

# Cliques in geometric inhomogeneous random graphs

Riccardo Michielan<sup>1</sup> and Clara Stegehuis<sup>1</sup>

<sup>1</sup>University of Twente - Faculty of Electrical Engineering, Mathematics and Computer Science

June 4, 2021

## Abstract

Many real-world networks were found to be highly clustered, and contain a large amount of small cliques. We here investigate the number of cliques of any size  $k$  contained in a geometric inhomogeneous random graph: a scale-free network model containing geometry. The interplay between scale-freeness and geometry ensures that connections are likely to form between either high-degree vertices, or between close by vertices. At the same time it is rare for a vertex to have a high degree, and most vertices are not close to one another. This trade-off makes cliques more likely to appear between specific vertices. In this paper, we formalize this trade-off and prove that there exists a predominant type of clique in terms of the degrees and the positions of the vertices that span the clique. Moreover, we show that the asymptotic number of cliques as well as the predominant clique type undergoes a phase transition, in which only  $k$  and the degree-exponent  $\tau$  are involved. Interestingly, this phase transition shows that for small values of  $\tau$ , the underlying geometry of the model is irrelevant: the number of cliques scales the same as in a non-geometric network model.

## 1 Introduction

Real-world networks often share common characteristics. For example, many large real-world networks are *scale-free*, that is, there exist a small number of individuals with a large amount of connections, whereas most of the individuals only have a small connectivity. This feature is mathematically often described by assuming that the degrees of the vertices in the network follow a power-law distribution. Another common feature is that real-world networks typically have large clustering coefficient. That is, there is a high probability that two neighbors of the same vertex are connected, and thus the network contains many triangles. This structural property is highly related to a possible underlying geometry of the network. Indeed, when any two close individuals are more likely to connect, the triangle inequality ensures that triangles are more likely to form between close groups of three vertices. In fact, in two examples of network models with an underlying geometry it has been proven that the presence of this underlying network geometry guarantees a high clustering coefficient, for hyperbolic random graphs [10], and geometric inhomogeneous random graphs [6].

While clustering is typically measured in terms of the number of triangles in the network, the presence of larger cliques inside a network is also informative on the amount of network clustering. Indeed, in clustered networks, one would expect the number of cliques of size larger than three to be high as well. Therefore, the number of cliques of general sizes have been extensively studied in

several types of random graph models without underlying geometry, such as dense inhomogeneous random graphs [8], scale free inhomogeneous random graphs [13, 2], general rank-1 inhomogeneous random graphs [5]. Results on the number of cliques in random graphs with underlying geometry are less well-studied, as the presence of geometry creates correlations between the presence of different edges, making it difficult to compute the probability that a given clique is present. Still, some results are known for high-dimensional geometric random graphs [1] and hyperbolic random graphs [4], showing that these types of random graphs typically contain a larger number of cliques than non-geometric models as long as the dimension of the underlying space is not too large. Particular attention has been given to the clique number: the largest clique in the network [4, 12, 7].

In this paper, we study general clique structures that can indicate network clustering inside a more general geometric network model, and investigate the relation between the presence of geometry and the presence of cliques. In particular, we analyze the Geometric Inhomogeneous random graph (GIRG) [6], a random graph model that includes scale-free vertex weights describing (roughly) the vertex degrees, and an underlying geometric space that makes nearby vertices more likely to be connected. We analyze the number of  $k$ -cliques contained in this random graph model, by deriving and solving an optimization method, similarly to [11]. This optimization method allows to overcome the difficulties posed by the dependence of the presence of edges in geometric models by studying the connections between different regions separately. Interestingly, we show that  $k$ -cliques typically appear in specified regions of the network. We describe the specific properties satisfied by these most predominant cliques in terms of the vertex degrees and their geometric positions. Interestingly, our results show that geometry plays a central role in the number of cliques of any size when the degree-exponent of the power-law is at least  $7/3$ . For smaller degree-exponents however, we show that the predominant clique does not depend on the underlying geometry, and that the number of cliques of all sizes scales the same as in scale-free configuration models [11], random graph models without any form of geometry. Thus, our results show that geometry does not always cause larger clique numbers or clustering coefficients in the scale-free regime, even when the dimension of the geometric space is low.

