) = \sqrt{xy}$ (i.e., for the Second Zagreb index and Reciprocal Randić index), the graphs maximizing $GFI_g(G)$ among all bipartite graphs of size m are precisely the complete bipartite graphs $K_{q,y}$ with $q\cdot y = m$. When we restrict both the order and size, one can expect that a similar exposition implies that the extremal graphs are again near complete bipartite, i.e., $K_{y,x}$ with some small number r of removed (or added) edges. K. Xu, K. Tang, H. Liu and J. Wang [124] studied these for the First and Second Zagreb indices, for example.

Theorem 4.5. *Let G be a bipartite graph of size m, and let $g(x,y)$ be an increasing strictly concave function in xy.Then $GIF_g(G)$ attains the minimum value if and only if G is a complete bipartite graph.*

If $g(x,y) = x+y$ or $g(x,y) = (x+y)^2$, since $\deg_G(u)+\deg_G(v) \le m+1$ for every $uv \in E(G)$ the unique extremal graph (maximizing $GFI_g(G)$) among all bipartite graphs of size m is easily seen to be the star $K_{1,m}$. In particular, this implies that among bipartite graphs of a given size, the set of graphs maximizing the First Zagreb index (i.e., for $g(x,y) = x+y$) and Second Zagreb index (i.e., for $g(x,y) = xy$) are not equal. This contrasts some of the intuition in the concluding section of [124, Section 4].

Also among bipartite graphs of a given order and size, the bipartite graphs with the maximum First Zagreb index might be different from the one with the maximum Second Zagreb index. For the First Zagreb index these extremal bipartite graphs were determined by S. Zhang and C. Zhou [128]. Here the asymmetry (large degrees) plays a major role, and so there are examples for which the bipartite graphs maximizing the First Zagreb index are not complete bipartite, while there do exist complete bipartite graphs $K_{y,x}$ with $xy = m$ and $x + y < n$ that maximize the Second Zagreb index.

Finally, we prove that the graphs which attain the maximum degree-entropy can be easily determined, and are the graphs for which the degree sequence is as balanced as possible. We prove these statements by showing that the latter graphs are those which minimize $GFI_g(G)$ for any increasing strictly convex function $g(x, y)$ in xy. Thus, we give a short alternative proof for the result in [53] in this more general setting, and for the result of [71].

Note that $h_d(mK_2) = 0$. Among all graphs with fixed size m, without constraints on the order, we have the following result.

Proposition 4.3. *Let G be a graph of size m and order $n \geq 2m$. Then $I_d(G) \geq \log(2m)$, with equality if and only if $G \cong mK_2 \cup \overline{K}_{n-2m}$.*

If there is a condition on both the order and size, the extremal graphs and bipartite graphs are precisely the almost regular graphs, as proven in the following two propositions. C.E. Cheng, Y. Guo, S. Zhang and Y. Du [32] determined these bipartite graphs with the minimum value of the First Zagreb index for bipartite graphs of a given order and size. The proof is essentially a corollary of Karamata's inequality and noting that there is a degree sequence that is majorized by all other possible realizations. We also make use of a result attributed to Walecki. A *decomposition* of a graph is a collection of subgraphs such that each edge belongs to exactly one subgraph. The complete graph K_{2k+1} admits a decomposition into Hamiltonian cycles and K_{2k} can be decomposed into Hamiltonian cycles and a perfect matching, see [5, 6].

Proposition 4.4. *Let G be a graph of size m and order n, and let $g(x)$ be a strictly concave function in x. Among all graphs with fixed size m and order n, $GFI_g(G) = \sum_{v \in V(G)} g(\deg_G(u))$ attains the maximum value if and only if G is almost regular (i.e., $\deg_{max}(G) - \deg_{min}(G) \leq 1$).*

Proof. Since the degree sequence of any almost regular graph is majorized by the degree sequence of any other graph of order n and size m, the result follows by Karamata's inequality. The characterization of the extremal graphs is a consequence of g being strictly convex. The existence of an almost regular graph (even connected if $m \geq n-1$ when necessary) is immediate, and implied by the decomposition of K_n into Hamiltonian cycles and at most one perfect matching. If $m = an + b$, then one can take the union of a Hamiltonian cycles and a matching of size b (for $0 \leq b \leq \frac{n}{2}$) or take $a+1$ Hamiltonian cycles and remove a matching of size $n-b$ from one Hamiltonian cycle (if $b > \frac{n}{2}$). $\square$

Proposition 4.5. *Let $G = (U \cup V, E)$ be a bipartite graph of size m and order n, and let $g(x)$ be a strictly concave function in x. Among all bipartite graphs of size m and order n, $GFI_g(G) = \sum_{v \in V(G)} g(\deg_G(u))$ attains the maximum value if and only if $G = (U \cup V, E)$ satisfies $|U| = \lceil \frac{n}{2} \rceil$ and $|V| = \lfloor \frac{n}{2} \rfloor$ for which the degrees in one partition class differ by at most one.*

Proof. Let $G' = (U' \cup V', E')$ be any other bipartite graph of size m and order n whose partition classes have size $|U'| = u \geq v = |V'|$. The sum of the $i \leq v$ largest degrees in V' is at least equal to the sum of the i largest degrees of G, and the sum of the $i \leq u$ smallest degrees in U' is at most the sum of the i smallest degrees in G. Here we use that $u \geq \lceil \frac{n}{2} \rceil$ and $v \leq \lfloor \frac{n}{2} \rfloor$, and the degrees in G are as balanced as possible. Remark that if $u \geq i > \lceil \frac{n}{2} \rceil$, then the sum of the i degrees in U' is at most m. So we conclude that the degree sequence of G is majorized by any degree sequence of any other bipartite graph of order n and size m, from which we obtain the conclusion (also for the uniqueness statement) by Karamata's inequality and $g(x)$ being strictly concave.

We remark that one can always construct at least one such a balanced bipartite graph G. If n is even, just partition $K_{n/2,n/2}$ in perfect matchings and add the edges from one matching at a time up to the point you selected precisely m edges. When n is odd, for every $1 \leq m \leq \left\lfloor \frac{n^2}{4} \right\rfloor$, one can construct a graph by adding for every $1 \leq k \leq m$ an edge between the vertices a_i and b_j for which $k \equiv i \pmod{\lceil \frac{n}{2} \rceil}$ and $k \equiv j \pmod{\lfloor \frac{n}{2} \rfloor}$ respectively. $\square$

Let $G = (V, E)$ be a graph of size m and order n, and let $g(x) = c_1 - c_2(x \log(x))$ with $c_1 = \frac{\log(2m)}{n}$ and $c_2 = \frac{1}{2m}$. We have $I_d(G) = \sum\limits_{u \in V(G)} g(\deg_G(u))$. Because $g(x)$ is strictly concave for $x > 0$, by Propositions 4.4 and 4.5, we have the following two results, respectively.

Corollary 4.1. *Let G be a graph of size m and order n. Among all graphs of size m and order n, $I_d(G)$ attains the maximum value if and only if G is almost regular (i.e., $\deg_{max}(G) - \deg_{min}(G) \leq 1$).*

Corollary 4.2. *Let $G = (U \cup V, E)$ be a bipartite graph of size m and order n. Among all bipartite graphs of size m and order n, $I_d(G)$ attains the maximum value if and only if $G = (U \cup V, E)$ satisfies $|U| = \lceil \frac{n}{2} \rceil$ and $|V| = \lfloor \frac{n}{2} \rfloor$ for which the degrees in one partition class differ by at most one.*

A few examples have been presented in Figure 4.7. Since everything boils down to having a degree sequence that is as balanced as possible, i.e., it is majorized by any other degree sequence of a graph, the result only depends on the degree sequence.

We end with a few observations. There do exist both connected and disconnected extremal bipartite graphs (as presented for $m = n = 9$). If n is odd and m is fixed, it is possible there is no extremal bipartite graph attaining the maximum over all graphs, since for bipartite graphs the maximum degree might exceed the minimum degree by at least 2. Finally, we observe that the partition itself does not necessarily have to be balanced (e.g., there exists an example with $m = 30, n = 22, |U| = 10$ and $|V| = 12$).

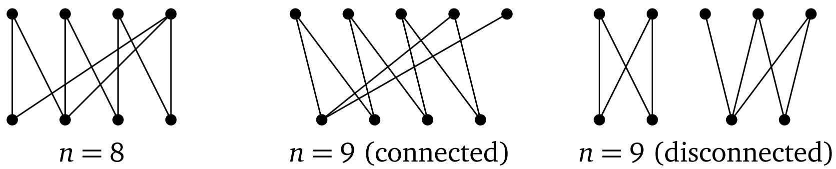


Figure 4.7: Bipartite graphs of size 9 with maximum degree-entropy.

4.6 Concluding remarks

In this chapter, we have shown that the bipartite graphs of size m that maximize the degree-entropy are complete bipartite graphs. If there is a restriction on the order, and m cannot be written as $x \cdot y$ where $x + y \leq n$, then the extremal graphs are of the form $B(n, m, y)$. Let us recall that $B(n, m, y)$ is a bipartite graph of order n and size m obtained from $K_{q,y}$ by adding one vertex adjacent to $m-yq$ vertices in one part of size y and adding $n-y-q-1$ isolated vertices, where $q=\left\lfloor \frac{m}{y} \right\rfloor$. Despite these observed facts, it seems hard to determine a general formula for y that minimizes the degree-entropy of $B(n,m,y)$, as we will explain here.

Using a computer program[1], we were able to find all extremal bipartite graphs for $n \le 50$ and $m \le \left\lfloor \frac{n^2}{4} \right\rfloor$. Interested readers can have a closer look at the output[2]. For every pair (n,m) and value $1 \le y \le \sqrt{m}$, we computed the values

$$q=\left\lfloor \frac{m}{y} \right\rfloor,\ x=\left\lceil \frac{m}{y} \right\rceil,\ m-qy,\ xy-m,\ h=h_d(B(n,m,y)),$$

as well as three Boolean values expressing whether $B(n,m,y)$ achieved the maximum value of h_d among all choices of y and whether $m-qy$ or $xy-m$ were minimal among all possible choices for y for that pair (n,m). Among all extremal graphs which are not complete bipartite, we found 547 cases for which both $xy-m$ and $m-qy$ are minimal among all possibilities, 375 cases where only $m-qy$ is minimal, 3635 cases where only $xy-m$ is minimal, and no cases for which neither $xy-m$ nor $m-qy$ is minimal.

Intuitively, when $xy-m$ or $m-qy$ is small, the graph $B(n,m,y)$ is nearly complete bipartite (up to the edges of a small star which have been added or removed), and so it might be extremal. Nevertheless, this was found not to be the case in general. When $n=17726$ and $m=318728$, the extremal graph is $B(n,m,18)$, which satisfies $q=17707$, $y=18$ and $r=m-qy=2$, even though $y=139$ and $q=2293$ would give $r=1$. Taking into account the estimates in Proposition 4.2, one can expect that there are more such examples. The difference with the upper bound depends on both r and y, i.e., one would like both y and r to be small, but the ratio has an influence on this. A precise statement here seems to be difficult to derive, as the above discussion is related to some hard number theoretic questions.

[1] `https://github.com/MatteoMazzamurro/extrema-graph-entropy/blob/main/bipartite_graphs_entropy.R`

[2] `https://github.com/MatteoMazzamurro/extrema-graph-entropy/blob/main/B_n_m_50.csv`

Consider the extremal bipartite graph associated with the extremal Young diagram of Proposition 4.1. The statement in this proposition implies that a complete bipartite graph $K_{x,y}$ with $m+r$ edges is obtained from the extremal graph after adding r edges. Hence, when m is large, one might hope that there is a relatively small value of r for which $m+r$ can be written as a product xy with $x+y \le n$. In such cases, the extremal graph will be $B(n,m,y)$, with y chosen in such a way that $y\left\lceil \frac{m}{y} \right\rceil - m$ is minimized under the constraint that $y + \left\lceil \frac{m}{y} \right\rceil \le n$, and so the extremal graph is a complete bipartite graph minus a small star. Furthermore, when there are multiple choices for y, we need to choose the smallest y, i.e., for fixed r, the complete bipartite graph will be as asymmetric (unbalanced in terms of the bipartition classes) as possible. To note the latter, one can compare the computations in Proposition 4.1 for some $x' > x > y > y'$ with $xy = x'y'$. For a fixed $1 \le k \le y-1$, we have $\Delta(x) + \Delta(y+1-k) \ge \Delta(x') + \Delta(y'+1-k)$, since $(x-u)(y+1-k-u) \ge (x'-u)(y'+1-k-u)$ for $u \in [0,1]$.

We can prove that the intuition that $r = xy - m$ is minimal for the extremal $B(n,m,y)$ is true in many cases, using the notion of M-smooth numbers. A positive integer is called M-smooth (or M-friable) if its largest prime divisor is at most M. At this point, the state-of-the-art on smooth numbers [95, 64] does not address the question if for fixed constants $0 < u < 1$ and $c > 0$, the interval $[m, m + c\log(m)]$ necessarily needs to contain an m^u-smooth number whenever m is sufficiently large, but it does so for almost all m by [95, Cor. 6].

Proposition 4.6. *Fix a $\delta > 0$. For almost every pair (n,m) with $\omega(n^{1+\delta}) < m < o(n^{3/2})$, the graph minimizing the entropy among all bipartite graphs of order (bounded by) n and size m is the graph $B(n,m,y)$ where y is chosen such that $r = xy - m \ge 0$ is minimal among all possibilities of x, y with $x + y \le n$.*

Proof. Let n and m be a random pair of values satisfying the inequality of the statement. Write $n = (m+\log(m))^u + (m+\log(m))^{1-u}$ for some real number $\frac{2}{3} \le u < 1$. Now with high probability, there exists a number $r \le \frac{1}{2}(1-u)\log(m)$ such that $m+r$ is m^u-smooth (by [95, Cor. 6]). Furthermore, $m+r$ being m^u-smooth implies that all its prime divisors are bounded by m^u and thus also $(m+r)^u$. Therefore $m+r$ has two divisors x, y for which $(m+r)^{1-u} \le y \le x \le (m+r)^u$; note that if the largest prime factor p of $m+r$

satisfies $(m+r)^{1-u} \leq p \leq (m+r)^u$, we can take $\{x, y\}$ to be equal to $\left\{p, \frac{m+r}{p}\right\}$ and otherwise we can take a product P of prime divisors of $m+r$ which is between $(m+r)^{1-u}$ and $(m+r)^{2(1-u)} \leq (m+r)^u$ and take $\{x, y\}$ to be equal to $\left\{P, \frac{m+r}{P}\right\}$. Since the function $P + \frac{m+r}{P}$ is increasing for $P \geq \sqrt{m+r}$ and $r < \log(m)$, we have $x+y < n$. Comparing Proposition 4.2 and Proposition 4.1, we conclude that the extremal graph is equal to the graph $B(n, m, y)$, where y has been chosen in such a way that $xy - m$ is minimized. The latter is because $y + \frac{m}{y} \leq n$ and $y < \sqrt{m}$ implies that $y \gg m^{0.5(1-u)}$, so $\log(y) - r$ is large. $\square$

On the other hand, there will be infinitely many examples which are not of this form. A simple example for this, where $m - qy$ can be arbitrarily large, can be constructed as follows. Let $c > 0$, and let $n = 2a + 1$ and $m = a^2 + c$ with $a = b^2 + b$. Now $B(n, m, a)$ is an example for which $m - qy = c$ is minimized (note that $q + y = 2a$ is needed). On the other hand $B(n, m, y)$ will always satisfy $xy - m \geq 2b - c$ since (note that $x + y = n$ is needed)

$$(a-b+1)(a+b)-m = a^2+a-b^2+b > m > a^2+a-b^2-b = (a-b)(a+b+1).$$

For fixed c, and $b \gg c$, we have $2b - c \gg c \log\left(\frac{2b-c}{c}\right)$.

It will be clear from the above discussion that a full characterization of the extremal bipartite graphs for degree-entropy remains a big challenge.

Chapter 5

Minimum values of degree-entropy of graphs

In this chapter, we continue our previous work in Chapters 2, 3 and 4 by considering extremal problems involving the degree-entropy I_d as it was defined in equation (2.1). For a graph with a given order and size achieving the minimum entropy value, we derive its unique structure.

5.1 Introduction

The main result of this chapter will be presented in the next subsection after recalling some definitions and notation.

5.1.1 Main result

In order to state our main result, we need to introduce some terminology and notation.

Since we consider graphs with a given order and size, it is convenient to use the term (n,m)-*graph* as shorthand for a graph of order n and size m.

Let G and G' be two disjoint graphs. Then the *union* of G and G', denoted by $G \cup G'$, is the graph with $E(G \cup G') = E(G) \cup E(G')$ and $V(G \cup G') =$

$V(G) \cup V(G')$. We use $\overline{G}$ to denote the *complement* of G. Let m be a positive integer. Then we let $k^* = \max\{k \in \mathbb{N} : \binom{k}{2} \leq m\}$ and $t^* = m - \binom{k^*}{2}$. By $\omega(G)$ denote the clique number of G. It is trivial that $\omega(G) \leq k^*$ for a graph G of size m.

We use $K(k,t)$ to denote the graph obtained from K_k by adding a new vertex and joining it to t vertices of K_k by edges. We also use the following indicator function σ defined by

$$\sigma(x) = \begin{cases} 0, & \text{if } x = 0; \\ 1, & \text{otherwise.} \end{cases}$$

Motivated by earlier results on extremal graphs and values for the degree-entropy in [53, 29, 63, 72], within the class of graphs with a given order and size, we determine the graphs which minimize the degree-entropy.

Theorem 5.1. *Let n and m be integers with $n \geq 2$ and $1 \leq m \leq \binom{n}{2}$, and let G be an (n,m)-graph. Then*

$$I_d(G) \geq \log(2m) - \frac{t^* k^* \log(k^*) + (k^* - t^*)(k^* - 1)\log(k^* - 1) + t^* \log(t^*)}{2m},$$

with equality if and only if $G \cong K(k^*, t^*) \cup \overline{K}_{n-k^*-\sigma(t^*)}$.

Our proof of Theorem 5.1 is postponed to Section 5.2. As in the previous chapter, we will study the extremal problems of $I_d(G)$ by making use of the function $h_d(G) = \sum_{i=1}^{n} \deg_G(v_i) \log \deg_G(v_i)$. This is again based on the observation following the equations (2.1) and (2.2).

In the next sections, we list some known results and deduce some auxiliary results that we will use in our proof of Theorem 5.1.

5.1.2 Preliminaries

We start with recalling some of the basic concepts, and we focus on so-called threshold graphs. They will play a key role in this chapter.

Recall that a degree sequence of a graph is called a *graphical sequence*, and a graph with degree sequence D is called a *realization* of D. Let $G = (V, E)$

be a graph, let $0 < d_1 < \cdots < d_s$ be all positive distinct degrees of G, and let $d_0 = 0$. Moreover, let $D_i = \{v \in V : \deg_G(v) = d_i\}$ for $i = 0, 1, \ldots, s$. Then the sequence $D_0, D_1, \ldots, D_s$ is called the *degree partition* of G. If the number of vertices in G with degree d_i is a_i for $i = 0, 1, \ldots, k$, then we denote by $D(G) = (\underbrace{d_k, \ldots, d_k}_{a_k}, \underbrace{d_{k-1}, \ldots, d_{k-1}}_{a_{k-1}}, \ldots, \underbrace{d_1, \ldots, d_1}_{a_1}) = [d_k^{a_k}, d_{k-1}^{a_{k-1}}, \ldots, d_1^{a_1}]$ the degree sequence of G in which $0 = d_0 < d_1 < \cdots < d_k$ and $a_0 + a_1 + \cdots + a_k = |V(G)|$. Note that we disregard any vertices with degree 0, as they have no influence on the value of I_d.

Recall that by $N_G(v)$ we denote the set of neighbors of vertex v in G. We also use $N_G[v] = N_G(v) \cup \{v\}$. Let S be a nonempty subset of $V(G)$. By $\deg_{\max}(S)$ and $\deg_{\min}(S)$ we denote the maximum degree and minimum degree among the vertices in S, respectively. Let A and B be two non-increasing (degree) sequences. We recall that by $A \succ B$ we indicate that A strictly majorizes B.

In [33], V. Chvátal and P.L. Hammer introduced the family of threshold graphs.

Definition 5.1. *A graph G with vertex set $V = \{v_1, \ldots, v_n\}$ is called a* threshold graph *if each vertex v_i of G can be assigned a non-negative real number w_i, and G can be assigned a non-negative real number r such that $v_i \in V(G)$ and $v_j \in V(G)$ are adjacent in G if and only if $w_i + w_j > r$.*

Using this definition, V. Chvátal and P.L. Hammer [33] established the following fact.

Fact 5.1 ([33]). *If G is a threshold graph, then every induced subgraph of G is a threshold graph.*

For stating the next known result about threshold graphs, we say that a degree sequence is a *threshold sequence* if it is the degree sequence of a threshold graph.

Theorem 5.2 ([94]). *A graphical sequence is a threshold sequence if and only if it has a unique labeled realization.*

Let G be a graph with degree partition $D_0, D_1, \ldots, D_s$. Combining some results from [33] and [94], we have the following three equivalent characterizations of threshold graphs.

Theorem 5.3 ([33, 94]). *The following three statements are equivalent:*

(a) *G is a threshold graph;*

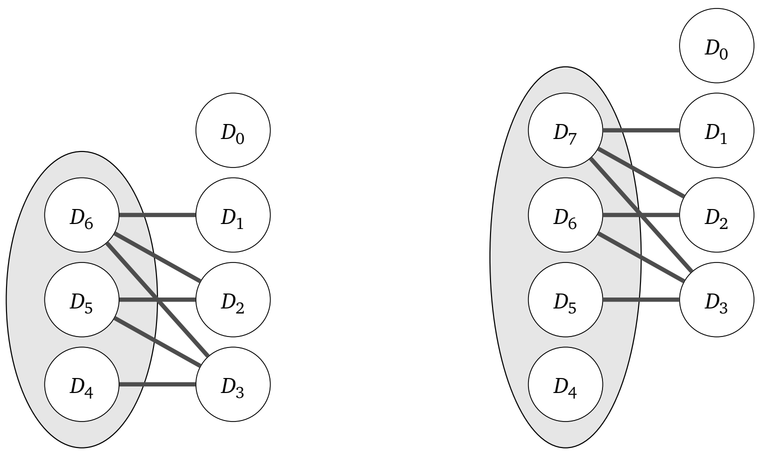
(b) *G does not have an alternating 4-cycle (i.e., there are no four vertices $u, v, w, x \in V(G)$ such that $uw, vx \notin E(G)$ and $uv, wx \in E(G)$);*

(c) *for each $v \in D_k$,*

$$N_G(v) = \cup_{j=1}^{k} D_{s+1-j} \text{ for } k = 1, 2, \ldots, \lfloor \frac{s}{2} \rfloor;$$

$$N_G[v] = \cup_{j=1}^{k} D_{s+1-j} \text{ for } k = \lfloor \frac{s}{2} \rfloor + 1, \lfloor \frac{s}{2} \rfloor + 2, \ldots, s,$$

in other words, for $u \in D_i$ and $v \in D_j$, u is adjacent to v if and only if $i + j > s$; Figure 5.1 illustrates this with $s = 6$ and $s = 7$.



(a) $s = 6$ (b) $s = 7$

Figure 5.1: An illustration of the degree partitions of two threshold graphs with $s = 6$ and $s = 7$.

Figure 5.1 illustrates the degree partitions of two threshold graphs with $s = 6$ and $s = 7$, respectively. A line between D_i and D_j indicates that every vertex of D_i is adjacent to every vertex of D_j. An oval indicates that the included vertices form a clique.

Theorem 5.3 (c) indicates the following theorem which shows the relation between degrees and degree partitions [94].

Theorem 5.4 ([94]). *For any threshold graph, we have*

$$d_{k+1} = d_k + |D_{s-k}| \text{ for } k = 0, 1, \ldots, s,\ k \neq \lfloor s/2 \rfloor;$$
$$d_{k+1} = d_k + |D_{s-k}| - 1 \text{ for } k = \lfloor s/2 \rfloor.$$

As a consequence of Theorem 5.2, we know that a threshold graph is uniquely determined by its degree sequence. Similarly, from Theorem 5.4, we learn that a threshold graph is also uniquely determined by its degree partition.

Before proving Theorem 5.1 in the next section, at the end of this section we first obtain the following four useful lemmas.

Lemma 5.1. *If G is a threshold graph with degree partition $D_0, D_1, \ldots, D_s$, then $|D_{\lceil \frac{s}{2} \rceil}| \geq 2$.*

Proof. We prove the statement by contradiction. Suppose that $|D_{\lceil \frac{s}{2} \rceil}| = 1$. By straightforward calculations, we have that $\lceil \frac{s+1}{2} \rceil = \lfloor \frac{s}{2} \rfloor + 1 > \lfloor \frac{s}{2} \rfloor$. Let d_i be the degree of the vertices in D_i for $i = 0, 1, \ldots, s$. Using Theorem 5.3 (c), we obtain $d_{\lceil \frac{s+1}{2} \rceil} = \sum_{i=\lceil \frac{s}{2} \rceil}^{s} |D_i| - 1 = \sum_{i=\lceil \frac{s}{2} \rceil + 1}^{s} |D_i| = d_{\lfloor \frac{s}{2} \rfloor}$, which contradicts the assumption that $d_i > d_j$ for $i > j$. □

We apply Lagrange's mean value theorem to prove the following result which will be used in our proof of Theorem 5.1.

Lemma 5.2. *Let a be a positive integer, and let $A = (a_1, a_2, \ldots, a_n)$ be a positive integer sequence. If f is a strictly concave function and $\sum_{i=1}^{n} a_i = a$, then $\sum_{i=1}^{n} f(a_i)$ attains its maximum value if and only if $a_i = \lceil \frac{a}{n} \rceil$ or $a_i = \lfloor \frac{a}{n} \rfloor$ for $i = 1, 2, \ldots, n$.*

Proof. We prove the statement by contradiction. Suppose that $\sum_{i=1}^{n} f(a_i)$ attains its maximum value and that $a_i - a_j \geq 2$ for some indices i and j. Let $A' = (a'_1, a'_2, \ldots, a'_n)$ be a sequence with $a'_i = a_i - 1$, $a'_j = a_i + 1$ and $a'_k = a_k$ for $k \neq i, j$. Because f is strictly concave (i.e., the first derivative f' is strictly decreasing), we obtain

$$\begin{aligned}
&\sum_{i=1}^{n} f(a_i) - \sum_{i=1}^{n} f(a'_i)\\
=&(f(a_i) - f(a_i - 1)) - \big(f(a_j + 1) - f(a_j)\big)\\
=&f'(\xi_1) - f'(\xi_2)\\
<&0,
\end{aligned}$$

where $\xi_1 \in (a_i - 1, a_i)$ and $\xi_2 \in (a_j, a_j + 1)$. This contradicts the assumption that $\sum_{i=1}^{n} f(a_i)$ attains its maximum value. □

The next result is a direct consequence of Jensen's inequality.

Lemma 5.3. *Let a be a positive integer, and let $A = (a_1, a_2, \ldots, a_n)$ be a positive integer sequence of. If f is strictly concave and $\sum_{i=1}^{n} a_i = a$, then $\sum_{i=1}^{n} f(a_i) \leq nf(\frac{a}{n})$ with equality if and only if $a_i = \frac{a}{n}$ for $i = 1, 2, \ldots, n$.*

Our final result of this section shows that threshold graphs play a key role in this chapter. We use the characterization of Theorem 5.3 (b) in terms of alternating 4-cycles and a lemma from Chapter 2 to prove this result.

Lemma 5.4. *Let n and m be integers with $n \geq 2$ and $1 \leq m \leq \binom{n}{2}$, and let G be an (n, m)-graph. If $I_d(G)$ attains the minimum value among all (n, m)-graphs, then G is a threshold graph.*

Proof. We prove the statement by contradiction. Suppose that G is not a threshold graph and that $I_d(G)$ attains the minimum value among all (n, m)-graphs. By Theorem 5.3 (b), there are four vertices $u, v, w, x \in V(G)$ such that $uv, wx \in E(G)$ and $ux, vw \notin E(G)$. Now assume without loss of generality that $\deg_G(v) \geq \deg_G(x)$. Consider the graph $G' = G - wx + wv$. It is straightforward to check that $D(G') \succ D(G)$. Using Lemma 2.1, we obtain that $I_d(G') < I_d(G)$, a contradiction. □

The remainder of this chapter is devoted to our proof of Theorem 5.1.

5.2 Proof of Theorem 5.1

Proof of Theorem 5.1. We assume that G^* is an (n,m)-graph with $I_d(G^*) = \min\{I_d(G) : G \text{ is an } (n,m)\text{-graph}\}$. Let $D_0, D_1, \ldots, D_s$ be the degree partition of G^*. By Lemma 5.4, G^* is a threshold graph. We want to show that G^* is isomorphic to $K(k^*, t^*) \cup \overline{K}_{n-k^*-\sigma(t^*)}$, which is indeed a threshold graph that is uniquely determined by its degree partition (or degree sequence), which is as follows:

(Ai) if $t^* = 0$ (so if $m = \binom{k^*}{2}$), then $|D_0| = n - k^*$ and $|D_1| = k^*$ (so then $D(K_{k^*} \cup \overline{K}_{n-k^*}) = [(k^*-1)^{k^*}]$);

(Aii) if $1 \le t^* < k^* - 1$ (so if $1 \le m - \binom{k^*}{2} < k^* - 1$), then $|D_0| = n - k^* - 1$, $|D_1| = 1$, $|D_2| = k^* - t^*$ and $|D_3| = t^*$ (so then $D(K(k^*, t^*) \cup \overline{K}_{n-k^*-1}) = [(k^*)^{t^*}, (k^*-1)^{k^*-t^*}, t^*]$);

(Aiii) if $t^* = k^* - 1$ (so if $m - \binom{k^*}{2} = k^* - 1$), then $|D_0| = n - k^* - 1$, $|D_1| = 2$ and $|D_2| = k^* - 1$ (so then $D(K(k^*, k^*-1) \cup \overline{K}_{n-k^*-1}) = [(k^*)^{k^*-1}, (k^*-1)^2]$).

In the remainder, we are going to show that the degree partition $D_0, D_1, \ldots, D_s$ of G^* matches the above degree partition of $K(k^*, t^*) \cup \overline{K}_{n-k^*-\sigma(t^*)}$. Since $|D_0| = n - \sum_{i=1}^{s} |D_i|$, the value of $|D_0|$ can be obtained from the values of all other $|D_i|$. Hence, it is sufficient to determine $|D_i|$ for $i = 1, 2, \ldots, s$. Let $K = \cup_{i=\lceil \frac{s+1}{2} \rceil}^{s} D_i$, and let $k = \omega(G^*)$. By Theorem 5.3 (c), we have $k = \sum_{i=\frac{s+2}{2}}^{s} |D_i| + 1$ (resp., $k = \sum_{i=\frac{s+1}{2}}^{s} |D_i|$) for s even (resp., odd). So we have $|K| = k - 1$ (resp., $|K| = k$) for s even (resp., odd). We continue by proving five claims, the first of which is the following.

Claim 1. *If* $\deg_{\max}(K) - \deg_{\min}(K) \le 1$*, then* $s \le 4$.

Proof. We prove the claim by contradiction. Suppose that $s \ge 5$. Then there are at least three distinct degrees of vertices in K of G^*. This implies $\deg_{\max}(K) - \deg_{\min}(K) \ge 2$, a contradiction. $\diamondsuit$

Before we state the next claim, we first define the unique non-negative integer c satisfying $(c-1)(k-1) < m-\binom{k}{2} \leq c(k-1)$ (i.e., $c = \lceil \frac{m-\binom{k}{2}}{k-1} \rceil$). This integer c appears as follows in the next claim.

Claim 2. *We have* $\deg_{\max}(K) - \deg_{\min}(K) \leq 1$ *if and only if*

(i) $|D_1| = k$ *for* $m - \binom{k}{2} = c(k-1)$ *and* $c = 0$*;*

(ii) $|D_1| = c+1$ *and* $|D_2| = k-1$ *for* $m - \binom{k}{2} = c(k-1)$ *and* $c \geq 1$*;*

(iii) $|D_1| = 1$*,* $|D_2| = \binom{k+1}{2} - m$ *and* $|D_3| = m - \binom{k}{2}$ *for* $(c-1)(k-1) < m - \binom{k}{2} < c(k-1)$ *and* $c = 1$*;*

(iv) $|D_1| = 1$*,* $|D_2| = c$*,* $|D_3| = c(k-1) + \binom{k}{2} - m$ *and* $|D_4| = m - \binom{k}{2} - (c-1)(k-1)$ *for* $(c-1)(k-1) < m - \binom{k}{2} < c(k-1)$ *and* $c \geq 2$*.*

Proof. The if part of the equivalence follows directly from Theorem 5.4, since its statement implies

(i) $\deg_{\max}(K) = \deg_{\min}(K) = d_1 = k-1$ for $m - \binom{k}{2} = c(k-1)$ and $c = 0$;

(ii) $\deg_{\max}(K) = \deg_{\min}(K) = d_2 = c+k-1$ for $m - \binom{k}{2} = c(k-1)$ and $c \geq 1$;

(iii) $\deg_{\max}(K) = d_3 = k$ and $\deg_{\min}(K) = d_2 = k-1$ for $(c-1)(k-1) < m - \binom{k}{2} < c(k-1)$ and $c = 1$;

(iv) $\deg_{\max}(K) = d_4 = c+k-1$ and $\deg_{\min}(K) = d_3 = c+k-2$ for $(c-1)(k-1) < m - \binom{k}{2} < c(k-1)$ and $c \geq 2$.

Thus $\deg_{\max}(K) - \deg_{\min}(K) \leq 1$ in all four cases.

For proving the only-if part of the statement, we assume that $\deg_{\max}(K) - \deg_{\min}(K) \leq 1$. By Claim 1, this implies $s \leq 4$. Now we distinguish four cases, depending on the value of s. In each of these four cases we show that the corresponding statement of the claim holds.

Case 1. $s = 1$.

This implies $K = D_1$ and $|D_1| = k$. By Theorem 5.4, we have $\deg_{\max}(K) = \deg_{\min}(K) = d_1 = k-1$. By Theorem 5.3 (c), every vertex in D_1 is adjacent to all other vertices in D_1, which implies $m = \binom{k}{2}$, that is, $m - \binom{k}{2} = c(k-1)$ and $c = 0$.

Case 2. $s = 2$.

This implies $K = D_2$ and $|D_2| = k-1$. By Theorem 5.4, we have $\deg_{\max}(K) = \deg_{\min}(K) = d_2 = d_1 + |D_1| - 1$ and $d_1 = d_0 + |D_2| = k-1$. Since $d_1|D_1| + d_2|D_2| = 2m$ by the degree sum argument, we have $|D_1| = \frac{m-\binom{k-1}{2}}{k-1} = \frac{m-\binom{k}{2}}{k-1} + 1$. This implies $c = \frac{m-\binom{k}{2}}{k-1}$, or equivalently that $m - \binom{k}{2} = c(k-1)$. By Lemma 5.1, we have $|D_1| \geq 2$, which implies $c \geq 1$.

Case 3. $s = 3$.

This implies $K = D_2 \cup D_3$ and $|K| = k$. This in turn implies $\deg_{\max}(K) = d_3$, $\deg_{\min}(K) = d_2$ and $|D_2| + |D_3| = k$. By Theorem 5.4, we have $d_1 = d_0 + |D_3| = |D_3|$, $d_2 = d_1 + |D_2| - 1$ and $d_3 = d_2 + |D_1|$. If $|D_1| \geq 2$, then $d_3 - d_2 = |D_1| \geq 2$, which contradicts $\deg_{\max}(K) - \deg_{\min}(K) \leq 1$. So we have $|D_1| = 1$. Because $|D_2| + |D_3| = k$, $d_2 = d_1 + |D_2| - 1$ and $d_1 = |D_3|$, we obtain $d_2 = k-1$. Since $d_3 = d_2 + 1$, we also get $d_3 = k$. We use the degree sum argument to deduce that $2m = d_1|D_1| + d_2|D_2| + d_3|D_3| = kd_1 + (k-1)(k-d_1) + d_1$. By straightforward calculations, we obtain $d_1 = m - \binom{k}{2}$. Now it follows from $|D_3| = d_1 = m - \binom{k}{2}$ and $|D_2| + |D_3| = k$ that $|D_2| = k - (m - \binom{k}{2}) = \binom{k+1}{2} - m$. By Lemma 5.1, we have $|D_2| \geq 2$. Because $|D_2| + |D_3| = k$, we have $1 \leq d_1 \leq k-2$. This implies $0 < m - \binom{k}{2} < k-1$, so $(c-1)(k-1) < m - \binom{k}{2} < c(k-1)$ and $c = 1$.

Case 4. $s = 4$.

This implies $K = D_3 \cup D_4$ and $|K| = k-1$. This in turn implies $\deg_{\max}(K) = d_4$, $\deg_{\min}(K) = d_3$ and $|D_3| + |D_4| = k-1$. By Theorem 5.4, we have $d_1 = d_0 + |D_4| = |D_4|$, $d_2 = d_1 + |D_3|$, $d_3 = d_2 + |D_2| - 1$ and $d_4 = d_3 + |D_1|$. Since $|D_3| + |D_4| = k-1$, $d_1 = |D_4|$ and $d_2 = d_1 + |D_3|$, we have $d_2 = k-1$. If $|D_1| \geq 2$, then $d_4 - d_3 = |D_1| \geq 2$, which contradicts $\deg_{\max}(K) - \deg_{\min}(K) \leq 1$. So we have $|D_1| = 1$. Because $d_2 = d_1 + |D_3|$, $d_1 = |D_4|$ and $|D_3| + |D_4| = k-1$, we have $d_2 = k-1$. It follows from $d_3 = d_2 + |D_2| - 1$ and $d_2 = k-1$ that $d_3 = |D_2| + k - 2$. Now the degree sum argument yields $2m = d_1|D_1| + d_2|D_2| + d_3|D_3| + d_4|D_4| = 2d_1 + 2(k-1)|D_2| + (k-1)(k-2)$. By straightforward calculations, we obtain

$|D_2| = \frac{m-\binom{k-1}{2}-d_1}{k-1}$. Because $0 < d_1 < d_2 = k-1$, we have $|D_2| = \lfloor\frac{m-\binom{k-1}{2}}{k-1}\rfloor = \lfloor\frac{m-\binom{k}{2}}{k-1}+1\rfloor = \lceil\frac{m-\binom{k}{2}}{k-1}\rceil$. Thus $c = \lceil\frac{m-\binom{k}{2}}{k-1}\rceil = |D_2| = \frac{m-\binom{k-1}{2}-d_1}{k-1}$. So we have $d_1 = m-\binom{k-1}{2}-c(k-1) = m-(\binom{k-1}{2}+k-1)+k-1-c(k-1) = m-\binom{k}{2}-(c-1)(k-1) = |D_4|$. This implies $|D_3| = k-1-|D_4| = c(k-1)+\binom{k}{2}-m$. By Lemma 5.1, we have $|D_2| \geq 2$, which implies $c \geq 2$. Thus $(c-1)(k-1) < m-\binom{k}{2} < c(k-1)$ and $c \geq 2$. $\diamondsuit$

In the next claim, we will gather more information on the degree partition of G^*. To prepare for the induction proof of this claim, we need the following set up. Let u be a vertex in $D_{\lceil\frac{s}{2}\rceil}$. By Theorem 5.4, we have $\deg_{G^*}(u) = k-1$. Set $G' = G^* - u$. Since G' is an induced subgraph of G^*, by Fact 5.1, we know that G' is also a threshold graph. Let $\omega(G') = k'$. If s is even, then there exists a maximum clique in G^* which does not contain u, since we can use another vertex of $D_{\frac{s}{2}}$ to replace u. So we have $k' = k$ if s is even. If s is odd, then $k' = k-1$, since u must be contained in the maximum clique. Our next claim provides us with more information on the degree partition of G^*.

Claim 3. *The degree partition of G^* satisfies*

(i) $|D_1| = k$ *for* $m-\binom{k}{2} = c(k-1)$ *and* $c = 0$*;*

(ii) $|D_1| = c+1$ *and* $|D_2| = k-1$ *for* $m-\binom{k}{2} = c(k-1)$ *and* $c \geq 1$*;*

(iii) $|D_1| = 1$*,* $|D_2| = \binom{k+1}{2}-m$ *and* $|D_3| = m-\binom{k}{2}$ *for* $(c-1)(k-1) < m-\binom{k}{2} < c(k-1)$ *and* $c = 1$*;*

(iv) $|D_1| = 1$*,* $|D_2| = c$*,* $|D_3| = c(k-1)+\binom{k}{2}-m$ *and* $|D_4| = m-\binom{k}{2}-(c-1)(k-1)$ *for* $(c-1)(k-1) < m-\binom{k}{2} < c(k-1)$ *and* $c \geq 2$*.*

Proof. By induction on m. The statement is trivial for $m = 1$, and easy to check for $m = 2$. In the latter case, $G^* \cong K_{1,2} \cup \overline{K}_{n-3}$, that is, $|D_1| = 2$ and $|D_2| = 1$ (i.e., $m-\binom{k}{2} = k-1$ and $k = 2$). So, we may assume $m \geq 3$ and that the statement holds for all extremal graphs with fewer than m edges.

Adopting the above set up, it follows that G' is a threshold graph with $n' = n-1$ vertices, $m' = m+1-k$ edges and $\omega(G') = k'$. Let $D'_0, D'_1, \ldots, D'_{s'}$ be the degree partition of G'. Let $c' = \lceil\frac{m'-\binom{k'}{2}}{k'-1}\rceil$. We assume that G' satisfies

(i) $|D'_1| = k'$ for $m' - \binom{k'}{2} = c'(k'-1)$ and $c' = 0$;

(ii) $|D'_1| = c'+1$, $|D'_2| = k'-1$ for $m' - \binom{k'}{2} = c'(k'-1)$ and $c' \geq 1$;

(iii) $|D'_1| = 1$, $|D'_2| = \binom{k'+1}{2} - m'$ and $|D'_3| = m' - \binom{k'}{2}$ for $(c'-1)(k'-1) < m' - \binom{k'}{2} < c'(k'-1)$ and $c' = 1$;

(iv) $|D'_1| = 1$, $|D'_2| = c'$, $|D'_3| = c'(k'-1) + \binom{k'}{2} - m'$ and $|D'_4| = m' - \binom{k'}{2} - (c'-1)(k'-1)$ for $(c'-1)(k'-1) < m - \binom{k'}{2} < c'(k'-1)$ and $c' \geq 2$.