Specifically, we find that for any clique size, there is a threshold degree-exponent such that for scale-free networks with degree-exponent below the threshold, geometry does not influence the clique counts. When the degree exponent is above the threshold, the clique counts of that size are influenced by the underlying geometry of the model. Furthermore, for larger cliques, this threshold degree-exponent becomes higher, so that the geometry of the model becomes irrelevant for a larger range of the degree-exponent. Therefore, for small cliques, the range in which optimal subgraph structures depend significantly on the geometric features of the vertices is larger.

**Organization of the paper.** In Section 2 we describe our main result, concerning the characterization of the optimal clique structures in the GIRG and their scaling. Moreover, we provide simulations in support of our result, and we provide a short discussion. In Section 3 we formulate the problem we are analyzing as an optimization problem, where the feasible region is formed by the pairs of vectors  $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ , which express the properties (weights and distances) of the vertices involved in a clique. Then in Section 4 we prove the solution structure of the optimization problem, and finally we provide the proof of the main theorem in Section 5.

**Notation.** We now describe the notation that will be used throughout this paper. A  $k$ -clique is a subset of vertices of size  $k$  such that they are all pairwise connected, and it is denoted by  $\mathcal{K}_k$ . In

this paper, we analyze cliques contained in GIRGs where the number of vertices  $n$  tends to  $\infty$ . We say that a sequence of events  $(\mathcal{E}_n)_{n \geq 1}$  happens with high probability (w.h.p.) of  $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E}_n) = 1$ . We denote with  $\mathbb{P}_n(\mathcal{E})$  the probability that the event  $\mathcal{E}$  happens in a GIRG with  $n$  vertices, and with  $\mathbb{E}_n[X]$  the mean value of the random variable  $X$  when the number of vertices is  $n$ . We write

- $f(n) = o(g(n))$  if  $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$
- $f(n) = O(g(n))$  if  $\limsup_{n \rightarrow \infty} |f(n)|/g(n) < \infty$
- $f(n) = \Theta(g(n))$  if  $f(n) = O(g(n))$  as well as  $g(n) = O(f(n))$

Moreover, we write

- $X_n = o_{\mathbb{P}}(a(n))$  if  $\lim_{n \rightarrow \infty} \mathbb{P}(\frac{X_n}{a(n)} \geq \varepsilon) = 0$  for all  $\varepsilon > 0$
- $X_n = O_{\mathbb{P}}(a(n))$  if for any  $\varepsilon > 0$  there exist  $M > 0$  and  $N \in \mathbb{N}$  such that  $\mathbb{P}(|X_n|/a(n) \geq M) < \varepsilon$  for all  $n > N$ .
- $X_n = \Theta_{\mathbb{P}}(a(n))$  if for any  $\varepsilon > 0$  there exist  $m, M > 0$  and  $N \in \mathbb{N}$  such that  $\mathbb{P}(|X_n|/a(n) \notin [m, M]) < \varepsilon$  for all  $n > N$ .

**Geometric Inhomogeneous random graph.** We now define the geometric inhomogeneous random graph [6]. Let  $n \in \mathbb{N}$  denote the number of vertices in the graph, and call  $V = [n] = \{1, 2, \dots, n\}$  the set of vertices of the GIRG. In the GIRG, each vertex  $i$  is associated with a weight,  $w_i$  and a position  $x_i$ . The weights  $w_1, \dots, w_n$  are independent and identically distributed random variables that follow a Pareto power law distribution with exponent  $\tau \in (2, 3)$ . That is, for any  $v \in V$

$$\mathbb{P}(w_v > w) = cw^{-(\tau-1)}, \quad (1)$$

for  $w \geq w_{\min}$ , for some  $w_{\min} \geq 0$ , where  $c$  is a normalization constant. We denote by  $W := \sum_{v=1}^n w_v$  the sum of the weights.