To relate G^* to G', we use the following recurrence relation

$$\begin{aligned} h_d(G^*) =& h_d(G') + (k-1)\log(k-1) \\ &+ \sum_{v \in K \setminus \{u\}} (\deg_{G^*}(v)\log(\deg_{G^*}(v)) \\ &- (\deg_{G^*}(v)-1)\log(\deg_{G^*}(v)-1)). \end{aligned}$$

Let $f(x) = x\log(x) - (x-1)\log(x-1)$ for $x \geq 2$. It follows that $f(x)$ is strictly concave. Since $\sum_{v \in K \setminus \{u\}} \deg_{G^*}(v) = m + \binom{k-1}{2}$ and $|K \setminus \{u\}| = k-1$, by Lemma 5.2, $\sum_{v \in K \setminus \{u\}} (\deg_{G^*}(v)\log(\deg_{G^*}(v)) - (\deg_{G^*}(v)-1)\log(\deg_{G^*}(v)-1))$ attains the maximum value if and only if the maximum degree exceeds the minimum degree by at most 1, that is, $\deg_{\max}(K) - \deg_{\min}(K) \leq 1$. The statement of the claim now follows from Claim 1 and the induction hypothesis. $\diamondsuit$

We next prove that $k = k^*$. We first make some observations, and we define an auxiliary graph and a help function to facilitate our proof. By Claim 3, for $s = 1$, we have $|D_1| = k^*$ in which $\binom{k^*}{2} = m$; for $s = 3$, we have $|D_0| = n - k^* - 1$, $|D_1| = 1$, $|D_2| = k^* - t^*$ and $|D_3| = t^*$ in which $1 \leq m - \binom{k^*}{2} < k^* - 1$. By Claim 3, if $m - \binom{k}{2} = c(k-1)$, then the degree sequence of G^* is $[(c+k-1)^{k-1}, (k-1)^{c+1}]$ (i.e., $\deg_{\max}(K) - \deg_{\min}(K) = 0$). Let $G_{n,m,k}$ be the (n,m)-graph satisfying $D(G_{n,m,k}) = [(c+k-1)^{k-1}, (k-1)^{c+1}]$ in which $c = \frac{m - \binom{k}{2}}{k-1}$. Notice that there are $n - c - k$ isolated vertices in $G_{n,m,k}$. By Theorem 5.2, this threshold sequence has a unique labeled realization. Therefore $G_{n,m,k} \cong G^*$ if $m - \binom{k}{2} = c(k-1)$. Let the function $g(m,k) = (k-1)(c+k-1)\log(c+k-1) + (c+1)(k-1)\log(k-1)$, in which $c = \frac{m-\binom{k}{2}}{k-1}$. It follows that $h_d(G_{n,m,k}) = g(m,k)$. Now we are ready to prove the final two claims.

Claim 4. *We have $h_d(G^*) \le g(m,k)$, with equality if and only if $G^* \cong G_{n,m,k}$.*

Proof. Because $\sum_{v \in K\setminus\{u\}} \deg_{G^*}(v) = m + \binom{k-1}{2}$, $|K\setminus\{u\}| = k-1$ and $f(x+1) = (x+1)\log(x+1) - x\log(x)$ is strictly concave, by Lemma 5.3, we have

$$\begin{aligned}&\sum_{v \in K\setminus\{u\}} (\deg_{G^*}(v)\log(\deg_{G^*}(v)) - (\deg_{G^*}(v)-1)\log(\deg_{G^*}(v)-1))\\ \le&(k-1)\left(\left(\frac{m+\binom{k}{2}}{k-1}\right)\log\left(\frac{m+\binom{k}{2}}{k-1}\right) + \left(\frac{m+\binom{k}{2}}{k-1}-1\right)\log\left(\frac{m+\binom{k}{2}}{k-1}-1\right)\right),\end{aligned}$$

with equality if and only if $\deg_{G^*}(v) = \frac{m+\binom{k-1}{2}}{k-1}$. For $m = 2$, $G^* \cong K_{1,2} \cup \overline{K}_{n-3}$ and $h_d(G^*) = 2 = g(2,2)$. By the same set up and similar induction arguments as in the proof of Claim 3, we assume $h_d(G') \le g(m',k')$. Using these arguments, we conclude that $h_d(G^*) \le g(m,k)$, with equality if and only if $G^* \cong G_{n,m,k}$. $\diamondsuit$

Next we prove that $g(m,k)$ is strictly increasing in k.

Claim 5. *We have $h_d(G_{n,m,k}) \le h_d(G_{n,m,k^*})$, with equality if and only if $k = k^*$.*

Proof. By straightforward calculations, we have

$$\begin{aligned}h_d(G_{n,m,k}) &= g(m,k)\\ &= \left(m + \binom{k-1}{2}\right)\log\left(\frac{m+\binom{k-1}{2}}{k-1}\right) + \left(m - \binom{k-1}{2}\right)\log(k-1).\end{aligned}$$

For $k \ge 2$ and $m \ge \binom{k}{2}$, by taking the derivative and after some reshuffling, we obtain

$$\frac{\partial g(m,k)}{\partial k} = \left(\frac{2k-3}{2}\right)\log(\frac{(k-1)(k-2)+2m}{2(k-1)^2}) + \frac{\log(e)}{2} > 0,$$

in which e is the natural constant. So, we conclude that $h_d(G_{n,m,k})$ is strictly increasing in k. Since $k \leq k^*$, this implies $h_d(G_{n,m,k}) = g(m,k) \leq h_d(G_{n,m,k^*})$, with equality if and only if $k = k^*$. $\diamondsuit$

We now finish the proof by going through the four cases of Claim 3, and showing that in each case the degree partition of G^* indeed matches the statements in (Ai)–(Aiii) of the beginning of this proof. Since $\omega(G^*) \leq k^*$, we get $0 \leq t^* = m - \binom{k^*}{2} \leq k^* - 1$. In particular, this shows that Case (iv) of Claim 3 cannot occur. Note that Case (i) and (iii) correspond to the statements in (Ai) and (Aii), respectively. We are done for $t^* = 0$ (i.e., for the statement in (Ai)) and $0 < t^* < k^* - 1$ (i.e., for the statement in (Aii)). So we consider $t^* = m - \binom{k^*}{2} = k^* - 1$ (i.e., $\frac{m - \binom{k^*}{2}}{k^* - 1} = 1$) in the remainder of this proof. By Claims 4 and 5, we have $h_d(G^*) \leq g(m, k^*)$, with equality if and only if $G^* \cong G_{n,m,k^*}$. Recall that by definition, $D(G_{n,m,k}) = [(c+k-1)^{k-1}, (k-1)^{c+1}]$ in which $c = \frac{m - \binom{k}{2}}{k-1}$. This implies $D(G_{n,m,k^*}) = [(c^* + k^* - 1)^{k^*-1}, (k^* - 1)^{c^*+1}]$ in which $c^* = \frac{m - \binom{k^*}{2}}{k^* - 1} = 1$. Substituting $c^* = 1$, we directly obtain $D(G_{n,m,k^*}) = [(k^*)^{k^*-1}, (k^* - 1)^2]$. Thus $|D_1| = 2$ and $|D_2| = k^* - 1$ for $t^* = k^* - 1$. This matches the statement in (Aiii), and completes the proof. $\square$

Chapter 6

Extremal problems on two distance-based entropies

In this chapter, we mainly study the behavior of two distance-based entropies. We first recall the definitions of the Wiener-entropy and the eccentricity-entropy. By deriving their (asymptotic) extremal behavior, we conclude that the values of the Wiener-entropy of graphs of a given order are more spread than the values of the eccentricity-entropy. We resolve three known conjectures on the eccentricity-entropy. We propose two new conjectures (in one statement) on the Wiener-entropy, based on some surprising observed behavior of the graphs minimizing it.

6.1 Introduction

In this section, we repeat some of the definitions related to distance-based entropies, for convenience. In Subsection 6.1.1, we explain our notation and give an overview of the main distance-based graph entropies. This is followed by Subsection 6.1.2, in which we present an elementary result, showing that the upper bound of many graph entropies is $\log(n)$. As a consequence, we also show that the extremal graphs attaining this upper bound exhibit some kind of regularity. Finally, an overview of our contributions is summarized in

Subsection 6.1.3 of this section.

6.1.1 Distance-based graph entropies

All graphs considered in this chapter are connected. Before introducing the main concepts, we need to recall and introduce some related terminology and notation.

Let $G=(V,E)$ be a graph. We recall that the distance between two vertices u and v is denoted by $d(u,v)$. In particular, it satisfies the *triangle inequality* $d(u,v)\leq d(u,w)+d(w,v)$ for all vertices $u,v,w\in V$. Recall that by $\mathrm{diam}(G)$ we denote the diameter of G. As before, the eccentricity of a vertex v is denoted by $\mathrm{ecc}(v)$. A *central vertex* of G is a vertex with the minimum eccentricity. The *radius*, denoted by $\mathrm{rad}(G)$, is the minimum eccentricity among all vertices of G. Recall that $S_j(v,G)=\{u\in V : d(u,v)=j\}$ is called the j-sphere of $v\in V$. The *transmission* of a vertex v is denoted by $\sigma_G(v)$, or $\sigma(v)$ if the graph G is clear, and equals the sum of distances towards all other vertices: $\sigma(v)=\sum_{u\in V}d(v,u)$.

Let us recall some graph entropies which will be studied in this chapter. Let $G=(V,E)$ be a graph with vertex set $\{v_1,v_2,\ldots,v_n\}$. Dehmer [43] defined a general form of graph entropy for G using an information functional $f(v_i)$ by the formula

$$I_f(G)=-\sum_{i=1}^{n}\frac{f(v_i)}{\sum_{j=1}^{n}f(v_j)}\log\left(\frac{f(v_i)}{\sum_{j=1}^{n}f(v_j)}\right).$$

The entropy $I_{s_2}(G)$ of G was defined in [47, 84] by the formula

$$I_{s_2}(G)=-\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{\mathrm{diam}(G)}c_j|S_j(v_i,G)|}{\sum_{t=1}^{n}\sum_{j=1}^{\mathrm{diam}(G)}c_j|S_j(v_t,G)|}\right)\log\left(\frac{\sum_{j=1}^{\mathrm{diam}(G)}c_j|S_j(v_i,G)|}{\sum_{t=1}^{n}\sum_{j=1}^{\mathrm{diam}(G)}c_j|S_j(v_t,G)|}\right),$$

for some fixed real numbers $c_j>0$ with $j=1,2,\ldots,\mathrm{diam}(G)$.

When $c_j = j$ for every $1 \le j \le \operatorname{diam}(G)$, we define it as the Wiener-entropy by the formula

$$I_w(G) = -\sum_{i=1}^{n} \frac{\sigma(v_i)}{\sum_{j=1}^{n} \sigma(v_j)} \log\left(\frac{\sigma(v_i)}{\sum_{j=1}^{n} \sigma(v_j)}\right),$$

where $\sigma(v_i)$ is the transmission of v_i. The Wiener index of G, proposed by Wiener [122], is defined by

$$W(G) = \sum_{\{v_i, v_j\} \subseteq V} d(v_i, v_j).$$

Since $\sum_{j=1}^{n} \sigma(v_j) = 2W(G)$, we have

$$I_w(G) = \log(2W(G)) - \frac{1}{2W(G)} \sum_{i=1}^{n} \sigma(v_i) \log(\sigma(v_i)).$$

In the literature [47, 84], the following entropy regarding the eccentricity of G was defined by

$$I_e(G) = -\sum_{i=1}^{n} \frac{c_i \operatorname{ecc}(v_i)}{\sum_{j=1}^{n} c_j \operatorname{ecc}(v_j)} \log\left(\frac{c_i \operatorname{ecc}(v_i)}{\sum_{j=1}^{n} c_j \operatorname{ecc}(v_j)}\right),$$

for some fixed real numbers $c_i > 0$ and $i = 1, 2, \ldots, n$. For $c_i = 1$, the eccentricity-entropy was defined in [47, 84] by the formula

$$I_{\mathrm{ecc}}(G) = -\sum_{i=1}^{n} \frac{\operatorname{ecc}(v_i)}{\sum_{j=1}^{n} \operatorname{ecc}(v_j)} \log\left(\frac{\operatorname{ecc}(v_i)}{\sum_{j=1}^{n} \operatorname{ecc}(v_j)}\right).$$

6.1.2 Maximum graph entropies

We state the following elementary result for clarity.

Theorem 6.1. *Let $G = (V, E)$ be a graph, and let $f : V \to \mathbb{R}^+$ be an information functional. Then $I_f(G) \le \log(n)$, with equality if and only if G is f-regular, in the sense that $f(v)$ is a constant for every $v \in V$.*

Proof. In general, the Shannon entropy of a discrete variable X with set of possible outcomes $\{x_1, x_2, \ldots, x_n\}$ is upper bounded by $\log(n)$, with equality if and only if X has a uniform distribution. This is a well-known fact and consequence of Jensen's inequality (since $f(x) = -x \log x$ is concave on $[0, 1]$). As a corollary, this implies that $I_f(G) \le \log(n)$, with equality if and only if $f(v)$ is a constant for every $v \in V$. $\square$

As a corollary, we obtain the following statements for the main examples of the degree-entropy I_d (see the equation (2.1)) and distance-based graph entropies I_w and I_{ecc}. Before giving statements, we need some definitions of some classes of graphs. A graph is *transmission-regular* if $\sigma(v) = \sigma(u)$ for every $u, v \in V$. Such graphs have been studied in the past, also under the name distance-balanced graphs, see e.g., [1, 12, 68]. If all the vertices in G have the same eccentricity, then G is a *self-centered* graph [27].

Corollary 6.1. *For a graph G of order n,*

- $I_d(G) \le \log(n)$*, with equality if and only if G is a regular graph,*
- $I_w(G) \le \log(n)$, *with equality if and only if G is a transmission-regular graph,*
- $I_{\text{ecc}}(G) \le \log(n)$, *with equality if and only if G is a self-centered graph.*

If one is restricting the class to trees, determining the maximum given the order is not trivial anymore. For I_d, it is known that the path is extremal [29, Thm. 1]. In Section 6.3, we prove that the trees maximizing I_{ecc} among all trees of order $n \ge 4$ are precisely the trees with diameter 3. For I_w, we claim that the extremal tree is again a different one, being the star, whenever $n \ge 5$.

Conjecture 6.1. *Let T be an arbitrary tree of order $n \ge 5$. Then $I_w(T) \le I_w(K_{1,n-1})$ with equality if and only if $T \cong K_{1,n-1}$.*

6.1.3 Contributions

When one is considering graphs, eccentricity and transmission are the local analogues of diameter and total distance (linearly related with average distance), the two main distance measures for graphs. As such, among distance-based entropies, the eccentricity-entropy and the Wiener-entropy seem to be among the most natural ones (besides the more general versions). In this chapter, we focus on these two.

Before that, in Section 6.2, we collect and prove some general results and observations about the Shannon entropy of normalized sequences (the probability distribution linearly related with the sequence). These general results are useful to apply and give intuition on questions for graph entropies, but might also be handy when working on general problems about (Shannon) entropy.

In Section 6.3, we focus on the eccentricity-entropy and resolve the following three conjectures.

Conjecture 6.2 ([84]). *Among graphs of order n, the minimum value of I_{ecc} is attained by the graph obtained by removing a small number of edges from the complete graph of order n. In particular, the extremal graphs of order n will have $k \ge \frac{n}{2}$ vertices of degree $n-1$.*

Conjecture 6.3 ([47]). *Among trees of order n, the maximum value of I_{ecc} is attained by the tree obtained by attaching $n-3$ vertices to a pendant vertex of the path of length 2.*

Conjecture 6.4 ([47]). *Among trees of order n and diameter $d \ll n$, the maximum value of I_{s_2} and I_e are attained by the tree obtained by identifying the central vertex of the star of order $n-d$ with a central vertex of the path of length d.*

Below is a brief statement of the relationship between our results and the above conjectures.

- Conjecture 6.2([84, Conj. 6.2]) is proven to be true (see Theorem 6.3), except from the statement about the removal of a few edges (see Remark 6.2).
- Conjecture 6.3 ([47, Conj. 4.6]) is true for I_{ecc} but false for general I_e, since one can choose the c_i in such a way that ones favorite tree is extremal (see Proposition 6.7).

- Conjecture 6.4 ([47, Conj. 4.3]) is refuted (see Proposition 6.8 and Corollary 6.2).

Our proofs mainly rely on the general results in Section 6.2. Intuitively, the eccentricity sequence has to be as unbalanced or balanced as possible to attain the minimum or maximum value for the entropy.

In Section 6.4, we prove that the minimum of $I_w(G)$ among all graphs of order n is of the form $\left(\frac{3}{4}+o(1)\right)\log(n)$. Since the minimum of $I_{\text{ecc}}(G)$ is $(1-o(1))\log(n)$, the Wiener-entropy has a better distinguishing character, i.e., the difference between the maximum and minimum is larger.

Finally, in Section 6.5 we give some remarks on the trees and graphs that conjecturally attain the minimum Wiener-entropy. For this, we define the graph $G_{n,k,j}$ formally as follows. Take the disjoint union of a path P_k and clique K_{n-k}, and join one pendant vertex of the path with j vertices of the clique by edges. An example of such a graph $G_{n,k,j}$ has been presented in Figure 6.1.

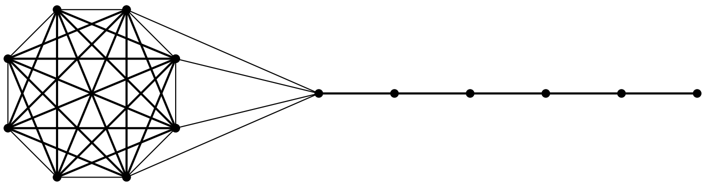


Figure 6.1: The graph $G_{n,k,j}$ for $n=14, k=6$ and $j=4$.

When we consider the value of j that minimizes $I_{\text{ecc}}(G_{n,k,j})$, there is some surprising behavior. When n ranges from 16 to 100, the value of j fluctuates a lot, which intuitively can be explained as the clique at the end growing in a less discretized way. Nevertheless, for n large, it seems that $j=1$ is always true, i.e., we conjecture that I_w for graphs of order n is maximized by a graph of the form $G_{n,k,1}$ whenever n is sufficiently large. Explaining this evolution in behavior of the graphs $G_{n,k,j}$ seems to be already interesting.

6.2 General results on entropy

The *normalized vector* (or *unit vector*) of a non-zero vector $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, is a vector in the same direction with norm 1. It is denoted by $\widehat{\mathbf{a}}$ and given by $\widehat{\mathbf{a}} = \frac{\mathbf{a}}{||\mathbf{a}||_1}$, where $||\mathbf{a}||_1 = \sum_{i=1}^{n} |a_i|$ is the 1-norm of $\mathbf{a}$.

Definition 6.1. *Let* $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ *be a positive real vector, and let* $\mathbf{p} = \widehat{\mathbf{a}} = (p_1, p_2, \ldots, p_n)$*. We define the entropy of* $\mathbf{a}$ *by*

$$\begin{aligned} H(\mathbf{a}) &= -(p_1, p_2, \ldots, p_n)(\log(p_1), \log(p_2), \ldots, \log(p_n))^{\mathrm{T}} \\ &= -\sum_{j=1}^{n} p_j \log(p_j). \end{aligned}$$

The following theorem is a direct consequence of Karamata's inequality [75] applied on the (strictly) concave function $-x\log(x)$ (for $x \in [0,1]$). Here we omit the definition of majorization that was introduced in Subsection 2.1.2.

Theorem 6.2. *Let* $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ *and* $\mathbf{b} = (b_1,\ b_2, \ldots,\ b_n)$ *be two different sequences of positive reals such that* $\mathbf{a}$ *majorizes* $\mathbf{b}$*. Then*

$$H(\mathbf{a}) < H(\mathbf{b}).$$

Proposition 6.1. *Let* $S \subset (\mathbb{R}^+)^n$ *be the convex hull of* N *points. If* $H(\mathbf{s})$ *attains the minimum value among* $\mathbf{s} \in S$*, then* $\mathbf{s}$ *is one of the extremal points of the convex hull.*

Proof. Assume this is not the case. So $\mathbf{s}$ is not one of the extremal points of the convex hull. In that case there is a non-zero vector $\mathbf{t}$ for which the interval $[\mathbf{s}-\mathbf{t}, \mathbf{s}+\mathbf{t}]$ is fully contained in S. Since $(\mathbf{s}-\mathbf{t}) + (\mathbf{s}+\mathbf{t}) = 2\mathbf{s}$, we have $||\mathbf{s}-\mathbf{t}||_1 \widehat{\mathbf{s}-\mathbf{t}} + ||\mathbf{s}+\mathbf{t}||_1 \widehat{\mathbf{s}+\mathbf{t}} = 2||\mathbf{s}||_1 \widehat{\mathbf{s}}$. So $\widehat{\mathbf{s}} = \lambda \widehat{\mathbf{s}-\mathbf{t}} + (1-\lambda)\widehat{\mathbf{s}+\mathbf{t}}$, where $0 < \lambda < 1$.

Since the entropy is strictly concave, this immediately implies that

$$\min\{H(\mathbf{s}-\mathbf{t}), H(\mathbf{s}+\mathbf{t})\} < H(\mathbf{s}).$$

This leads to a contradiction with the choice of $\mathbf{s}$. $\square$

Proposition 6.2. *Let $a_1, a_2, \dots, a_s$ be fixed positive reals and $b_1, b_2, \dots, b_t$ be positive variables. Let β be the solution for $\beta = \exp\left(\frac{\sum_{j=1}^s a_j \log(a_j)}{\sum_{j=1}^s a_j \log(e)}\right)$. Then the entropy of (the normalized vector of) $(a_1, a_2, \dots, a_s, b_1, \dots, b_t)$ is maximized if and only if $b_1 = b_2 = \dots = b_t = \beta$. Furthermore, $H\left((a_1, a_2, \dots, a_s, \underbrace{b, \dots, b}_{t})\right)$ is increasing when $b < \beta$ and decreasing when $b > \beta$.*

Proof. First note that due to concavity of $f(x) = -x\log(x)$, once $\sum b_i$ is fixed, we know that the maximum occurs when all b_i are equal to a value b. In that case

$$H((a_1, a_2, \dots, a_s, b_1, \dots, b_t)) = \log\left(\sum a_i + bt\right) - \frac{\sum a_i \log(a_i) + tb\log(b)}{\sum a_i + bt}.$$

The maximum can be found by taking the derivative towards b and setting this equal to zero:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}b} H\left((a_1, a_2, \dots, a_s, \underbrace{b, \dots, b}_{t})\right) =& \frac{t}{\sum a_i + bt} - \frac{t(1+\log(b))}{\sum a_i + bt} \\ &+ t\frac{\sum a_i \log(a_i) + tb\log(b)}{(\sum a_i + bt)^2} \\ =& \frac{\sum a_i \log(a_i) - \log(b)\sum a_i}{(\sum a_i + bt)^2}.\end{aligned}$$

So this is zero when $\log(b) = \frac{\sum_{j=1}^s a_j \log(a_j)}{\sum_{j=1}^s a_j}$, i.e., $b = \beta$, and it is positive respectively negative if b is smaller or larger, implying it indeed attains a maximum for this choice. □

Proposition 6.3. *Let $a_1, a_2, \dots, a_s > 1$ be fixed reals and $b_1, b_2, \dots, b_t$ be variables that are positive integers. Let β be the real solution for the equality $\beta = \exp\left(\frac{\sum_{j=1}^s a_j \log(a_j)}{\sum_{j=1}^s a_j \log(e)}\right)$. Then the entropy of $(a_1, a_2, \dots, a_s, b_1, \dots, b_t)$ is maximized for $b_1 = b_2 = \dots = b_t = b$, where b is the optimal choice in $\{\lfloor \beta \rfloor, \lceil \beta \rceil\}$.*

Proof. Let $(b_i)_{1\le i\le t}$ be positive integers maximizing $H((a_1,a_2,\dots,a_s,b_1,\dots,b_t))$. As a first step, we prove that all b_i are equal to $\lceil\beta\rfloor$. Assume to the contrary that $b_1\le b_2\le\cdots\le b_t$ (i.e., they are ordered) and $b_{u-1}<b_u=b_{u+1}=\cdots=b_t>\lceil\beta\rceil$ (the case where some numbers are smaller than $\lfloor\beta\rfloor$ is analogous). Now applying Proposition 6.2 with fixed reals $(a_1,a_2,\dots,a_s,b_1,\dots,b_{u-1})$ and variables $(t-u+1)$ times a variable b, we conclude that $H\big((a_1,a_2,\dots,a_s,b_1,\dots,b_{u-1},\underbrace{b,\dots,b}_{t-u+1})\big)$ is decreasing when $b>\beta'$. Here $\beta'\le\max\{\beta,b_{u-1}\}\le b_u-1$. Hence decreasing $b_u=b_{u+1}=\cdots=b_t$ by one increases the entropy. This contradiction implies the result.
Alternatively, one can repeat the above argument of increasing the entropy by decreasing all occurrences of the largest value by one if it is larger than $\lceil\beta\rceil$. This implies that $\max\{b_i\}\le\lceil\beta\rceil$ for an optimal choice of $(b_i)_{1\le i\le t}$. Analogously an optimal sequence satisfies $\min\{b_i\}\ge\lfloor\beta\rfloor$.

In the second step, we prove that all b_i are equal. Assume this is not the case and t_1 of the b_i equal $b_1=\lfloor\beta\rfloor$ and $t_2=t-t_1$ equal (possible after renaming the index) $\lceil\beta\rceil=b_2$. Consider $(a_1,a_2,\dots,a_s,b_3,\dots,b_t)$ as fixed values and b_1 and b_2 as the variable ones. Now there do exist positive reals c_1,c_2 for which

$$c_1b_1+c_2b_2=\sum_{i=1}^{s}a_i+\sum_{i=3}^{t}b_i$$
$$c_1b_1\log(b_1)+c_2b_2\log(b_2)=\sum_{i=1}^{s}a_i\log(a_i)+\sum_{i=3}^{t}b_i\log(b_i).$$

We will need the following claim in the remaining of the proof.

Claim 1. *For fixed positive reals n and $b\ge1$, consider the function*

$$h(c)=\log(nb+c)-\frac{c(b+1)\log(b+1)+(n-c)b\log(b)}{nb+c}.$$

Then $h(c)$ is a strictly convex function on $[0,n]$.

Proof. Since $h(c)$ is a function that is twice continuously differentiable on $[0,n]$, to prove $h(c)$ is strictly convex, it is sufficient to prove its second derivative is positive for all c in $[0, n]$. By calculating, its second derivative is $\frac{2b(b+1)n(\ln(b+1)-\ln(b))-bn-c}{(bn+c)^3\ln(2)}$. This second derivative is positive since $bn+c \leq (b+1)n$ and $2b(\ln(b+1)-\ln(b)) > 1$ for every $b \geq 1$. $\diamondsuit$

We apply Claim 1 with fixed $b = \lfloor \beta \rfloor$ and $n = c_1 + c_2 + 2$. Let $c = c_2 + 1$. Then due to convexity of h, $h(c) < \max\{h(c-1), h(c+1)\}$. But this is exactly telling that changing b_1 into $\lceil \beta \rceil$ or b_2 into $\lfloor \beta \rfloor$ will increase the entropy $H((a_1, a_2, \dots, a_s, b_1, \dots, b_t))$. This contradicts the choice of the sequence having the maximum entropy and thus we can conclude that all b_i are equal. $\square$

Definition 6.2. *Let* $H^n((a_1, a_2, \dots, a_s)) = H\left((a_1, a_2, \dots, a_s, \underbrace{b, \dots, b}_{n-s})\right)$ *where* b *is chosen such that* $b = \exp\left(\frac{\sum_{j=1}^s a_j \log(a_j)}{\sum_{j=1}^s a_j \log(e)}\right)$.

Lemma 6.1. $H^n((a_1, a_2, \dots, a_s)) = \log(n-r)$ *where* $r = s - \frac{\sum_{j=1}^s a_j}{b}$.

Proof. Note that

$$H\left((a_1, a_2, \dots, a_s, \underbrace{b, \dots, b}_{n-s})\right) = \log\left(\sum_{j=1}^s a_j + (n-s)b\right) - \frac{\sum_{j=1}^s a_j \log(a_j) + (n-s)b\log(b)}{\sum_{j=1}^s a_j + (n-s)b}$$
$$= \log\left(\sum_{j=1}^s a_j + (n-s)b\right) - \log(b)$$
$$= \log\left(n - s + \frac{\sum_{j=1}^s a_j}{b}\right). \qquad \square$$

Proposition 6.4. *Let* $\mathbf{a} = (a_1, a_2, \dots, a_s)$ *and* $\mathbf{c} = (c_1, c_2, \dots, c_s)$ *be two sequences of positive reals such that* $\mathbf{a}$ *majorizes* $\mathbf{c}$. *Then for every* $n \geq s$,

$$H^n(\mathbf{a}) \leq H^n(\mathbf{c}).$$

Proof. Let $\log(b) = \frac{\sum_{j=1}^{s} a_j \log(a_j)}{\sum_{j=1}^{s} a_j}$ and $\log(b') = \frac{\sum_{j=1}^{s} c_j \log(c_j)}{\sum_{j=1}^{s} c_j}$. By Theorem 6.2 and Proposition 6.2 (with b' as maximizer) respectively,

$$
\begin{aligned}
&H\left((a_1, a_2, \ldots, a_s, \underbrace{b, \ldots, b}_{n-s})\right) \\
\leq &H\left((c_1, c_2, \ldots, c_s, \underbrace{b, \ldots, b}_{n-s})\right) \\
\leq &H\left((c_1, c_2, \ldots, c_s, \underbrace{b', \ldots, b'}_{n-s})\right).
\end{aligned}
$$

□

Remark 6.1. *Observe that by definition, if* $\mathbf{c}$ *is a subsequence of* $\mathbf{a}$, *then* $H^n(\mathbf{a}) \leq H^n(\mathbf{c})$.

6.3 On eccentricity-entropy

This section is devoted to the three conjectures mentioned in Section 6.1. We start by proving Conjecture 6.2.

Theorem 6.3. *The minimum value for* I_{ecc} *among all graphs of order* n *is attained by graphs of diameter* 2.

Proof. Let G be a graph attaining the minimum value for $I_{\mathrm{ecc}}(G)$ among all graphs of order n. Let r be the radius of the graph and let $S = [r, 2r]^n$. The sequence of eccentricities $(\mathrm{ecc}(v))_{v \in V}$ belongs to S and by Proposition 6.1 every eccentricity is r or $2r$ if it attains the minimum in S. If all eccentricities are equal, then the entropy equals $\log(n)$ and so it would be the maximum instead of minimum. If $r > 1$, then if there are vertices with eccentricity r, as well as with $2r$, there is also a vertex with eccentricity $2r - 1$. When $r = 1$, equality can clearly be attained and since the normalization made the precise value of r being unimportant, we conclude that the graphs with radius 1 indeed attain the minimum. □

Remark 6.2. *A vertex has eccentricity* 1 *if and only if it has degree* $n-1$. *The entropy is completely determined once the number* k *of vertices of degree* $n-1$ *is* *known. The other $n-k$ vertices have eccentricity* 2*. The entropy of the graph equals* $\log(2n-k)-\frac{2(n-k)}{2n-k}$*. This is a concave function for* $0 \le k \le n$*, with the minimum being attained by* $k=(2-2\ln(2))n$ *(over the reals, and so* k *will be of the form* $\lceil(2-2\ln(2))n\rfloor$*). This minimum is roughly* $\log(n)-0.086=(1-o(1))\log(n)$*. As such, for* $n \ge 10$*, we immediately have that* $k > \frac{n}{2}$*, so together with the verification for small* n*, this addresses Conjecture 6.2 completely. Since there is a set* S *of* $s=\lceil(2\ln(2)-1)n\rfloor$ *vertices with eccentricity* 2*, the complement of* $G[S]$ *needs to have degree at least* 1*. In particular, taking* $G[S]$ *to be the empty graph, we observe that there are extremal graphs with at least* $\binom{\lfloor(2\ln(2)-1)n\rfloor}{2}$ *pairs of non-adjacent vertices. It also implies that there are many minimal graphs,* $2^{\Theta(n^2)}$ *(roughly the number of non-isomorphic connected graphs of order* s*), contrasting the ideas of [84].*

Then we turn attention to Conjecture 6.3. When restricting to the class of trees, we will observe that the star (the only tree with diameter 2) is not the graph minimizing the eccentricity-entropy, mainly due to the reason that there is no possibility to play with the ratio of eccentricities with values 1 and 2. Actually the trees minimizing I_{ecc} will be caterpillars with many pendant vertices attached to both the central vertices and pendant vertices of a path, such that most eccentricities are $r+1$ and $2r$, for some value of $r(n)$. On the other hand, the star has the third largest possible value for I_{ecc} among all trees of order n. This will be verified by deriving the three largest possible values of the eccentricity-entropy for the class of trees, and as such we also confirm Conjecture 6.3. To do so, we first observe that the eccentricities of vertices on a diametrical path (a path between two furthest apart vertices) are fixed.

Remark 6.3. *Let* P *be a diametrical path of length* d *in a tree. If* v *is a vertex on the path at distance* i *from one end vertex of* P*, then* $\operatorname{ecc}(v)=\max\{i,d-i\}$*. Otherwise, there exists a path of length larger than* d*.*

Lemma 6.2. *For a star* S_n*,* $I_{\text{ecc}}(S_n) > \log\left(n-\frac{1-\ln(2)}{2}\right)$*.*

Proof. Observe that $\frac{d}{dx}\log(x)=\frac{1}{\ln(2)x}$, and thus

$$I_{\text{ecc}}(S_n)=\log\left(n-\frac{1}{2}\right)+\frac{\frac{1}{2}}{n-\frac{1}{2}}$$

$$\begin{aligned} &\geq \log\left(n-\frac{1}{2}\right)+\frac{1}{\ln(2)}\int_{n-\frac{1}{2}}^{n-\frac{1-\ln(2)}{2}}\frac{1}{x}\,\mathrm{d}x \\ &= \log\left(n-\frac{1-\ln(2)}{2}\right) \end{aligned}$$ □

Proposition 6.5. *If T is a graph of order n and diameter $d \geq 6$, then $I_{\mathrm{ecc}}(T) < I_{\mathrm{ecc}}(S_n)$.*

Proof. By combining Remark 6.1, Proposition 6.4 and Lemma 6.1, we conclude that for even diameter $d = 2r$ where $r \geq 3$ (r is the radius) and $n \geq 2r+1$,

$$\begin{aligned} & H^n((r, r+1, r+1, r+2, r+2, \ldots, 2r-1, 2r-1, 2r, 2r)) \\ \leq & H^n((r, r+1, r+1, 2r-1, 2r-1, 2r, 2r)) \\ \leq & H^n\left((r, \frac{4}{3}r, \frac{4}{3}r, \frac{5}{3}r, \frac{5}{3}r, 2r, 2r)\right) \\ = & H^n((3,4,4,5,5,6,6)) \\ < & \log(n-0.17). \end{aligned}$$

For the second inequality, it is sufficient to note that $r+1 < \frac{4}{3}r < \frac{5}{3}r < 2r-1$ when $r \geq 4$. Analogously, for $d = 2r-1$ odd and $n \geq 2d$ we have

$$\begin{aligned} & H^n((r, r, r+1, r+1, 2r-2, 2r-2, 2r-1, 2r-1)) \\ \leq & H^n((4,4,5,5,6,6,7,7)) \\ < & \log(n-0.16). \end{aligned}$$

Since $\frac{1-\ln(2)}{2} < 0.154$, we conclude by Remark 6.3. □

Proposition 6.6. *If a tree T of order n has diameter 5 and $I_{\mathrm{ecc}}(T) < I_{\mathrm{ecc}}(S_n)$, then it has at most 2 vertices with eccentricity 5. Every tree T of order n with diameter 4 satisfies $I_{\mathrm{ecc}}(T) < I_{\mathrm{ecc}}(S_n)$.*

Proof. For diameter 5, by Lemma 6.2 it is sufficient to compute with Lemma 6.1 that $H^n((3,3,4,4,5,5,5,5)) < \log(n-0.157)$ for every $n \geq 8$, and

$H^n\left((3,3,\underbrace{4,\dots,4}_{40},5,5,5)\right) < \log(n-0.15345)$ for every $n \geq 45$, and verify that $H\left((3,3,\underbrace{4,\dots,4}_{n-5},5,5,5)\right) < I_{\mathrm{ecc}}(S_n)$ for every $n \leq 44$.

For diameter 4, similarly it is sufficient to compute that $H^n((2,3,3,4,4,4)) < \log(n-0.18)$,
$H^n\left((2,\underbrace{3,\dots,3}_{10},4,4)\right) < \log(n-0.154)$ (for n sufficiently large) and
$H\left((2,\underbrace{3,\dots,3}_{n-3},4,4)\right) < I_{\mathrm{ecc}}(S_n)$ for every $n \leq 12$. □

By Propositions 6.5 and 6.6 there are only 3 candidates of extremal graphs.

The next result determines the first three maximum values of eccentricity-entropy for trees of a given order by their corresponding trees or eccentricity sequences.

Proposition 6.7. *Among all trees of order n, the three largest possible values of I_{ecc} are obtained (in order) by a tree $T_{3,n}$ of diameter* 3*, a tree $T_{5,n}$ with eccentricity sequence* $(3,3,\underbrace{4,\dots,4}_{n-4},5,5)$*, and the star S_n.*

Proof. We compute that (computations analogous to the ones in the proof of Lemma 6.2)

$$\begin{aligned}
I_{\mathrm{ecc}}(T_{3,n}) - I_{\mathrm{ecc}}(T_{5,n}) &= \log(n-2/3) + \frac{4\log(3)-4\log(2)}{3n-2} \\
&\quad - \log(n) - \frac{16-3\log(3)-5\log(5)}{2n} \\
&\geq \log(n-2/3) - \log(n) \\
&\quad + \frac{(4/3+3/2)\log(3)+5/2\log(5)-(8+4/3)}{n-2/3} > 0
\end{aligned}$$

and

$$I_{\mathrm{ecc}}(T_{5,n}) - I_{\mathrm{ecc}}(S_n) = \log(n) + \frac{16 - 3\log(3) - 5\log(5)}{2n} - \log\left(n - \frac{1}{2}\right) - \frac{1}{2n-1}$$

is a strictly decreasing function (for $n \geq 6$) for which the limit is zero, so always positive. $\square$

To end this section, we consider trees of a given order and diameter to refute Conjecture 6.4.

Proposition 6.8. *There exists a value b such that the maximum value for I_{ecc} among all trees of diameter d and order n is obtained by the trees with eccentricity sequence*

- $(\frac{d}{2}, \frac{d}{2}+1, \frac{d}{2}+1, \dots, d, d, \underbrace{b, \dots, b}_{n-d-1})$ *for even d*
- $(\frac{d+1}{2}, \frac{d+1}{2}, \dots, d, d, \underbrace{b, \dots, b}_{n-d-1})$ *for odd d.*

Here $b \sim \frac{\sqrt[3]{2}}{\sqrt{e}} d$ as $d \to \infty$.

Proof. Let T be a tree attaining the maximum value for I_{ecc} among trees of order n. Since we can add pendant vertices with any eccentricity in $[\mathrm{rad(T)}+1, \mathrm{diam(T)}]$ to a diametrical path, we conclude by Remark 6.3 and Proposition 6.3. For d sufficiently large, we have the following computations, which we compute exactly up to a $1 + O\left(\frac{1}{d}\right)$ factor.

$$\begin{aligned}
\log(b) &\sim \frac{2\sum_{i=\frac{d}{2}}^{d} i\log(i)}{2\sum_{i=\frac{d}{2}}^{d} i} \\
&\sim \frac{\int_{\frac{d}{2}}^{d} x log(x)\,\mathrm{d}x}{\int_{\frac{d}{2}}^{d} x\,\mathrm{d}x} \\
&\sim \frac{\frac{3}{8}d^2\log(d) - \frac{3}{16\ln(2)}d^2 + \frac{1}{8}d^2}{\frac{3}{8}d^2}
\end{aligned}$$

$$= \log(d) - \frac{1}{2\ln(2)} + \frac{1}{3}.$$

Since $\exp(\log(d)/d) = 1 + o(1)$, we conclude that

$$b \sim \frac{\sqrt[3]{2}}{\sqrt{e}} d \sim 0.764d. \qquad \square$$

Corollary 6.2. *Since the eccentricity-entropy I_{ecc} is a special case of I_e, Conjecture 6.4 is not true when d is sufficiently large, as then the value b in Proposition 6.8 satisfies $b > \lceil \frac{d}{2} \rceil + 1$.*

6.4 Asymptotic minimum Wiener-entropy of graphs

We start by proving the following two elementary lemmas.

Lemma 6.3. *Let G be a connected graph and $v \in V$. Then $(n-1)\sigma(v) \geq W(G)$, with equality if and only if G is a star and v is its center.*

Proof. By rewriting the sum and applying the triangle-inequality,

$$\begin{aligned}
(n-1)\sigma(v) &= \sigma(v) + (n-2) \sum_{u \in V \backslash v} d(v,u) \\
&= \sigma(v) + \sum_{w,u \in V \backslash v} (d(u,v) + d(v,w)) \\
&\geq \sigma(v) + \sum_{w,u \in V \backslash v} d(u,w) \\
&= W(G).
\end{aligned}$$

Equality only occurs if for every $u, w \in V \backslash v$, there is a shortest path from u to w containing v. In particular, $d(u,w) \geq 2$ for all $u, w \in V \backslash v$ implies that $V \backslash v$ is an independent set. Since G is a connected graph, this implies that G is a star with center v. $\square$

Lemma 6.4. *Let G be a connected graph and $uv \in E$. Then $\sigma(v) \geq n-1$. Also $|\sigma(u) - \sigma(v)| \leq n-2$, with equality if and only if u or v is a pendant vertex.*

Proof. The first observation is trivial, since $\sigma(v)$ is the sum of $n-1$ distances that are all at least one. Let w be a vertex different from u and v. Then $d(u,w) \le d(u,v)+d(v,w) = d(v,w)+1$ and vice versa, so $|d(u,w)-d(v,w)| \le 1$. Hence we conclude, by applying the triangle inequality again: $|\sigma(u)-\sigma(v)| = |\sum_{w\in V\setminus\{u,v\}}(d(u,w)-d(v,w))| \le \sum_{w\in V\setminus\{u,v\}}|d(u,w)-d(v,w)| \le n-2$. □

Proposition 6.9. *Let T be the broom consisting of a path P_k with one of its pendant vertices c joined to $n-k$ pendant vertices by edges. For fixed $\frac{1}{3} > \epsilon > 0$, let $k = n^{1/2+\epsilon}$ and n sufficiently large. Then $I_w(T) \sim \frac{3+2\epsilon}{4}\log(n)$.*

Proof. Note that

$$\begin{aligned} W(T) &= W(P_k) + W(S_{n-k+1}) + (n-k)\left(\sum_{i=1}^{k} i\right) \\ &= \binom{k+1}{3} + (n-k)^2 + (n-k)\binom{k+1}{2} \\ &\sim \frac{nk^2}{2}. \end{aligned}$$

If ℓ is one of the pendant vertices of the star, then $\sigma(\ell) = 2(n-k-1) + \sum_{i=1}^{k} i \sim \frac{k^2}{2}$.