As ground space for positions of the vertices, we consider the  $d$ -dimensional torus  $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ . Then, the positions  $x_1, \dots, x_n$  are independent and identically distributed random variables, with uniform distribution on  $\mathbb{T}^d$ . That is, for any  $v \in V$

$$\mathbb{P}(x_v \in [a_1, b_1] \times \dots \times [a_d, b_d]) = \prod_{i=1}^d (b_i - a_i) \quad (2)$$

for any  $a_1, \dots, a_d, b_1, \dots, b_d$  in  $[0, 1]$  such that  $a_i \leq b_i$  for every  $i = 1, \dots, d$ .

We denote by  $E$  the set of edges in the GIRG. An edge between any two vertices  $u, v \in V$  of the GIRG appears with a probability  $p_{uv}$  determined by the weights and the positions of the vertices

$$p_{uv} := \mathbb{P}((u, v) \in E) = \min \left\{ \frac{1}{\|x_u - x_v\|^{\gamma d}} \left( \frac{w_u w_v}{W} \right)^{\gamma}, 1 \right\}, \quad (3)$$

where the exponent  $\gamma > 1$  is a fixed parameter. Here we use the  $\infty$ -norm to measure  $\|x_u - x_v\|$ . That is, for any  $x, y \in \mathbb{T}^d$

$$\|x - y\| := \max_{1 \leq i \leq d} |x_i - y_i|_C. \quad (4)$$

Here  $|\cdot|_C$  denotes the distance on the circle  $\mathbb{T}^1$ , namely, for any  $a, b \in \mathbb{T}^1$

$$|a - b|_C := \min\{|a - b|, 1 - |a - b|\} \quad (5)$$

Equation (3) shows an interesting relation between the properties of the vertices and their connection probabilities: Two vertices with high weights are more likely to connect. However, vertices with high weights are *rare* due to (1). Moreover, two vertices are more likely to connect if their positions are close. However, again, the probability for two vertices to have close locations is small due to (2). In our main results, we will exploit this trade-off formally in the form of an optimization problem.

## 2 Main result

The aim of this paper is to study the emerging subgraph structures inside the GIRG, as the number of vertices  $n$  goes to infinity. In particular we are interested in computing  $N(\mathcal{K}_k)$ , the number of complete subgraphs (cliques) appearing in the GIRG. In our computations we assume that  $k$  is small compared to  $n$ , in particular  $k = O(1)$ .

Our results do not only investigate the number of cliques, but also show where in the GIRG these cliques are most likely to be located in terms of their positions and their weights. In particular, we will consider weights and distances between the vertices as quantities scaling with  $n$ , and show that most cliques are found on vertices whose weights and distances scale as specific values of  $n$ . Indeed, as the positions of the vertices enter in the edge probabilities introduced in (3) only through their distances, we are not interested in the position of each vertex, but rather on their distances.

We now introduce some notation for our main theorem that introduces the number of cliques in regions of  $\text{GIRG}^{(n)}$  with specific vertex weights and distances. We say that  $x \in I_\varepsilon(n^\nu)$  if  $x \in [\varepsilon n^\nu, \frac{1}{\varepsilon} n^\nu]$  for some  $0 < \varepsilon < 1$ . Fix a sequence  $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_k\}$  of non-negative real numbers and a sequence  $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_k\}$  of non-positive real vectors of length  $d$ , where  $d$  denotes the dimension of the GIRG model. Moreover, set  $\beta_1 := [-\infty, \dots, -\infty]$ , and define

$$M^{(\boldsymbol{\alpha}, \boldsymbol{\beta})} = \left\{ (v_1, \dots, v_k) : w_{v_i} \in I_\varepsilon(n^{\alpha_i}), |x_{v_i} - x_{v_1}| \in \left[ I_\varepsilon(n^{\beta_i^{(1)}}), \dots, I_\varepsilon(n^{\beta_i^{(d)}}) \right], \forall i \in [k] \right\} \quad (6)$$