If v is a vertex on the path at distance $i-1$ from c, then $\sigma(v) = i(n-k) + \binom{i}{2} + \binom{k-i+1}{2}$. The sum of the transmissions for the vertices at distance $i-1$ from c for $2 \le i \le n^{2\epsilon}$ is bounded by

$$\sum_{i=2}^{n^{2\epsilon}} \left(in + \binom{k}{2}\right) < n\frac{n^{4\epsilon}}{2} + n^{2\epsilon}\frac{n^{1+2\epsilon}}{2} = n^{1+4\epsilon}.$$

Hence the sum of associated values $p_i = \frac{\sigma(v)}{2W(G)}$ is of the order $\frac{n^{1+4\epsilon}}{nk^2} = O(n^{2\epsilon-1}) = o(1)$. By Lemma 6.3, we have that every p_i is at least $\frac{1}{2(n-1)}$ and thus $-\log(p_i) \le \log(2(n-1))$. This implies that the contribution to $I_w(G)$ of the vertices at distance $i-1$ from c for $2 \le i \le n^{2\epsilon}$ is $o(\log n)$.

When $i > n^{2\epsilon}$, then $\sigma(v) \sim in$. As such, the probability $p_i = \frac{\sigma(v)}{2W(G)} \sim \frac{i}{k^2}$. For a pendant vertex ℓ of the star, we have that the associated probability for the functional $\frac{\sigma(i)}{2W(G)} \sim \frac{\frac{k^2}{2}}{k^2 n} \sim \frac{1}{2n}$.

All together, this implies that

$$\begin{aligned}
I_w(G) &\sim \sum_{i=n^{2\epsilon}}^{k} \frac{i}{k^2}(2\log(k)-\log(i)) + n\frac{1}{2n}\log(2n)\\
&\sim \log(k) - \frac{1}{k^2}\sum_{i=2}^{k} i\log(i) + \frac{1}{2}\log n\\
&\sim \frac{1}{2}\log n + \frac{1}{2}\log(k)\\
&= \frac{3+2\epsilon}{4}\log(n).
\end{aligned}$$

Here we used that $\int_1^k x\log x\,\mathrm{d}x \sim \frac{k^2}{2}\log(k)$. $\square$

Theorem 6.4. *Let G be a connected graph of order n. Then $I_w(G) > \frac{3}{4}(1+o(1))\log(n)$.*

Proof. Let $p_1, p_2, \dots, p_n$ be the n fractions of the form $\frac{\sigma(v)}{2W(G)}$, ordered in a decreasing order, i.e., $p_i \geq p_{i+1}$ for every i. By Lemma 6.3, we know $p_n \geq \frac{1}{2(n-1)} > \frac{1}{2n}$. By Lemma 6.4 and $2W(G) \geq n(n-1)$, we also note that $p_i - p_{i+1} < \frac{n-2}{n(n-1)} < \frac{1}{n}$ for every $1 \leq i \leq n-1$.

Let k be the largest number for which $k^2 - k \leq n$. Then the sequence $(p_1, p_2, \dots, p_n)$ is majorized by the sequence $(a_1, a_2, \dots, a_n)$ with $a_1 = \frac{n-k^2+3k-1}{2n}$, $a_i = \frac{2k-2i+1}{2n}$ for $2 \leq i \leq k$ and $a_i = \frac{1}{2n}$ whenever $n \geq i > k$.

Since the function $f(x) = -x\log(x)$ is concave, by Karamata's inequality we have

$$\sum_i f(p_i) \geq \sum_i f(a_i).$$

Note that one term separately is negligible, i.e., $f(a_1) = o(\log(n))$, and so we can replace it by $f\left(\frac{2k-1}{2n}\right)$ in the estimation.

Now

$$(n-k)f\left(\frac{1}{2n}\right) = \frac{1}{2}(1+o(1))\log(2n) = \frac{1}{2}(1+o(1))\log(n)$$

and

$$\begin{aligned}\sum_{i=1}^{k-1} f\left(\frac{2i-1}{2n}\right) \geq \sum_{i=1}^{k-1} f\left(\frac{i}{n}\right) \\ &= \frac{1}{2}(1+o(1))\log(n) - \frac{1}{n}\sum_{i=1}^{k-1} i\log(i) \\ &\sim \frac{1}{2}\log(n) - \frac{1}{n}\int_1^{k-1} x\log x\,\mathrm{d}x \\ &\sim \frac{1}{2}\log(n) - \frac{1}{4}\log n \\ &= \frac{1}{4}\log(n).\end{aligned}$$

Together, this implies that

$$\sum_i f(p_i) \geq \frac{3}{4}(1+o(1))\log(n).$$

□

6.5 Further thoughts on the extremal graphs for the Wiener-entropy

In Section 6.4, we determined the minimum value of $I_w(G)$ among graphs of order n asymptotically. A precise result, or characterizing the extremal graphs seems to be much harder. In this section, we present some thoughts about the extremal graphs.

From the idea of Theorem 6.4, for the class of trees, the trees of order n with minimum Wiener-entropy are expected to be brooms for sufficiently large n. If this intuition is true, the extremal broom would be completely determined by the length k of the path in Proposition 6.9, and the optimal choice $k(n)$ can be expected to be a step-wise increasing function which behaves like $n^{1/2+o(1)}$. For small n, the asymptotic analysis does not give a clear indication of the extremal tree. In particular, for $n \leq 16$ the extremal trees are not brooms.

Let $T_s(n)$ be a tree of order n which attains the minimum Wiener-entropy. For $3 \leq n \leq 16$, the trees with the minimum Wiener-entropy are listed in Table 6.1[1].

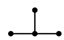
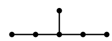
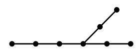
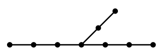
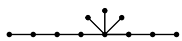
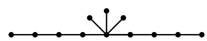
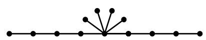
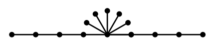
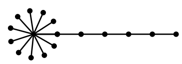
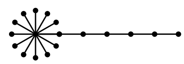
n	$T_s(n)$	n	$T_s(n)$
3		4	
5		6	
7		8	
9		10	
11		12	
13		14	
15		16	
17		18	

Table 6.1: Trees with the minimum Wiener-entropy.

Next, we focus on graphs instead of trees. As a corollary of Proposition 6.2, one can easily observe that if there are two vertices u, v for which $\sigma(u)$ and $\sigma(v)$ are small (smaller than the corresponding value β) and uv does not affect $\sigma(w)$ for $w \notin \{u, v\}$, then uv is always present. The latter also holds for a set of vertices. Starting from a broom, by the previous argument, all edges should be present between the leaves of the star, and we obtain the concatenation of a path and a clique. It has to be observed that for $n \leq 15$, the extremal graphs are not of this form. This is not surprising, as the asymptotic estimates and intuition in the proof of Theorem 6.4 are only about big order behavior, and also for trees the broom was not extremal for small n. For $5 \leq n \leq 9$, the extremal graphs are presented in Figure 6.2.

If the extremal graphs are the concatenation of a path and a clique, they would again be defined by a stepwise increasing function $k(n)$ on the integers

[1] See `https://github.com/yndongmath/wiener-entropy` for verification.

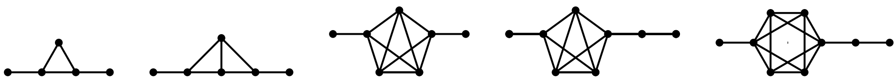


Figure 6.2: Graphs on n vertices which attain the minimum Wiener-entropy for $5 \le n \le 9$.

n. One can expect that if it would be plausible, the optimal function $k(x)$ on the reals would be more continuous. By joining one pendant vertex of the path with only a portion of the vertices of the clique, there is this more continuous behavior. Recall that $G_{n,k,j}$ is the disjoint union of a path P_k and a clique K_{n-k}, with j vertices of the clique joined to one pendant vertex of the path (Figure 6.1).

Restricted to the class of graphs of the form $G_{n,k,j}$, we computed the extremal graphs for small n. For $16 \le n \le 94$, the results reveal behavior one can expect. When n is growing, k grows stepwise and in these steps j decreases. For larger n, we observe that $j = 1$ appears more often, e.g., when $208 \le n \le 222$ this happens for 6 out of the 13 values for which $k(n) = 26$. All of these values are presented in Table 6.2.

n	(k,j)	$I_w(G_{n,k,j})$
32	(8, 22)	4.8418782994
33	(8, 20)	4.8824114556
34	(8, 18)	4.9217394089
35	(8, 15)	4.9599202002
36	(8, 12)	4.9970026044
37	(8, 8)	5.0330361551
38	(8, 4)	5.0680644063
39	(9, 26)	5.1020833397
40	(9, 23)	5.1352102662
41	(9, 20)	5.1675123079
42	(9, 16)	5.1990223046
43	(9, 12)	5.2297674906
44	(9, 8)	5.2597769036
45	(9, 3)	5.2890774027
46	(10, 31)	5.3176708476

n	(k,j)	$I_w(G_{n,k,j})$
208	(25, 1)	7.2287884533
209	(26, 159)	7.2347291497
210	(26, 133)	7.2406346487
211	(26, 108)	7.246505897
212	(26, 84)	7.2523437764
213	(26, 61)	7.2581490247
214	(26, 38)	7.2639222854
215	(26, 16)	7.2696641255
216	(26, 1)	7.2753755053
217	(26, 1)	7.281063551
218	(26, 1)	7.2867307557
219	(26, 1)	7.2923773056
220	(26, 1)	7.2980033842
221	(26, 1)	7.3036091723
222	(27, 175)	7.3091923114

Table 6.2: Minimum value of Wiener-entropy among graphs of the form $G_{n,k,j}$ for $32 \le n \le 46$ and $208 \le n \le 222$.

Maybe surprisingly, for large n, it seems that $j = 1$ always yields the extremal graphs. So, there seems to be some stability result or discretization for large values of n, which was not there for the smaller values. Noting that the size $m = \binom{n-k}{2} + j + k - 1$, one can also plot the value $I_w(G_{n,k,j})$ in terms of the size, and expect some monotonicity below and above the optimal choice. This seems to be true in large regions, but is not always true. As an example, when $n = 48$, then the optimal size is $m = 736$ ($k = 10$ and $j = 24$) and around this value the function I_w behaves nicely, but for $k = 38$ we do not have monotonicity in terms of j. This is presented in Figure 6.3 of the appendix (Section 6.6).

For a given n, let k' be the optimal choice for which there is a j such that $G_{n,k',j}$ attains the minimum Wiener-entropy among all choices of graphs of the form $G_{n,k,j}$. Then for $k = k' \pm 1$, we observed the same behavior as was the case with $n = 48$ for larger values. Nevertheless, when plotting $I_w(G_{n,k',j})$ for larger n as a function of j, there are examples with multiple local minima. This leads to the following question.

Question 6.1. *Can one explain (give some intuition for) the difference in behavior, depending on the order, for the graphs of the form $G_{n,k,j}$ minimizing $I_w(G)$.*

We expect that the extremal graphs are of the form $G_{n,k,j}$ from a reasonably small constant onwards, and we conjecture that the extremal graphs for large n are indeed the concatenation of a path and a clique, i.e., graphs of the form $G_{n,k,1}$. Based on the above discussions, we propose the following conjecture (or in fact two conjectures combined into one statement).

Conjecture 6.5. *There exists a value n_0 such that for all $n \geq n_0$, among all trees and graphs of order n, the Wiener-entropy is minimized by respectively a broom and a $G_{n,k,1}$.*

Based on a verification among the graphs of the form $G_{n,k,j}$, it seems plausible that $n_0 = 1270$. The sporadic examples for which $1000 \leq n \leq 2540$[2] and $j > 1$ are given in Table 6.3.

For $n \geq 16$, the graphs of the form $G_{n,k,j}$ with the minimum value of $I_w(G_{n,k,j})$ have been computed and are summarized in Table 6.3 for some powers of 2. For these powers of 2, the value for the Wiener-entropy can be

[2]Computations have been done within a restricted range, based on an assumption of monotonicity in k, see `https://github.com/yndongmath/wiener-entropy/tree/main/CalE`.

n	(k,j)	$I_w(G_{n,k,j})$
1003	$(67,401)$	9.1328643808
1004	$(67,75)$	9.1340468031
1029	$(68,152)$	9.163275375
1054	$(69,389)$	9.1917887671
1055	$(69,29)$	9.1929126061
1080	$(70,323)$	9.2207190796
1133	$(72,112)$	9.2775546382
1269	$(77,37)$	9.4118343668

Values of $1000 \leq n \leq 2540$ for which the graph $G_{n,k,j}$ attaining the minimum value of I_w satisfies $j > 1$

n	(k,j)	$I_w(G_{n,k,j})$
16	$(5,9)$	3.9126433225
32	$(8,22)$	4.8418782994
64	(12,26)	5.744804111
128	(19,69)	6.624593606
256	(29,4)	7.4845154156
512	(44,1)	8.32786753
1024	(67,1)	9.1574755626
2048	(101,1)	9.9757653248
4096	(152,1)	10.7847443225
8192	(225,1)	11.5860993918

Minimum values of I_w among graphs of the form $G_{n,k,j}$ in which n is a power of 2

Table 6.3: Minimum values of $I_w(G_{n,k,j})$ for some special cases.

easily compared with $\log(n)$ and as such, we observe that the $o(1)$ part in Theorem 6.4 tends to zero rather slowly.

6.6 Appendix

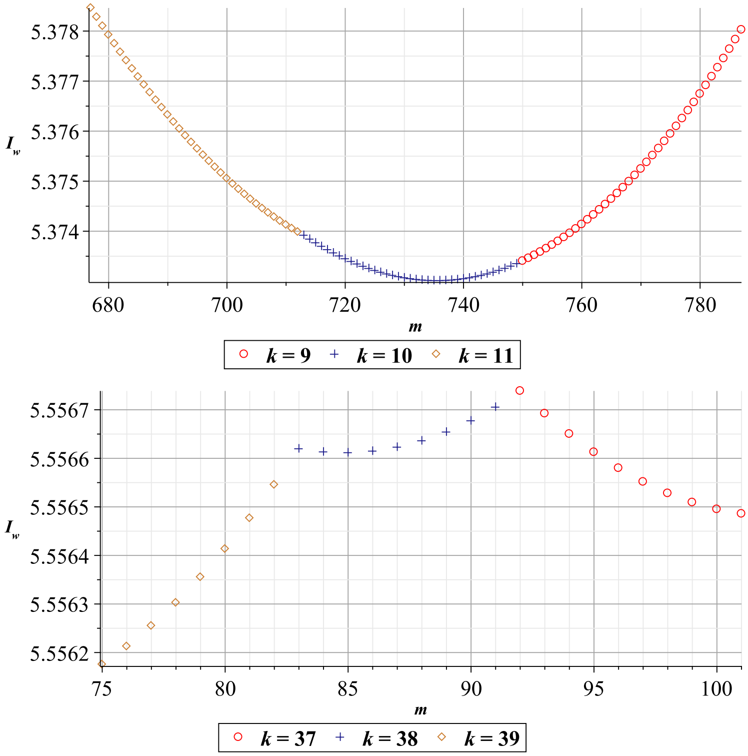


Figure 6.3: Plots of $I_w(G_{48,k,j})$ for $k \sim 10$ and $k \sim 38$.

Chapter 7

The complexity of spanning tree problems

This chapter differs from the other chapters, in the sense that it deals with algorithmic problems and their computational complexity. However, it is related to the previous chapters, in the sense that the algorithmic problems we address involve graphical function-indices, including graph entropies. In particular, in this chapter we study the computational complexity of decision and optimization problems concerning maximum and minimum spanning trees for graphical function-indices that are computable in polynomial time. Among trees of a given size and order, many topological indices attain either their maximum or minimum value for the unique case that the tree is a path. We show that either the maximum or the minimum spanning tree problems for such topological indices are $\mathcal{NP}$-complete. We also prove that if the corresponding functions are strictly convex or concave, then the minimum and maximum spanning tree problems for these graphical function-indices are $\mathcal{NP}$-complete, and their optimization versions are $\mathcal{APX}$-complete, respectively.

7.1 Introduction

Within the popular area of chemical graph theory, so-called graphical indices (also known as chemical indices or topological indices) play an important role in capturing the structural properties of molecules. Our main focus here is on graphical indices which are based on functions of the degrees of the vertices of the graphs that represent these molecules. Adopting the terminology of Li and Peng [90], for a symmetric real function $f(x, y)$, we use the unifying term graphical function-index of a graph $G = (V, E)$ for the expression $\sum_{uv \in E} f(\deg_G(u), \deg_G(v))$, where $\deg_G(\cdot)$ denotes the degree. This definition captures many well-studied graphical indices, some of which we included with their commonly used name in Table 7.1 in the appendix (Section 7.8).

The above term also captures graphical indices defined in [114, 126] by a real function $f(x)$ and the expression $\sum_{u \in V} f(\deg_G(u))$. This follows from $\sum_{uv \in E}(\frac{f(\deg_G(u))}{\deg_G(u)} + \frac{f(\deg_G(v))}{\deg_G(v)}) = \sum_{u \in V} f(\deg_G(u))$ (i.e., by choosing $f(x, y) = \frac{f(x)}{x} + \frac{f(y)}{y}$ in the above expression).

We are interested in the computational complexity of decision problems and optimization problems related to topological indices. In particular, we aim to unify decision problems concerning lower and upper bounds on the value of the graphical function-index of spanning trees, as well as the associated optimization problems, for different choices of the function $f(x, y)$. For this reason, we assume that all graphs are simple and connected. We refer to [62] for any undefined notation and terminology related to complexity.

7.2 Spanning tree problems and their complexity

As in Chapter 4, we use $GFI_f(G)$ to denote the graphical function-index of a graph $G = (V, E)$, where $GFI_f(G) = \sum_{uv \in E} f(\deg_G(u), \deg_G(v))$ for a suitable choice of the symmetric real function $f(x, y)$. Since we deal with complexity questions, throughout the chapter we assume that the chosen function $f(x, y)$ is computable in polynomial time, without explicitly mentioning it.

In this contribution, we are interested in the computational complexity of determining a spanning tree T of G that maximizes (minimizes) $GFI_f(T)$. For this purpose, we define the two associated decision problems as follows.

MAXST-GFI$_f$
INSTANCE: A graph G of order n, and a real number k.
QUESTION: Does G have a spanning tree T with $GFI_f(T) \geq k$?

MINST-GFI$_f$
INSTANCE: A graph G of order n, and a real number k.
QUESTION: Does G have a spanning tree T with $GFI_f(T) \leq k$?

Our first main observation is that the above problems are $\mathscr{NP}$-complete if paths are the unique extremal trees, in the following sense.

Theorem 7.1. *Suppose P_n is the unique tree with the largest value of $GFI_f(T)$ among all spanning trees T of G for every connected graph G of order n. Then* MAXST-GFI$_f$ *is $\mathscr{NP}$-complete.*

Theorem 7.2. *Suppose P_n is the unique tree with the smallest value of $GFI_f(T)$ among all spanning trees T of G for every connected graph G of order n. Then* MINST-GFI$_f$ *is $\mathscr{NP}$-complete.*

Since the proofs of both results are similar, we only present the proof of Theorem 7.1.

Proof. MAXST-GFI$_f$ is clearly in $\mathscr{NP}$, since it is straightforward to check in polynomial time whether $GFI_f(T) \geq k$ for any given spanning tree T of G and real number k. Note that here we implicitly use the assumption that $GFI_f(T)$ can be computed in polynomial time.

To show the $\mathscr{NP}$-completeness of MAXST-GFI$_f$, we use a reduction from the well-known $\mathscr{NP}$-complete problem HAMILTON PATH [62], which is defined as follows.

HAMILTON PATH
INSTANCE: A graph G of order n.
QUESTION: Does G have a Hamilton path, i.e., a subgraph isomorphic to P_n?

Suppose that P_n is the unique tree with the largest value of $GFI_f(T)$ among all spanning trees T of G for every connected graph G of order n. Then an arbitrary graph G is a YES-instance of HAMILTON PATH if and only if G is connected and G has a spanning tree T with $GFI_f(T) \geq GFI_f(P_n)$. This completes the proof of Theorem 7.1. $\square$

We note here that $\mathscr{NP}$-completeness results similar to the statements in Theorem 7.1 or Theorem 7.2 hold for other topological indices, as long as these indices are computable in polynomial time and similar extremal results are known or can be proved. We come back to this in Section 7.7.

We first continue with some complexity results on $GFI_f(G)$ in case $f(x,y) = \frac{f(x)}{x} + \frac{f(y)}{y}$ for a real function $f(x)$. Recall that this implies $GFI_f(G) = \sum_{u \in V} f(\deg_G(u))$. We focus on the cases for which we know that $f(x)$ is strictly concave or convex. In these cases we can prove that the above problems MAXST-GFI$_f$ and MINST-GFI$_f$ remain $\mathscr{NP}$-complete when restricted to cubic graphs, i.e., instance graphs in which all vertices have degree 3.

7.3 More complexity results

In this section, throughout we assume $GFI_f(G) = \sum_{u \in V} f(\deg_G(u))$ for a real function $f(x)$ which is computable in polynomial time. We consider the below special cases of MAXST-GFI$_f$ and MINST-GFI$_f$ for such $GFI_f(G)$ and restricted to cubic graphs.

C-MAXST-GFI$_f$
INSTANCE: A cubic graph G of order n, and a real number k.
QUESTION: Does G have a spanning tree T with $GFI_f(T) \geq k$?

C-MINST-GFI$_f$
INSTANCE: A cubic graph G of order n, and a real number k.
QUESTION: Does G have a spanning tree T with $GFI_f(T) \leq k$?

In the next section, we show how we can use the following problem 1,3-ST to prove complexity results if $f(x)$ is assumed to be either strictly concave or convex in the above problems.

1,3-ST
INSTANCE: A cubic graph G of order n.
QUESTION: Does G have a spanning tree with no vertices of degree 2?

The problem 1,3-ST is known to be $\mathscr{NP}$-complete by a result of Lemke [86]. We use it to deduce the following complexity results.

Theorem 7.3. *If f is a strictly concave function, then* C-MINST-GFI$_f$ *is $\mathscr{NP}$-complete.*

Theorem 7.4. *If f is a strictly convex function, then* C-MAXST-GFI$_f$ *is $\mathscr{NP}$-complete.*

We only prove Theorem 7.3; the counterpart for strictly convex functions can be proved in a similar way.

The above results directly imply that MINST-GFI$_f$ (resp., MAXST-GFI$_f$) are $\mathscr{NP}$-complete if f is a strictly concave (resp., convex) function and $GFI_f(G) = \sum_{u \in V} f(\deg_G(u))$ for a real function $f(x)$ which is computable in polynomial time. We show examples of known topological indices satisfying these conditions in Section 7.7.

In the next section, we introduce some additional terminology and notation, and present our proof of Theorem 7.3.

7.4 Proof of Theorem 7.3

We start this section with two useful lemmas, but first need to recall some additional terminology and notation.

For a graph G, we use the same notation $D(G)$ and the same way to denote the degree sequence of G as in Subsection 5.1.2. Let A and B be two non-increasing (degree) sequences of the same length. Then, as previously, we use $A \succ B$ to denote that A strictly majorizes B. If T is any tree of order n different from P_n (resp., $K_{1,n-1}$), then clearly $D(T) \succ D(P_n)$ (resp., $D(K_{1,n-1}) \succ D(T)$). So, by Karamata's inequality (Theorem 4.2), we immediately obtain the following result as a consequence.

Lemma 7.1. *Let T be a tree of order n. If f is a strictly concave (resp., convex) function, then*

(a) *$GFI_f(T) \leq GFI_f(P_n)$ (resp., $GFI_f(T) \geq GFI_f(P_n)$), with equality holding in the inequality if and only if $T \cong P_n$;*

(b) *$GFI_f(T) \geq GFI_f(K_{1,n-1})$ (resp., $GFI_f(T) \leq GFI_f(K_{1,n-1})$), with equality holding in the inequality if and only if $T \cong K_{1,n-1}$.*

Note that the statements in Lemma 7.1 (a) can be considered as special cases of the statements in Theorems 7.1 and 7.2, for the case that $GFI_f(G) = \sum_{u\in V} f(\deg_G(u))$. Theorem 7.3 and its counterpart show the stronger versions of the latter case when restricted to cubic graphs. We complete this section with our proof of Theorem 7.3.

Proof of Theorem 7.3. We assume that f is a strictly concave function, and our aim is to prove that the problem C-MINST-GFI$_f$ is $\mathcal{NP}$-complete.

The problem is clearly in $\mathcal{NP}$.

Let T^* be a tree of order n with degree sequence $[3^{\frac{n-2}{2}}, 1^{\frac{n+2}{2}}]$ (where n is even by the well-known fact that the degree-sum is twice the number of edges). We consider the problem C-MINST-GFI$_f$ for $k = GFI_f(T^*)$. We prove the required $\mathcal{NP}$-completeness by a reduction from the 1,3-ST problem. Recall that the latter problem is $\mathcal{NP}$-complete. Let G be a cubic graph with vertex set $\{v_1, v_2, \ldots, v_n\}$. It is sufficient to prove that G has a spanning tree with no vertices of degree 2 if and only if G has a spanning tree T with $GFI_f(T) \leq GFI_f(T^*)$.

Suppose first that T' is a spanning tree of G with no vertices of degree 2. Set $D(T') = [3^a, 1^b]$. Then $a + b = n$ and $3a + b = 2(n-1)$. By straightforward calculations, we obtain $a = \frac{n-2}{2}$ and $b = \frac{n+2}{2}$. So we have $D(T') = [3^{\frac{n-2}{2}}, 1^{\frac{n+2}{2}}] = D(T^*)$. Thus $GFI_f(T') = GFI_f(T^*)$, and hence G has a spanning tree $T = T'$ with $GFI_f(T) \leq GFI_f(T^*)$.

For the other implication, suppose that all spanning trees of G have at least one vertex of degree 2. Let T be an arbitrary spanning tree of G. We complete the proof by showing that $GFI_f(T) > GFI_f(T^*)$.

Set $D(T) = [3^{a_1}, 2^{a_2}, 1^{a_3}] = (d_1, d_2, \ldots, d_n)$. So we have $a_2 \geq 1$. Then $a_1 + a_2 + a_3 = n$ and $3a_1 + 2a_2 + a_3 = 2(n-1)$. By straightforward calculations, we have $a_1 = \frac{n-a_2-2}{2}$ and $a_3 = \frac{n-a_2+2}{2}$. Since $a_2 \geq 1$, we have $a_1 < \frac{n-2}{2}$ and $a_3 < \frac{n+2}{2}$. Recall that $D(T^*) = [3^{\frac{n-2}{2}}, 1^{\frac{n+2}{2}}] = (d'_1, d'_2, \ldots, d'_n)$. It is easy to check the validity of the following inequalities.

$$\sum_{i=1}^{k} d_i = 3k = \sum_{i=1}^{k} d'_i \text{ for } k = 1, 2, \ldots, a_1;$$

$$\sum_{i=1}^{k} d_i = 3a_1 + 2(k-a_1) < 3k = \sum_{i=1}^{k} d'_i \text{ for } k = a_1+1, a_1+2, \ldots, \frac{n-2}{2};$$

$$\sum_{i=1}^{k} d_i = 2n-2-2t-a_3 \leq 2n-2-t-a_3 = \sum_{i=1}^{k} d'_i$$
$$\text{for } t = n-a_3-k \text{ and } k = \frac{n}{2}, \frac{n}{2}+1, \ldots, a_1+a_2;$$

$$\sum_{i=1}^{k} d_i = 2n-2-\ell = \sum_{i=1}^{k} d'_i$$
$$\text{for } \ell = n-k \text{ and } k = a_1+a_2+1, a_1+a_2+2, \ldots, n.$$

Hence $D(T^*) \succ D(T)$. Since we assume f is strictly concave, using Theorem 4.2, we conclude that $GFI_f(T) > GFI_f(T^*)$. This completes the proof of Theorem 7.3. □

7.5 $\mathscr{APX}$−completeness

In this and the next section, we continue to consider $GFI_f(G) = \sum_{u \in V} f(\deg_G(u))$ for a real function $f(x)$ which is computable in polynomial time.

However, we will turn our attention to the optimization problems associated with the decision problems we considered in the previous sections.

Adopting the way optimization problems are presented in a classic paper by D.S. Johnson [73], we list the optimization versions of C-MINST-GFI$_f$ and

C-MAXST-GFI$_f$ as follows.

C-MINST-GFI$_f$
INSTANCE: A cubic graph G of order n.
FEASIBLE SOLUTION: A spanning tree T of G.
OBJECTIVE FUNCTION: $GFI_f(T)$.
OPT: Min.

C-MAXST-GFI$_f$
INSTANCE: A cubic graph G of order n.
FEASIBLE SOLUTION: A spanning tree T of G.
OBJECTIVE FUNCTION: $GFI_f(T)$.
OPT: Max.

The following results deal with the $\mathscr{APX}$-completeness of the above optimization versions of C-MINST-GFI$_f$ and C-MAXST-GFI$_f$. These results imply that there exists an $\epsilon > 0$ such that no polynomial time $(1+\epsilon)$-approximation algorithm is possible for these two problems, unless $\mathscr{P} = \mathscr{NP}$ [9].

Theorem 7.5. *If f is a strictly concave function, then the optimization version of* C-MINST-GFI$_f$ *is $\mathscr{APX}$-complete.*

Theorem 7.6. *If f is a strictly convex function, then the optimization version of* C-MAXST-GFI$_f$ *is $\mathscr{APX}$-complete.*

We only prove Theorem 7.5 below, since the proof of Theorem 7.6 is similar. We first give some remarks. The above results directly imply that the optimization version of MINST-GFI$_f$ (resp., MAXST-GFI$_f$) is $\mathscr{APX}$-complete if f is a strictly concave (resp., convex) function and $GFI_f(G) = \sum_{u \in V} f(\deg_G(u))$ for a real function $f(x)$ which is computable in polynomial time.

There exists an extensive literature on proofs for $\mathscr{APX}$-completeness of optimization problems by L-reductions (See, e.g., [2, 38, 102, 105]). Given two optimization problems F and G, and a polynomial time transformation h from instances of F to instances of G, we say that h is an *L-reduction* if there are positive constants α and β such that for every instance x of F

1. $\mathrm{opt}_G(h(x)) \le \alpha \cdot \mathrm{opt}_F(x)$;

2. for every feasible solution y of $h(x)$ with objective value $g_G(h(x), y) = c_2$, we can in polynomial time find a solution y' of x with $g_F(x, y') = c_1$ such that $|\mathrm{opt}_F(x) - c_1| \leq \beta \cdot |\mathrm{opt}_G(h(x)) - c_2|$.

We next prove Theorem 7.5 by an L-reduction; the counterpart for strictly convex functions can be proved in a similar way. For the full proof of Theorem 7.5 we also need to show that C-MINST-GFI$_f \in \mathscr{APX}$. This will be done in Section 7.6.

Proof of Theorem 7.5. We present our proof that C-MINST-GFI$_f \in \mathscr{APX}$ in Section 7.6. Next we prove the $\mathscr{APX}$-hardness by an L-reduction from the optimization problem C-MLST to C-MINST-GFI$_f$. The c-MLST problem is defined as follows; it is known to be $\mathscr{APX}$-complete by a result due to Bonsma [23].

C-MLST
INSTANCE: A cubic graph G of order n.
FEASIBLE SOLUTION: A spanning tree T of G.
OBJECTIVE FUNCTION: The number of leaves of T.
OPT: Max.

Let $G = (V, E)$ be a cubic graph of order n. Let T_1 and T_2 be two spanning trees of G on distinct numbers of leaves ℓ_1 and ℓ_2, respectively. It is easy to see that $D(T_i) = [3^{\ell_i - 2}, 2^{n+2-2\ell_i}, 1^{\ell_i}]$ for $i = 1, 2$. Under the majorization relation, any pair of degree sequences of two spanning trees of G are comparable. This implies either $D(T_1) \succ D(T_2)$ or $D(T_2) \succ D(T_1)$. By straightforward calculations, $\ell_1 > \ell_2$ if and only if $D(T_1) \succ D(T_2)$. By Theorem 4.2, we have $GFI_f(T_1) < GFI_f(T_2)$ if and only if $\ell_1 > \ell_2$. This implies that a spanning tree T^* of G has the maximum number ℓ^* of leaves if and only if $GFI_f(T^*)$ attains the minimum value among all spanning trees of G. Let T be a spanning tree of G with $GFI_f(T) = (\ell - 2)f(3) + (n + 2 - 2\ell)f(2) + \ell f(1)$. This implies that T has ℓ leaves. Since $f(x)$ is strictly concave, we have

$$\left| \frac{(\ell^* - 2)f(3) + (n + 2 - 2\ell^*)f(2) + \ell^* f(1)}{\ell^*} \right|$$

$$
\begin{aligned}
&\leq \left|\frac{\ell^*(f(3)-2f(2)+f(1))}{\ell^*}\right| + \left|\frac{(n+2)f(2)}{\ell^*}\right| + \left|\frac{2f(3)}{\ell^*}\right| \\
&\leq 2f(2)-f(3)-f(1)+\frac{(n+2)|f(2)|}{2}+|f(3)| \\
&\leq 2|f(3)|+\frac{(n+6)|f(2)|}{2}-f(1)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\frac{\ell^*-\ell}{(\ell^*-\ell)f(3)-2(\ell^*-\ell)f(2)+(\ell^*-\ell)f(1)}\right| \\
&= \frac{1}{2f(2)-f(3)-f(1)}.
\end{aligned}
$$

Therefore,

$$
|(\ell^*-2)f(3)+(n+2-2\ell^*)f(2)+\ell^* f(1)| \leq \alpha \cdot |\ell^*|
$$

and

$$
|\ell^*-\ell| \leq \beta \cdot |(\ell^*-\ell)f(3)-2(\ell^*-\ell)f(2)+(\ell^*-\ell)f(1)|,
$$

where $\alpha = 2|f(3)| + \frac{(n+6)|f(2)|}{2} - f(1)$ and $\beta = \frac{1}{2f(2)-f(3)-f(1)}$. □

7.6 Approximation algorithm

In this section, we prove the following results for the optimization problems we introduced in the previous section. As before, we only give the details for the minimization version.

Theorem 7.7. *If f is a strictly concave function, then* C-MINST-GFI$_f \in \mathscr{APX}$.

Theorem 7.8. *If f is a strictly convex function, then* C-MAXST-GFI$_f \in \mathscr{APX}$.

We prove Theorem 7.7 by showing that the next algorithm is a polynomial-time approximation algorithm for the optimization version of C-MINST-GFI$_f$. An explanation of the notation and rationale of the algorithm follows.

Algorithm 1 LOCAL SEARCH-C-MINST-GFI$_f$

Initial: a cubic graph G.
Output: a spanning tree T' and $GFI_f(T')$.

Initialize a spanning tree T of G;
$\widehat{E} \leftarrow E(G) \setminus E(T)$;
while $\widehat{E} \neq \emptyset$ **do**
    randomly choose an edge $e \in \widehat{E}$;
    if $e \in E^*(T+e)$ **then**
        $\widehat{E} \leftarrow \widehat{E} \setminus \{e\}$;
    else
        randomly choose an edge $e^* \in E^*(T+e)$;
        $T \leftarrow T + e - e^*$;
        $\widehat{E} \leftarrow E(G) \setminus E(T)$;
    end if
end while
return $T' \leftarrow T$ and $GFI_f(T') \leftarrow GFI_f(T)$.

Let $G = (V, E)$ be a graph. Define $h_G(w) = f(\deg_G(w)) - f(\deg_G(w) - 1)$ for $w \in V(G)$. Suppose that $e \in E(G)$ is an edge joining vertices $u \in V(G)$ and $v \in V(G)$. Define $\Delta_G(e) = h_G(u) + h_G(v)$. Let $\Delta_G^* = \max_{e \in E(G)} \{\Delta_G(e) | G - e \text{ is connected}\}$ and $E^*(G) = \{e | \Delta_G(e) = \Delta_G^* \text{ and } G - e \text{ is connected}\}$. Our local search algorithm is based on the following observation: if $e^* \in E^*(G)$, then $GFI_f(G - e^*)$ attains the minimum value among all connected spanning subgraphs of size $|E(G)| - 1$. If we add a new edge e to a tree T, then we obtain a unicyclic graph. Deleting an edge $e^* \in E^*(T+e)$ gives us a new tree $T + e - e^*$ with $GFI_f(T + e - e^*) \leq GFI_f(T)$. This is the rationale behind the algorithm LOCAL SEARCH-C-MINST-GFI$_f$.

Before proving Theorem 7.7, we take on the job of analyzing its running time on an input cubic graph G of order n. We may call the breadth-first search algorithm of [37] to generate a spanning tree of G in running time $O(n)$. Let T be a spanning tree of G. Since G is a cubic graph, for any edge $e \in E(G) \setminus E(T)$, Δ_{T+e}^* is one value in the set $\{2f(3) - 2f(2), f(3) - f(1), 2f(2) - 2f(1)\}$. This implies that, for $e^* \in E^*(T+e)$, $GIF_f(T) - GIF_f(T + e - e^*) \geq 2f(2) - f(3) - f(1)$ if $GIF_f(T + e - e^*) < GIF_f(T)$. By Lemma 7.1, we have $GIF_f(T) \leq (n-2)f(2) + 2f(1)$. The number of iterations of the while-loop is at most

$\frac{(n-2)f(2)+2f(1)}{2f(2)-f(3)-f(1)}$, which is $O(n)$. Each iteration runs in time $(\frac{n}{2}+1)\times n = O(n^2)$. Therefore, LOCAL SEARCH-C-MINST-GFI$_f$ runs in time $O(n^3)$, where n is the order of the input cubic graph G.

Now we have all the ingredients to prove Theorem 7.7.

Proof of Theorem 7.7. Let G be a cubic graph of order n. Let T' be a spanning tree of G obtained by using Algorithm 1. Let a, b and c be the numbers of vertices of degree 3, 2 and 1 of T', respectively. We first state and prove two claims.

Claim 1. $a < c$.

Proof. Using the above notation, we have $3a+c = 2n-2b-2$ and $a+c = n-b$. This implies $a-c = -2 < 0$. $\diamondsuit$

An M-path P is a maximal path with all internal vertices of degree 2 and ends of degrees 1 or 3. It follows that the sum of the numbers of internal vertices of M-paths of T' is b.

Claim 2. $b \leq 4c$.

Proof. By Claim 1, the number of M-paths is less than $2c$. If each M-path has at most two internal vertices, then the claim holds. If all internal vertices of every M-path are adjacent to pendant vertices of T', then the claim holds as well. This follows since each pendant vertex of T' is adjacent to at most two internal vertices, hence $b \leq 2c \leq 4c$.

Let P be an M-path. Suppose that there exist at least three internal vertices of P, with one internal vertex not adjacent to any pendant vertex of T'.

Suppose that u is an internal vertex of P and w is a vertex in G with $\deg_{T'}(w) = 2$ satisfying $e = uw \in E(G) \setminus E(T')$. Then there exists a vertex $v \in N_P(u)$ such that $e' = uv \in E(P)$. It follows that $T' + e - e'$ is a spanning tree of G.

By straightforward calculations, we obtain the following expression for $\Delta_{T'+e}(e') - \Delta_{T'+e}(e)$:

$$\Delta_{T'+e}(e') - \Delta_{T'+e}(e)$$

$$
\begin{aligned}
=&(f(\deg_{T'+e}(u))-f(\deg_{T'+e}(u)-1)\\
&+f(\deg_{T'+e}(v))-f(\deg_{T'+e}(v)-1))\\
&-(f(\deg_{T'+e}(u))-f(\deg_{T'+e}(u)-1)\\
&+f(\deg_{T'+e}(w))-f(\deg_{T'+e}(w)-1))\\
=&f(\deg_{T'+e}(v))-f(\deg_{T'+e}(v)-1)-f(\deg_{T'+e}(w))+f(\deg_{T'+e}(w)-1)\\
=&f(2)-f(1)-f(3)+f(2)\\
=&2f(2)-f(3)-f(1).
\end{aligned}
$$

Recall that a real-valued function f on an interval is said to be strictly concave if $f((1-\alpha)x+\alpha y) > (1-\alpha)f(x)+\alpha f(y)$ for any $\alpha \in (0,1)$ and $x \neq y$. Setting $x=3$, $y=1$ and $\alpha=\frac{1}{2}$, we get $f(2) > \frac{1}{2}f(3)+\frac{1}{2}f(1)$ (i.e., $2f(2)-f(3)-f(1) > 0$). So we obtain $\Delta_{T'+e}(e') > \Delta_{T'+e}(e)$. This implies $GFI_f(T'+e-e') < GFI_f(T')$, which contradicts that T' is a minimal spanning tree of G constructed by Algorithm 1.

This completes the proof of Claim 2. $\diamondsuit$

Let $\widehat{T}$ be a spanning tree of G with the minimum graphical function-index. Let T^* be a tree with degree sequence $D(T^*)=[3^{\frac{n-2}{2}},1^{\frac{n+2}{2}}]$. From the proof of Theorem 7.3, we have $GFI_f(T^*) \leq GFI_f(\widehat{T})$ and $c \leq \frac{n+2}{2}$. By Claims 1 and 2, we have $n=a+b+c<6c$. This implies $a+b<\frac{5n}{6}$. Since G is a cubic graph, we have $n \geq 4$. If $f(3) \geq f(2)$, then we get

$$
\begin{aligned}
GFI_f(G)(T') &= af(3)+bf(2)+cf(1)\\
&\leq (a+b)f(3)+cf(1)\\
&\leq \frac{5n}{6}f(3)+\frac{n+2}{2}f(1)\\
&\leq \frac{10}{3}\left(\frac{n-2}{2}f(3)+\frac{n+2}{2}f(1)\right)\\
&= \frac{10}{3}GFI_f(T^*)\\
&\leq \frac{10}{3}GFI_f(\widehat{T}).
\end{aligned}
$$

If $f(3) < f(2)$, then we get

$$\begin{aligned}
GFI_f(G)(T') &= af(3) + bf(2) + cf(1) \\
&\leq \frac{5n}{6} f(2) + \frac{n+2}{2} f(1) \\
&\leq \frac{10f(2)}{3f(3)} \left(\frac{n-2}{2} f(3) + \frac{n+2}{2} f(1) \right) \\
&= \frac{10f(2)}{3f(3)} GFI_f(T^*) \\
&\leq \frac{10f(2)}{3f(3)} GFI_f(\widehat{T}).
\end{aligned}$$

It follows that $GFI_f(G)(T') \leq \gamma \cdot GFI_f(\widehat{T})$ in which $\gamma = \max\{\frac{10}{3}, \frac{10f(2)}{3f(3)}\}$. $\square$

7.7 Concluding remarks

To complete this chapter, we reflect on an earlier remark by listing some examples of well-studied topological indices in Table 7.2 of the appendix (Section 7.8) that do not fall under our general description, but for which similar complexity results hold. Following the table, we gathered the known associated extremal results from literature in Theorem 7.9 below. The consequences of these extremal results for the complexity of our studied decision and optimization problems are summarized in two corollaries.

For a better understanding of the expressions of the topological indices that are listed in the table, we introduce some additional notation, without going into the details.

Let G be a graph. As usual, with $d(u, v)$ we denote the distance between two vertices u and v in G, assuming that G is connected. Under the same assumption, we use $\delta_G(u)$ to denote the distance sum of a vertex $u \in V(G)$ to all other vertices in G. In the table, $\mu = |E(G)| - |V(G)| + 1$ is used for the cyclomatic number, and $\lambda_1, \lambda_2, \ldots, \lambda_{|V(G)|}$ indicate the eigenvalues of the adjacency matrix of G. Furthermore, α is used to indicate a real number, k is an integer, and $m(G, k)$ denotes the number of distinct matchings of G consisting of k edges. So, by definition, $m(G, 0) = 1$ and $m(G, 1) = |E(G)|$. Finally, in the table $i(G,k)$ denotes the number of distinct k-element independent vertex sets of G. So, by definition, $i(G,0)=1$ and $i(G,1)=|V(G)|$.