where  $|y - z|$  denotes the component-wise distance between the vectors  $y$  and  $z$ . The set  $M^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}$  contains all the lists of  $k$  vertices such that their weights scale with  $n$  according to  $\boldsymbol{\alpha}$ , and such that their distances from  $v_1$  scale with  $n$  according to  $\boldsymbol{\beta}$ . Observe that in (6) we are setting the origin of the torus as the position of vertex  $v_1$ , so that all the other vertices lie in some neighbourhood of the origin. This choice makes sense because the edge probability of the GIRG defined in (3) depends on the distances between the vertices, and not on their absolute positions and the positions are distributed uniformly. Thus, in order to study the number of  $k$ -cliques in the GIRG, we can fix the position of any vertex  $v_1$  without loss of generality (by symmetry), and allow the position parameters  $\beta_2, \dots, \beta_k$  of the remaining  $k - 1$  vertices to vary.

The number of  $k$ -cliques inside  $\text{GIRG}^{(n)}$  with vertices in  $M^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}$  will be denoted by  $N(\mathcal{K}_k, M^{(\boldsymbol{\alpha}, \boldsymbol{\beta})})$ . Then our aim is to prove that there exists a specific set of values  $\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$  such that the number of  $k$ -cliques  $N(\mathcal{K}_k, M^{(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)})$  is predominant among all others. The next theorem states that such optimal values are explicitly determined, depending on  $k$ , the size of the clique, and on  $\tau$ , the power law exponent for the weights of the vertices.

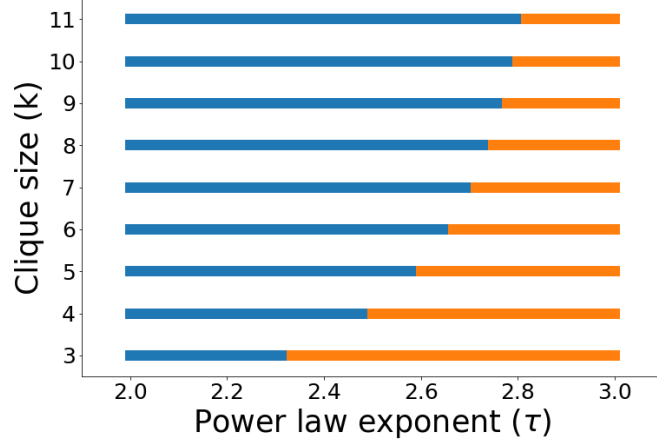


Figure 1: Phase transition described by Theorem 1. The blue region corresponds to  $k > \frac{2}{3-\tau}$  (non-geometric case), where most of the cliques generate independently from the geometry of the system; whereas, in the orange region  $k < \frac{2}{3-\tau}$  (geometric case), cliques appear most likely between vertices at distance  $\Theta(n^{-\frac{1}{d}})$ .

**Theorem 1.** *Suppose  $\gamma \neq \tau - 1$ . Then, there exist  $\alpha^*, \beta^*$  such that*

$$\frac{\mathbb{E}_n[N(\mathcal{K}_k, M(\alpha, \beta))]}{\mathbb{E}_n[N(\mathcal{K}_k, M(\alpha^*, \beta^*))]} = 0 \quad \text{w.h.p.}, \quad \forall (\alpha, \beta) \neq (\alpha^*, \beta^*). \quad (7)$$

Moreover,

$$N(\mathcal{K}_k, M(\alpha^*, \beta^*)) = \begin{cases} \Theta_{\mathbb{P}}(n) & \text{if } k < \frac{2}{3-\tau}, \\ \Theta_{\mathbb{P}}(n^{\frac{3-\tau}{2}k}) & \text{if } k > \frac{2}{3-\tau}. \end{cases} \quad (8)$$

and  $\alpha^*, \beta^*$  are uniquely determined by

$$\alpha_i^* = \begin{cases} 0 & \forall i \in [k], \text{ if } k < \frac{2}{3-\tau}, \\ \frac{1}{2} & \forall i \in [k], \text{ if } k > \frac{2}{3-\tau}. \end{cases}, \quad (9)$$

$$\beta_i^* = \begin{cases} [-\frac{1}{d}, \dots, -\frac{1}{d}] & \forall i \in [k] \setminus 1, \text{ if } k < \frac{2}{3-\tau}, \\ [0, \dots, 0] & \forall i \in [k] \setminus 1, \text{ if } k > \frac{2}{3-\tau}. \end{cases}. \quad (10)$$