Below we present the known extremal results with respect to the listed topological indices, together with their sources.

Theorem 7.9. *Among trees on n vertices, P_n is the unique extremal tree with the minimum Balaban index [51, 52], the maximum degree-based graph entropy for $k>0$ [72], the maximum energy [91], the minimum Estrada index [40], the minimum general first Zagreb index $M_1^\alpha(G)$ for $\alpha<0$ or $\alpha>1$ [92], the maximum general first Zagreb index $M_1^\alpha(G)$ for $0<\alpha<1$ [92], the minimum general Randić index $R_\alpha(G)$ for $0<\alpha\leq 1$ [70], the maximum general Randić $R_\alpha(G)$ for $\alpha<0$ [69], the maximum geometrical-arithmetic index [116], the maximum harmonic index [130], the maximum Hosoya index [117], the maximum hyper-Wiener index [66], the minimum Merrifield-Simmons index [117], the minimum second Zagreb index [85], the maximum sum-connectivity index [30], and the maximum Wiener index [58].*

In order to show these results more clearly, we list some known topological indices with P_n as the unique extremal tree in Tables 7.3 and 7.4 of the appendix (Section 7.8). The above extremal results have the following consequences. In the first corollary, the decision version of the maximization problem for a specific topological index $TI(G)$ is the problem of deciding whether G has a spanning tree T with $TI(T)\geq k$, for an arbitrary graph G and real number k. The decision versions of the minimization problems are defined analogously.

Corollary 7.1. *The decision version of the maximization problem is $\mathcal{NP}$-complete for the following topological indices: the degree-based entropy for $k>0$, the energy, the general first Zagreb index $M_1^\alpha(G)$ for $0<\alpha\leq 1$, the general Randić index R_α for $\alpha<0$, the harmonic index, the Hosoya index, the hyper-Wiener index, the sum-connectivity index, and the Wiener index.*

The decision version of the minimization problem is $\mathcal{NP}$-complete for the following topological indices: the Balaban index, the Estrada index, the general first Zagreb index $Z_1^\alpha(G)$ for $\alpha<0$ or $\alpha>1$, the general Randić index $R_\alpha(G)$ for $0<\alpha\leq 1$, the Merrifield-Simmons index, and the second Zagreb index.

The general first Zagreb index M_1^α (resp., the degree-entropy I_d) corresponds to $f(x) = x^\alpha$ (resp., $f(x) = \frac{x}{2|E(G)|}\log_2(\frac{2|E(G)|}{x})$). We immediately obtain that $f(x)$ regarding M_1^α is strictly concave for $1 < \alpha < 1$, and strictly convex for $\alpha < 0$ or $\alpha > 1$; $f(x)$ regarding I_d is strictly concave. By applying Theorems 7.3, 7.4, 7.5 and 7.6, we obtain the following result.

Corollary 7.2. *The maximization problem is $\mathscr{NP}$-complete and $\mathscr{APX}$-complete for the general first Zagreb index M_1^α for $\alpha < 0$ or $\alpha > 1$.*

The minimization problem is $\mathscr{NP}$-complete and $\mathscr{APX}$-complete for the following indices: the general first Zagreb index M_1^α for $0 < \alpha < 1$, and the degree-entropy I_d.

7.8 Appendix

Table 7.1: Some known graphical function-indices.

Name	$f(x,y) =$
First Zagreb index	$x+y$
Second Zagreb index	xy
First hyper-Zagreb index	$(x+y)^2$
Second hyper-Zagreb index	$(xy)^2$
Modified first Zagreb index	$x^{-3}+y^{-3}$
Albertson index	$\lvert x-y \rvert$
Extended index	$(x/y+y/x)/2$
Sigma index	$(x-y)^2$
Randić index	$1/\sqrt{xy}$
Reciprocal Randić index	$\sqrt{xy}$
Sum-connectivity index	$1/\sqrt{x+y}$
Reciprocal sum-connectivity index	$\sqrt{x+y}$
Harmonic index	$2/(x+y)$
Atom-bond connectivity index	$\sqrt{(x+y-2)/(xy)}$
Argumented Zagreb index	$x^3y^3/(x+y-2)^3$
Forgotten index	x^2+y^2
Inverse degree	$x^{-2}+y^{-2}$
Geometric-arithmetic index	$2\sqrt{xy}/(x+y)$
Arithmetic-geometric index	$(x+y)/2\sqrt{xy}$
Inverse sum index	$xy/(x+y)$
First Gourava index	$x+y+xy$
Second Gourava index	$(x+y)xy$
First hyper-Gourava index	$(x+y+xy)^2$
Second hyper-Gourava index	$x^2y^2(x+y)^2$
Sum-connectivity Gourava index	$1/\sqrt{x+y+xy}$
Product-connectivity Gourava index	$\sqrt{(x+y)xy}$
Sombor index	$\sqrt{x^2+y^2}$

Table 7.2: Some topological indices and their definition.

Name	Expression
Balaban index	$J(G) = \frac{\lvert E(G)\rvert}{\mu+1} \sum_{uv\in E(G)} \frac{1}{\sqrt{\delta_G(u)\delta_G(v)}}$
Degree-based entropy	$I_d^k(G) = -\sum_{u\in V(G)} \frac{\deg_G(u)^k}{\sum_{v\in V(G)} \deg_G(v)^k} \log\left(\frac{\deg_G(u)^k}{\sum_{v\in V(G)} \deg_G(v)^k}\right)$
Degree-entropy	$I_d(G) = -\sum_{u\in V(G)} \frac{\deg_G(u)}{2\lvert E(G)\rvert} \log\left(\frac{\deg_G(u)}{2\lvert E(G)\rvert}\right)$
Energy	$EN(G) = \sum_{i=1}^{n} \lvert\lambda_i\rvert$
Estrada index	$EE(G) = \sum_{i=1}^{n} e^{\lambda_i}$
First Zagreb index	$M_1(G) = \sum_{v\in V(G)} \deg_G(v)^2$
General first Zagreb index	$M_1^\alpha(G) = \sum_{v\in V(G)} \deg_G(v)^\alpha$
General Randić index	$R_\alpha(G) = \sum_{uv\in E(G)} (\deg_G(u)\deg_G(v))^\alpha$
Geometric-arithmetic index	$GA(G) = \sum_{uv\in E(G)} \frac{2\sqrt{\deg_G(u)\deg_G(v)}}{\deg_G(u)+\deg_G(v)}$
Harmonic index	$H(G) = \sum_{uv\in E(G)} \frac{2}{\deg_G(u)+\deg_G(v)}$
Hosoya index	$Z(G) = \sum_{k\geq 0} m(G,k)$
Hyper-Wiener index	$WW(G) = \frac{\sum_{u,v\in V(G)} d(u,v) + \sum_{u,v\in V(G)} d^2(u,v)}{2}$
Merrifield-Simmons index	$\sigma(G) = \sum_{k\geq 0} i(G,k)$
Randić index	$R(G) = \sum_{uv\in E(G)} \frac{1}{\sqrt{\deg_G(u)\deg_G(v)}}$
Second Zagreb index	$M_2(G) = \sum_{uv\in E(G)} \deg_G(u)\deg_G(v)$
Sum-connectivity index	$\chi(G) = \sum_{uv\in E(G)} \frac{1}{\sqrt{\deg_G(u)\deg_G(v)}}$
Wiener index	$W(G) = \sum_{\{u,v\}\subseteq V(G)} d(u,v)$

Table 7.3: Some known topological indices with P_n
as the unique extremal tree attaining the maximum value.

Name	Notation	Condition
Degree-based entropy	$I_d^k(G))$	$k > 0$
Energy	$EN(G)$	None
General first Zagreb index	$M_1^\alpha(G)$	$0 < \alpha < 1$
General Randić index	$R_\alpha(G)$	$\alpha < 0$
Geometric-arithmetic index	$GA(G)$	None
Harmonic index	$H(G)$	None
Hosoya index	$Z(G)$	None
Hyper-Wiener index	$WW(G)$	None
Sum-connectivity index	$\chi(G)$	None
Wiener index	$W(G)$	None

Table 7.4: Some known topological indices with P_n as the unique extremal tree attaining the minimum value.

Name	Notation	Condition
Balaban index	$J(G)$	None
Estrada index	$EE(G)$	None
General first Zagreb index	$M_1^\alpha(G)$	$\alpha < 0$ or $\alpha > 1$
General Randić index	$R_\alpha(G)$	$0 < \alpha \leq 1$
Merrifield-Simmons index	$\sigma(G)$	None
Randić index	$R(G)$	None
Second Zagreb index	M_2	None

Summary

This thesis focuses on extremal problems involving various graph parameters that were introduced in graph theory motivated by the well-known Shannon entropy in information theory. In particular, new results are obtained for extremal problems involving degree-based and distance-based entropies. Moreover, results on the computational complexity of decision and optimization problems concerning maximum and minimum spanning trees for graphical function indices are derived.

Determining extremal values of topological indices and the extremal graphs that attain these values is a popular research direction within graph theory. In Chapter 1 of this thesis we give a general introduction to this theme, together with some historical background and an overview of the contributions from this thesis. One of the approaches is to look at the effect of certain graph operations on the value of such indices. This approach is particularly useful if the extremal graphs can be obtained through a series of graph operations. In Chapter 2, we consider the effect of graph operations, including concepts like the weak product, the blow-up and the identification of vertices, on the degree-based entropy.

In Chapter 3 to Chapter 5, we study extremal problems involving the degree-based entropy restricted to some specific graph classes. In Chapter 3, we focus on trees and unicyclic graphs, and determine the maximum and minimum values of the degree-based entropy under some given parameters, including the diameter and a given bipartition. This is mainly done by analyzing the effect of graph operations on the degree-based entropy.

Chapter 4 studies the extremal values of the degree-based entropy of bipartite graphs, using the representation of bipartite graphs by Young diagrams. We prove that the extremal graph attaining the minimum value is a complete bipartite graph or nearly complete bipartite graph. We show that the general problem of characterizing the extremal graphs is related to a very complicated problem in number theory. Therefore, we believe that this problem is difficult to solve. Among bipartite graphs with a given order and size, we use the degree sequence to characterize the extremal graphs attaining the maximum value of the degree-based entropy. We extend these results to some generalized graphical function indices. We prove that the difference between the maximum degree and the minimum degree of the degree sequence of the extremal graphs does not exceed 2.

In Chapter 5, we fully characterize the extremal graphs for which the degree-based entropy attains the minimum value among all graphs with a given order and size. The extremal graphs turn out to be special so-called threshold graphs.

In Chapter 6, we study two kinds of distance-based entropies, namely the eccentricity-entropy, denoted by I_{ecc}, and the Wiener-entropy, denoted by I_w. By deriving the (asymptotic) extremal behavior, we conclude that the Wiener-entropy of graphs of a given order is more spread than the eccentricity-entropy. We resolve three known conjectures on I_{ecc} and propose two new conjectures on I_w, which we formulate at the end of this summary.

In Chapter 7, we study the computational complexity of decision and optimization problems concerning maximum and minimum spanning trees for graphical function indices that are computable in polynomial time. Among trees of a given size and order, many topological indices attain either their maximum or minimum value for the unique case that the tree is a path. We show that either the maximum or the minimum spanning tree problems for such topological indices are $\mathscr{NP}$-complete. We also prove that if the corresponding functions are strictly convex or concave, then the minimum and maximum spanning tree problems for these graphical function indices are $\mathscr{NP}$-complete, and their optimization versions are $\mathscr{APX}$-complete, respectively.

Although this thesis contains many new results, several questions and conjectures remain open. One of the challenging open problems is to fully characterize the bipartite graphs of a given size and order which minimize the degree-based entropy. Regarding the Wiener-entropy, we propose the following two conjectures (in one statement). For this, we define the graph $G_{n,k,j}$ and the broom of order $n > k \geq 2$ as follows. To construct $G_{n,k,j}$, take the disjoint union of a path P_k and a complete graph K_{n-k}, and join one end vertex of the path with j vertices of the complete graph by edges. The broom of order n is a tree obtained from a P_k by joining one end vertex with $n-k$ additional vertices by edges. We conjecture that there exists an integer n_0 such that for all $n \geq n_0$, among all trees and graphs of order n, the Wiener-entropy is minimized by respectively a broom and a $G_{n,k,1}$. Based on a computer-aided verification among graphs of the form $G_{n,k,j}$, it seems plausible that $n_0 = 1270$.

Apart from the open problems in the thesis, there are still many other problems to be explored and solved involving graph entropies. In this sense, the results of this thesis are only the tip of the iceberg.

Samenvatting

Dit proefschrift richt zich op extremaalproblemen betreffende diverse graafparameters die binnen de grafentheorie geïntroduceerd zijn gemotiveerd door de bekende Shannon-entropie uit de informatietheorie. In het bijzonder worden er nieuwe resultaten gepresenteerd op het gebied van graadgebaseerde en afstandsgebaseerde entropieën. Bovendien worden er resultaten afgeleid met betrekking tot de complexiteit van beslissings– en optimaliseringsproblemen betreffende maximale en minimale opspannende bomen voor graafgerelateerde functie-indices.

Het bepalen van extreme waarden voor topologische indices en het karakteriseren van de bijbehorende grafen is een populair onderwerp binnen de grafentheorie. In Hoofdstuk 1 van dit proefschrift geven we een algemene inleiding tot dit gebied, tezamen met wat historische achtergrond en een overzicht van de bijdragen uit dit proefschrift. Eén van de benaderingen is om te onderzoeken wat het effect is van bepaalde graafoperaties op de waarde van zulke indices. Deze aanpak is vooral nuttig als de extremale grafen kunnen worden bepaald door een serie van die graafoperaties. In Hoofdstuk 2 van dit proefschrift beschouwen we het effect van bepaalde graafoperaties op de waarde van de graadgebaseerde entropie.

In Hoofdstuk 3 tot en met Hoofdstuk 5 bestuderen we extremaalproblemen met betrekking tot de graadgebaseerde entropie voor grafen uit specifieke graafklassen. In Hoofdstuk 3 gaat het daarbij om bomen en unicyclische grafen. Voor die klassen worden de maximale en minimale waarde van de graadgebaseerde entropie bepaald, op basis van specifieke graafparameters zoals de diameter en een gegeven bipartitie van de graaf. Deze resultaten worden met name afgeleid door te analyseren wat het effect is van graafoperaties op de graadgebaseerde entropie.

Hoofdstuk 4 richt zich op de algemene klasse van bipartiete grafen, daarbij gebruikmakend van de representatie van bipartiete grafen door middel van zogenoemde 'Young-Diagramm'. We bewijzen dat de extremale grafen die de minimale waarde aannemen ofwel volledig bipartiet zijn, ofwel bijna volledig bipartiet, in een bepaalde zin. Bovendien laten we zien dat het algemene probleem van het karakteriseren van de extremale grafen gerelateerd is aan een zeer ingewikkeld probleem uit de getaltheorie. Naar aanleiding daarvan geloven we dat dit probleem moeilijk op te lossen is. Binnen de klasse van bipartiete grafen met een gegeven aantal punten en lijnen karakteriseren we de extremale grafen die de maximale graadgebaseerde entropie bereiken, door gebruik te maken van hun graadrij. Tevens breiden we deze resultaten uit tot meer algemene graaf-indices. Hiervoor wordt bewezen dat de maximale en minimale graden in de graadrij van de extremale grafen hooguit 2 kunnen verschillen.

In Hoofdstuk 5 wordt een volledige karakterisering gegeven van de extremale grafen met een gegeven aantal punten en lijnen die de minimale waarde van de graadgebaseerde entropie aannemen. Deze extremale grafen blijken te vallen binnen een bekende klasse van grafen die in het Engels worden aangeduid als 'threshold graphs'.

Hoofdstuk 6 bevat resultaten op het gebied van twee afstandsgebaseerde entropieën, namelijk de zogenoemde 'eccentricity-entropy', aangeduid met I_{ecc}, en de zogenoemde 'Wiener-entropy', aangeduid met I_w. Door naar het (asymptotische) extremale gedrag van die entropieën te kijken, concluderen we dat de spreiding van de waarden van I_w groter is dan die van I_{ecc}. We ontraadselen drie bekende vermoedens betreffende I_{ecc} en poneren twee nieuwe vermoedens betreffende I_w die we tegen het eind van deze samenvatting formuleren.

Hoofdstuk 7 gaat in op de rekenkundige complexiteit van beslissings– en optimaliseringsproblemen die verband houden met het bepalen van de maximale en minimale opspannende bomen met betrekking tot algemene graafgerelateerde functie-indices. Binnen de klasse van bomen blijken veel van die indices hun extreme waarden te bereiken in het unieke geval dat de boom een pad is. Voor indices met deze laatste eigenschap laten we zien dat de bijbehorende beslissingsproblemen $\mathcal{NP}$-volledig zijn. Tevens bewijzen we dat strikt convexe of strikt concave functies leiden tot $\mathcal{NP}$-volledige beslissingsproblemen en $\mathcal{APX}$-volledige optimaliseringsproblemen voor de corresponderende functie-indices.

Hoewel dit proefschrift een flink aantal nieuwe resultaten bevat, blijven er verscheidene onbeantwoorde vragen en open vermoedens die om een oplossing vragen. Eén van de uitdagende open problemen is het volledig karakteriseren van de bipartiete grafen met een gegeven aantal punten en lijnen die de kleinste graadgebaseerde entropie hebben.

Met betrekking tot de 'Wiener-entropy' poneren we de volgende twee vermoedens (in één formulering). Daartoe definiëren we eerst de graaf $G_{n,k,j}$ en de 'broom' met $n > k \geq 2$ punten, als volgt. Begin voor de constructie van $G_{n,k,j}$ met de disjuncte vereniging van een pad P_k en een volledige graaf K_{n-k}, en verbind één van de eindpunten van het pad met j punten van de volledige graaf door middel van lijnen. De 'broom' op n punten is een boom die uit een P_k ontstaat door één van de eindpunten te verbinden met $n-k$ nieuw toegevoegde punten door middel van lijnen. We vermoeden dat er een getal n_0 bestaat zodanig dat voor alle $n \geq n_0$ de 'Wiener-entropy' binnen de klasse van bomen, respectievelijk grafen met n punten de kleinste waarde aanneemt voor de 'broom', respectievelijk voor $G_{n,k,1}$. Uit computerondersteunde verificatie lijkt het aannemelijk dat dit geldt voor $n_0 = 1270$.

Naast de open problemen die betrekking hebben op dit proefschrift zijn er veel andere problemen op het gebied van graafentropieën die onderzocht kunnen worden. In die zin vormen de resultaten uit dit proefschrift slechts het topje van de ijsberg.

Bibliography

[1] A. Abiad, B. Brimkov, A. Erey, L. Leshock, X. Martínez-Rivera, S. O, S.-Y. Song and J. Williford, On the Wiener index, distance cospectrality and transmission-regular graphs, *Discrete Appl. Math.*, **230** (2017) 1–10.

[2] P. Alimonti and V. Kann, Some APX-completeness results for cubic graphs, *Theoret. Comput. Sci.*, **237** (2000) 123–134.

[3] E.B. Allen, Measuring graph abstractions of software: An information-theory approach, *in: Proceedings of the Eighth IEEE Symposium on Software Metrics, IEEE, Ottawa*, 2002, pp. 182–193.

[4] N. Alon and A. Orlitsky, Source coding and graph entropies, *IEEE Trans. Inform. Theory*, **42** (1996) 1329–1339.

[5] B. Alspach, The wonderful Walecki construction, *Bull. Inst. Combin. Appl.*, **52** (2008) 7–20.

[6] B. Alspach, J.-C. Bermond and D. Sotteau, Decomposition into Cycles I: Hamilton Decompositions, *in: Workshop on Cycles and Rays, Springer, Montreal*, 1987, pp. 9–18.

[7] G. Altay and F. Emmert-Streib, Revealing differences in gene network inference algorithms on the network level by ensemble methods, *Bioinformatics*, **26** (2010) 1738–1744.

[8] K. Anand and G. Bianconi, Entropy measures for networks: Toward an information theory of complex topologies, *Phys. Rev. E*, **80** (2009) 045102(R).

[9] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy, Proof verification and the hardness of approximation problems, *J. AMC*, **45** (1998) 501–555.

[10] M. Azari, Eccentric connectivity coindex under graph operations, *J. Appl. Math. Comput.,* **62** (2020) 23–35.

[11] R. Badii and A. Politi, Complexity: Hierarchical Structures and Scaling in Physics, *Cambridge University Press, London* (1997).

[12] K. Balakrishnan, M. Changat, I. Peterin, S.Špacapan, P. Šparl and A. R. Subhamathi, Strongly distance-balanced graphs and graph products, *European J. Combin.*, **30** (2009) 1048–1053.

[13] T. Balch, Hierarchic social entropy: An information theoretic measure of robot group diversity, *Auton. Robots*, **8** (2000) 209–237.

[14] L. Boltzmann, Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht, *Wien. Ber.,* **76** (1877) 373–435.

[15] D. Bonchev, Information indices for atoms and molecules, *Match*, **7** (1979) 65–112.

[16] D. Bonchev, Information Theoretic Indices for Characterization of Chemical Structures, *Research Studies Press, Herts* (1983).

[17] D. Bonchev, Kolmogorov's information, Shannon's entropy, and topological complexity of molecules, *Bulg. Chem. Commun.*, **28** (1995) 567–582.

[18] D. Bonchev, A.T. Balaban and O. Mekenyan, Generalization of the graph center concept, and derived topological centric indexes, *J. Chem. Inf. Comput. Sci.*, **20** (1980) 106–113.

[19] D. Bonchev and D.H. Rouvray, Complexity in Chemistry, Biology, and Ecology, *Mathematical and Computational Chemistry, Springer, New York* (2005).

[20] D. Bonchev and D.H. Rouvray, Complexity in Chemistry. Introduction and Fundamentals, *Taylor & Francis, London* (2003).

[21] D. Bonchev and N. Trinajstić, Information theory, distance matrix, and molecular branching, *J. Chem. Phys.*, **67** (1977) 4517–4533.

[22] J.A. Bondy and U.S.R. Murty, Graph Theory, Graduate Texts in Mathematics, Volume 244, *Springer, New York* (2008).

[23] P. Bonsma, Max-leaves spanning tree is APX-hard for cubic graphs, *J. Discrete Algorithms*, **12** (2012) 14–23.

[24] S.L. Braunstein, S. Ghosh and S. Severini, The Laplacian of a graph as a density matrix: A basic combinatorial approach to separability of mixed states, *Ann. Combinat.*, **10** (2006) 291–317.

[25] L. Brillouin, Science and Information Theory, Second Edition, *Academic Press Inc., New York* (1962).

[26] F.P. Brooks, Three great challenges for half-century-old computer science, *J. ACM*, **50** (2003) 25–26.

[27] F. Buckley, Self-centered graphs, *in: Graph theory and its applications: East and West, Ann. New York Acad. Sci., Jinan*, 1986, pp. 71–78.

[28] S. Cambie, Five results on maximizing topological indices in graphs, *Discrete Math. Theor. Comput. Sci.,* **23** (2021) #10.

[29] S. Cao, M. Dehmer and Y. Shi, Extremality of degree-based graph entropies, *Inform. Sci.,* **278** (2014) 22–33.

[30] G. Caporossi, I. Gutman, P. Hansen and L. Pavlović, Graphs with maximum connectivity index, *Comput. Biol. Chem.,* **27** (2003) 85–90.

[31] S. Carnot, Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance, *Chez Bachelier, Paris* (1824)

[32] T.C.E. Cheng, Y. Guo, S. Zhang and Y. Du, Extreme values of the sum of squares of degrees of bipartite graphs, *Discrete Math.* **309** (2009) 1557–1564.

[33] V. Chvátal and P.L. Hammer, Aggregation of inequalities in integer programming, *Ann. of Discrete Math.*, **1** (1977) 145–162.

[34] R. Clausius, On a modified form of the second fundamental theorem in the mechanical theory of heat, *Phil. Mag S.*, **12** (1856) 81–98.

[35] R. Clausius, The Mechanical Theory of Heat, *Macmillan, London* (1879).

[36] R. Clausius, Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen, *Ann. Phys.*, **155** (1850) 368–397.

[37] T.H. Cormen, C.E. Leiserson, R.L. Rivest and C. Stein, Introduction to Algorithms, Third Edition, *The MIT Press, London* (2009).

[38] P. Crescenzi, A short guide to approximation preserving reductions, *in: Twelfth Annual IEEE Conference on Computational Complexity, IEEE Computer Soc., Los Alamitos*, 1997, pp. 262–273.

[39] I. Csiszár, J. Körner, L. Lovász, K. Marton and G. Simonyi, Entropy splitting for antiblocking corners and perfect graphs, *Combinatorica*, **10** (1990) 27–40.

[40] K.C. Das and S.-G. Lee, On the Estrada index conjecture, *Linear Algebra Appl.*, **431** (2009) 1351–1359.

[41] M. Dehmer, A novel method for measuring the structural information content of networks, *Cybern. Syst.*, **39** (2008) 825–842.

[42] M. Dehmer, Information processing in complex networks: Graph entropy and information functionals, *Appl. Math. Comput.*, **201** (2008) 82–94.

[43] M. Dehmer, Information-theoretic concepts for the analysis of complex networks, *Appl. Artif. Intell.*, **22** (2008) 684–706.

[44] M. Dehmer, Z. Chen, A. Mowshowitz, H. Jodlbauer, F. Emmert-Streib, Y. Shi, S. Tripathi and C. Xia, On the degeneracy of the Randić entropy and related graph measures, *Inform. Sci.*, **501** (2019) 680–687.

[45] M. Dehmer and F. Emmert-Streib, Towards network complexity, *in: Complex Sciences, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, Shanghai*, 2009, pp. 707–714.

[46] M. Dehmer, F. Emmert-Streib, Z. Chen, X. Li and Y. Shi, Mathematical Foundations and Applications of Graph Entropy, *Wiley, Weinheim* (2016).

[47] M. Dehmer and V. Kraus, On extremal properties of graph entropies, *MATCH Commun. Math. Comput. Chem.*, **68** (2012) 889–912.

[48] M. Dehmer and A. Mowshowitz, A history of graph entropy measures, *Inform. Sci.*, **181** (2011) 57–78.

[49] M. Dehmer and A. Mowshowitz, Inequalities for entropy-based measures of network information content, *Appl. Math. Comput.*, **215** (2010) 4263–4271.

[50] M. Dehmer, K. Varmuza, S. Borgert and F. Emmert-Streib, On entropy-based molecular descriptors: Statistical analysis of real and synthetic chemical structures, *J. Chem. Inf. Model.*, **49** (2009) 1655–1663.

[51] H. Deng, On the Balaban index of trees, *MATCH Commun. Math. Comput. Chem.*, **66** (2011) 253–260.

[52] H. Dong and X. Guo, Character of graphs with extremal Balaban index, *MATCH Commun. Math. Comput. Chem.*, **63** (2010) 799–812.

[53] Y. Dong, S. Qiao, B. Chen, P. Wan and S. Zhang, Maximum values of degree-based entropies of bipartite graphs, *Appl. Math. Comput.*, **401** (2021) 12094.

[54] Z. Du, Wiener indices of trees and monocyclic graphs with given bipartition, *Int. J. Quantum Chem.,* **112** (2012) 1598–1605.

[55] F. Emmert-Streib and M. Dehmer, Global information processing in gene networks: Fault tolerance, *in: 2007 2nd Bio-Inspired Models of Network, Information and Computing Systems, IEEE, Budapest* 2007, pp. 326–329.

[56] F. Emmert-Streib and M. Dehmer, Robustness in scale-free networks: Comparing directed and undirected networks, *Int. J. Mod. Phys. C,* **19** (2008) 717–726.

[57] F. Emmert-Streib and M. Dehmer, Fault tolerance of information processing in gene networks, *Physica A,* **338** (2009) 541–548.

[58] R.C. Entringer, D.E. Jackson and D.A. Snyder, Distance in graphs, *Czechoslovak Math. J.*, **26** (1976) 283–296.

[59] L. Euler, Solutio problematis ad geometriam situs pertinentis, *Commentarii Academiae Scientiarum Imperialis Petropolitanae,* **8** (1736) 128–140.

[60] M. Fredman and J. Komlós, On the size of separating systems and perfect hash functions, *SIAM J. Algebraic Disc. Meth.*, **5** (1984) 61–68.

[61] J.I. Fujii and S. Yuki, Entropy and coding for graphs, *Int. J. Math. Stat. Sci.*, **6** (2008) 63–77.

[62] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, *A Series of Books in the Mathematical Sciences, Macmillan, San Francisco* (1979).

[63] A. Ghalavand, M. Eliasi and A.R. Ashrafi, First degree-based entropy of graphs, *J. Appl. Math. Comput.,* **59** (2019) 37–46.

[64] E. Goudout, Lois locales de la fonction ω dans presque tous les petits intervalles, *Proc. Lond. Math. Soc.,* **115** (2017) 599–637.

[65] I. Gutman, Degree-based topological indices, *Croat. Chem. Acta,* **86** (2013) 351–361.

[66] I. Gutman, W. Linert, I. Lukovits and A.A. Dobrynin, Trees with extremal hyper-Wiener index: Mathematical basis and chemical applications, *J. Chem. Inf. Comput. Sci.*, **37** (1997) 349–354.

[67] P. L. Hammer, U. N. Peled and X. Sun, Difference graphs, *Discrete Appl. Math.,* **28** (1990) 35–44.

[68] K. Handa, Bipartite graphs with balanced (a, b)-partitions, *Ars Combin,* **51** (1999) 113–119.

[69] Y. Hu, X. Li and Y. Yuan, Trees with maximum general Randić index, *Commun. Math. Comput. Chem.,* **52** (2004) 129–146.

[70] Y. Hu, X. Li and Y. Yuan, Trees with minimum general Randić index, *Commun. Math. Comput. Chem.,* **52** (2004) 119–128.

[71] Z. Hu, X. Li and D. Peng, Graphs with minimum vertex-degree function-index for convex functions, *MATCH Commun. Math. Comput. Chem.,* **88** (2022) 521–533.

[72] A. Ilić, On the extremal values of general degree-based graph entropies, *Inform. Sci.,* **370–371** (2016) 424–427.

[73] D.S. Johnson, Approximation algorithms for combinatorial problems, *J. Comput. System Sci.,* **9** (1974) 256–278.

[74] J. Kahn, An entropy approach to the hard-core model on bipartite graphs, *Combin. Probab. Comput.*, **10** (2001) 219–237.

[75] J. Karamata, Sur une inégalité rélative aux fonctions convexes, *Publ. Math. Univ. Belgrade,* **1** (1932) 145–148.

[76] M.H. Khalifeh, H. Yousefi-Azari and A.R. Ashrafi, The first and second Zagreb indices of some graph operations, *Discrete Appl. Math.,* **4** (2009) 804–811.

[77] M.H. Khalifeh, H. Yousefi-Azari and A.R. Ashrafi, The hyper-Wiener index of graph operations, *Comput. Math. Appl.,* **56** (2008) 1402–1407.

[78] E.V. Konstantinova and A.A. Paleev, Sensitivity of topological indices of polycyclic graphs, *Vychisl. Sistemy.*, **136** (1990) 38–48.

[79] J. Körner, Coding of an information source having ambiguous alphabet and the entropy of graphs, *in: Transactions of the Sixth Prague Conference on Information Theory, Springer, Prague,* 1973, pp. 411–426.

[80] J. Körner, Fredman-Komlós bounds and information theory, *SIAM J. Algebraic Discrete Methods,* **7** (1986) 560–570.

[81] J. Körner and K. Marton, Graphs that split entropies, *SIAM J. Discrete Math.*, **1** (1988) 71–79.

[82] J. Körner and K. Marton, New bounds for perfect hashing via information theory, *European J. Combin.*, **9** (1988) 523–530.

[83] J. Körner, G. Simonyi and Z. Tuza, Perfect couples of graphs, *Combinatorica*, **12** (1992) 179–192.

[84] V. Kraus, M. Dehmer and M. Schutte, On sphere-regular graphs and the extremality of information-theoretic network measures, *MATCH Commun. Math. Comput. Chem.*, **70** (2013) 885–900.

[85] R. Lang, X. Li and S. Zhang, Inverse problem for the Zagreb index of molecular graphs, *Appl. Math. J. Chinese Univ. Ser. A*, **18** (2003) 487–493.

[86] P. Lemke, The Maximum-Leaf Spanning Tree Problem in Cubic Graphs is NP-complete, *IMA Preprinter Series #428, University of Minnesota, Minneapolis* (1988).

[87] A. Li and Y. Pan, Structural information and dynamical complexity of networks, *IEEE Trans. Inform. Theory*, **62** (2016) 3290–3339.

[88] A. Li, X, Yin, B. Xu, D. Wang, J. Han, Y. Wei, Y. Deng, Y. Xiong and Z. Zhang, Decoding topologically associating domains with ultra-low resolution Hi-C data by graph structural entropy, *Nat. Commun.*, **9** (2018) 3265.

[89] M. Li and P. Vitaanyi, An Introduction to Kolmogorov Complexity and Its Applications, *Springer, New York* (1997).

[90] X. Li and D. Peng, Extremal problems for graphical function-indices and f-weighted adjacency matrix, *Discrete Math. Lett.*, **9** (2022) 57–66.

[91] X. Li, Y Shi and I. Gutman, Graph Energy, *Springer, New York* (2012).

[92] X. Li and J. Zheng, A unified approach to the extremal trees for different indices, *MATCH Commun. Math. Comput. Chem.,* **50** (2005) 195–208.

[93] L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory*, **IT-25** (1979) 1–7.

[94] N.V.R. Mahadev and U. N. Peled, Threshold Graphs and Related Topics, Annals of Discrete Mathematics, *Elsevier, Amsterdam* (1995).

[95] K. Matomäki and M. Radziwiłł, Multiplicative functions in short intervals, *Ann. of Math.*, **183** (2016) 1015–1056.

[96] A. Mehler, M. Dehmer and R. Gleim, Towards logical hypertext structure. A graph-theoretic perspective, *in: Proceedings of the Fourth International Workshop on Innovative Internet Computing Systems, Springer, Guadalajara*, 2004, pp. 136–150.

[97] A. Mowshowitz, Entropy and the complexity of graphs: I. An index of the relative complexity of a graph, *Bull. Math. Biophys.*, **30** (1968) 175–204.

[98] A. Mowshowitz, Entropy and the complexity of graphs: II. The information content of digraphs and infinite graphs, *Bull. Math. Biophys.*, **30** (1968) 225–240.

[99] A. Mowshowitz, Entropy and the complexity of graphs: III. Graphs with prescribed information content, *Bull. Math. Biophys.*, **30** (1968) 387–414.

[100] A. Mowshowitz, Entropy and the complexity of graphs: IV. Entropy measures and graphical structure, *Bull. Math. Biophys.*, **30** (1968) 533–546.

[101] I. Newman and A. Wigderson, Lower bounds on formula size of Boolean functions using hypergraph entropy, *SIAM J. Discrete Math.*, **30** (1995) 536–542.

[102] C.H. Papadimitriou and M. Yannakakis, Optimization, approximation, and complexity classes, *J. Comput. System Sci.*, **43** (1991) 425–440.

[103] M. Planck, Uber das Gesetz der Energieverteilung im Normalspektrum, *Annalen der Physik.*, **309** (1901) 553–563.

[104] N. Rashevsky, Life, information theory, and topology, *Bull. Math. Biophys.*, **17** (1955) 229–235.

[105] A. Reich, Complexity of the maximum leaf spanning tree problem on planar and regular graphs, *Theoret. Comput. Sci.*, **626** (2016) 134–143.

[106] E. Schrödinger, What Is Life? The Physical Aspect of the Living Cell with Mind and Matter & Autobiographical Sketches, *Cambridge University Press, New York* (1992).

[107] F. Schweitzer, W. Ebeling, H. Rosé and O. Weiss, Network optimization using evolutionary strategies, *in: Parallel Problem Solving from Nature - PPSN IV Lecture Notes in Computer Science, Springer, Berlin,* 1996, pp. 940–949.

[108] C.E. Shannon, A mathematical theory of communication, *Bell System Tech. J.*, **27** (1948) 379–423, 623–656.

[109] C.E. Shannon, The zero error capacity of a noisy channel, *IRE Trans. Inform. Theory*, **2** (1956) 8–19.

[110] G. Simonyi, Graph entropy: A survey, *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.,* **20** (1995) 1–38.

[111] G. Simonyi, Perfect graphs and graph entropy. An updated survey, *in: Perfect Graphs, Wiley, Chichester,* 2001, pp. 293–328.

[112] R.V. Solé and S. Valverde, Information theory of complex networks: On evolution and architectural constraints, *in: Lecture Notes in Physics, Springer, Berlin, Heidelberg,* 2004, pp. 189–207.

[113] R. Todeschini, V. Consonni and R. Mannhold, Handbook of Molecular Descriptors, *Wiley, Weinheim* (2002).

[114] I. Tomescu, Extremal vertex-degree function index for trees and unicyclic graphs with given independence number, *Discrete Appl. Math.,* **306** (2022) 83–88.

[115] E. Trucco, A note on the information content of graphs, *Bull. Math. Biophys.,* **18** (1956) 129–135.

[116] D. Vukičević and B. Furtula, Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges, *J. Math. Chem.,* **46** (2009) 1369–1376.

[117] S. Wagner and I. Gutman, Maxima and minima of the Hosoya index and the Merrifield-Simmons index: A survey of results and techniques, *Acta Appl. Math.,* **112** (2010) 323–346.

[118] P. Wan, X. Chen, J. Tu, M. Dehmer, S. Zhang and F. Emmert-Streib, On graph entropy measures based on the number of independent sets and matchings, *Inform. Sci.,* **516** (2020) 491–504.

[119] P. Wan, J. Tu, M. Dehmer, S. Zhang and F. Emmert-Streib, Graph entropy based on the number of spanning forests of c-cyclic graphs, *Appl. Math. Comput.,* **363** (2019) 124616.

[120] P. Wan, J. Tu, S. Zhang and B. Li, Computing the numbers of independent sets and matchings of all sizes for graphs with bounded treewidth, *Appl. Math. Comput.,* **332** (2018) 42–47.

[121] N. Wiener, Cybernetics: Or Control and Communication in the Animal and the Machine, *The MIT Press, London* (1961).

[122] H. Wiener, Structural determination of paraffin boiling points, *Journal of the American Chemical Society*, **6** (1947) 17–20.

[123] H.W. Xiang, The Corresponding-States Principle and Its Practice: Thermodynamic, Transport and Surface Properties of Fluids, *Elsevier, Amsterdam* (2005).

[124] K. Xu, K. Tang, H. Liu and J. Wang, The Zagreb indices of bipartite graphs with more edges, *J. Appl. Math. Inform.,* **33** (2015) 365–377.

[125] J. Yan, Topological structure of extremal graphs on the first degree-based graph entropies, *MATCH Commun. Math. Comput. Chem.,* **85** (2021) 275–284.

[126] Y. Yao, M. Liu, F. Belardo and C. Yang, Unified extremal results of topological indices and spectral invariants of graphs, *Discrete Appl. Math.,* **271** (2019) 218–232.

[127] H. Zhang and S. Li, On the (reverse) cover cost of trees with some give parameters, *Discrete Math.,* **344** (2021) 112226.

[128] S. Zhang and C. Zhou, Bipartite graphs with the maximum sum of squares of degrees, *Acta Math. Appl. Sin. Engl. Ser.,* **30** (2014) 801–806.

[129] Y. Zhao, The number of independent sets in a regular graph, *Combin. Probab. Comput.,* **19** (2010) 315–320.

[130] L. Zhong, The harmonic index for graphs, *Appl. Math. Lett.,* **25** (2013) 562–566.

Acknowledgements

The initial dream poured sweat into my hard work, and my sweat-soaked body was so exhausted that I could not sleep all night. The power of love kept pushing me forward. Whenever I stop and think about the road I have traveled, whether I have missed the scenery and forgotten you, I hope you can forgive me for needing to regain my memory. But nothing affects my gratitude to you.

I would like to thank my supervisor Professor Shenggui Zhang at Northwestern Polytechnical University. Your insightful feedback motivates me to improve my thinking and take my work to the next level. In particular, I would like to thank you for all the opportunities you have provided for my research.

I am incredibly grateful to my supervisor, Professor Hajo Broersma, for giving me the opportunity to study and work in this warm group at the University of Twente. Your expertise is invaluable in formulating research writing and methods. Thank you for guiding me when I lost my way.

I would like to convey my heartfelt gratitude towards my supervisor Doctor Maximilien Gadouleau at Durham University. I have had invaluable discussions and your patient support and encouragement in my research.

I especially want to thank the head of FMT group Professor Marieke Huisman in the University of Twente. Thank you for giving me hugs and strength when I was down in my research. I would like to thank the management assistant of FMT, Marion Steenbergen-Boeringa, for your meticulous care and devotion to us from scientific research to life.

I am immensely grateful to my colleagues whose dedication and positive work ethic have continuously inspired and motivated me. I would like to express special appreciation to my friends across the globe, whose presence has provided a much-needed respite from my research endeavors.

In addition, I extend heartfelt thanks to my parents for the invaluable guidance and unwavering support. Your wise counsel and compassionate listening have been a constant source of strength, propelling me forward on my research journey.

Finally, I would not have been able to complete this thesis without the support of my coauthors and referees who provided valuable suggestions for improving my papers and thesis. I would like to give thanks to all those who provided comments and suggestions for my papers and this thesis. I also want to give thanks to the library staff for helping me with searching and purchasing literature.

Although the regained memory has no end, I have to put an end to my acknowledgements. Please forgive me if you read this and realize I have forgotten who you are to be grateful for. Maybe it is because I have sealed this gratitude to you in my heart, or maybe it is because I cannot express my gratitude to you in words, maybe it is because you have faded out of my memory. So please allow me to thank everything with a grateful heart!

Yanni Dong
June 2023, Enschede

About the Author

Yanni Dong was born in Lanzhou City, Gansu Province, P.R. China. From 1999 to 2005, Yanni Dong completed six years elementary education in Primary School attached to Lanzhou University of Technology. From 2005 to 2008, she completed her three years of junior middle school in Junior middle school attached to Lanzhou University of Technology. She spent three years as a senior middle school student in Lanlian No.1 Middle School. After that she started to study at the School of Mathematics, Lanzhou City University in September 2011. After obtaining her Bachelor degree in June 2015, she spent three years as a graduated student in the School of Science at Lanzhou University of Technology.

In September 2018, she became a Ph.D. student directly under the supervision of Professor Shenggui Zhang at Northwestern Polytechnical University. From October 2019 to June 2020, she visited the Department of Computer Science, Durham University as a visiting student. Starting from June 2021, she became a member of the group of Formal Methods and Tools, University of Twente as a joint Ph.D. student to perform research on graph entropy under the supervision of Professor Hajo Broersma. The research has been sponsored by the China Scholarship Council. The main results obtained from her research during her Ph.D. work have been collected in the current thesis.