In particular, Theorem 1 shows that there exists a phase transition for the number of cliques in the GIRG, depending on  $k$  and  $\tau$ . When  $k < \frac{2}{3-\tau}$ , the predominant number of cliques scales as  $n$ . Furthermore, most cliques appear between vertices with small distances, proportional to  $n^{\beta^*} = n^{-1/d}$ , and low degrees, that do not grow in  $n$ , as  $n^{\alpha^*} = 1$ . On the contrary, when  $k > \frac{2}{3-\tau}$  the number of cliques scales as  $n^{\frac{3-\tau}{2}k} \gg n$ . In this case, the most predominant cliques are formed on vertices with high weights, proportional to  $n^{\alpha^*} = \sqrt{n}$ , but arbitrarily distant from each other, as  $n^{\beta^*} = 1$ , which is also the maximal distance in the GIRG model.

In the latter case the geometry does not influence the dominant clique structure. That is, the number of cliques does not depend on geometric features (the positions of the vertices) anymore.

This is further illustrated by the fact that our result is equivalent to the analogous result (see Theorem 2.2 in [11]) for scale-free configuration models, in which geometry is not involved. For this reason, we call the two different cases *geometric* and *non-geometric* (see Figure 1).

## 2.1 Simulations

We provide here simulations of the GIRG model. For each sample of the GIRG we count the number of triangles, and compare it to the expected asymptotic behaviour predicted by Theorem 1. In [3] Bläsius et al. provide an algorithm to sample GIRGs efficiently, with expected running time  $\Theta(n + m)$  (where  $n, m$  denote the number of vertices and edges of the GIRG). We make use of a C++ library which implements this algorithm [15, 9].

The required parameters for each sample of the GIRG model are:  $n$ ,  $\tau$ ,  $d$  and the *temperature*  $1/\gamma$ . Moreover, in the code we add an additional parameter  $c$ , corresponding to the constant factor in the weight distribution (1). This factor affects the edge probability in (3): the mean number of edges in the graph increases as  $c$  increases.

To count the number of triangles in a GIRG, we use the *forward algorithm* [16, 14], which has a running time of  $O(m^{3/2})$ , where  $m$  denotes the number of edges. As in the GIRG model the number of edges scales as the number of vertices [6], this is therefore equivalent to a running time of  $O(n^{3/2})$ . Figure 2 shows the number of triangles obtained from simulations, against the number of vertices  $n$  for the two regimes of  $\tau$  distinguished by Theorem 1 for  $k = 3$ :  $\tau < 7/3$  and  $\tau > 7/3$ . Indeed, for  $\tau = 2.1$ , Figure 2a and Theorem 1 show that the optimal weights and distances are  $\alpha^* = \frac{1}{2}, \beta^* = 0$ , and that the asymptotic behaviour of  $N(\Delta, M^{(\frac{1}{2}, 0)})$  is  $\Theta_{\mathbb{P}}(n^{1.35})$ . Instead, in Figure 2b,  $\tau = 2.7$ , so that Theorem 1 predicts that the optimal values are  $\alpha^* = 0, \beta^* = -\frac{1}{d}$ , and the asymptotic behaviour of  $N(\Delta, M^{(0, -\frac{1}{d})})$  is  $\Theta_{\mathbb{P}}(n)$ . These different asymptotic slopes are shown in Figures 2a and 2b, and in both cases, our simulations follow these asymptotic slopes quite well. In particular, while the results from Theorem 1 are asymptotic in  $n$ , our simulations show that these asymptotics are visible for networks with sizes of only thousands of vertices, and even less.

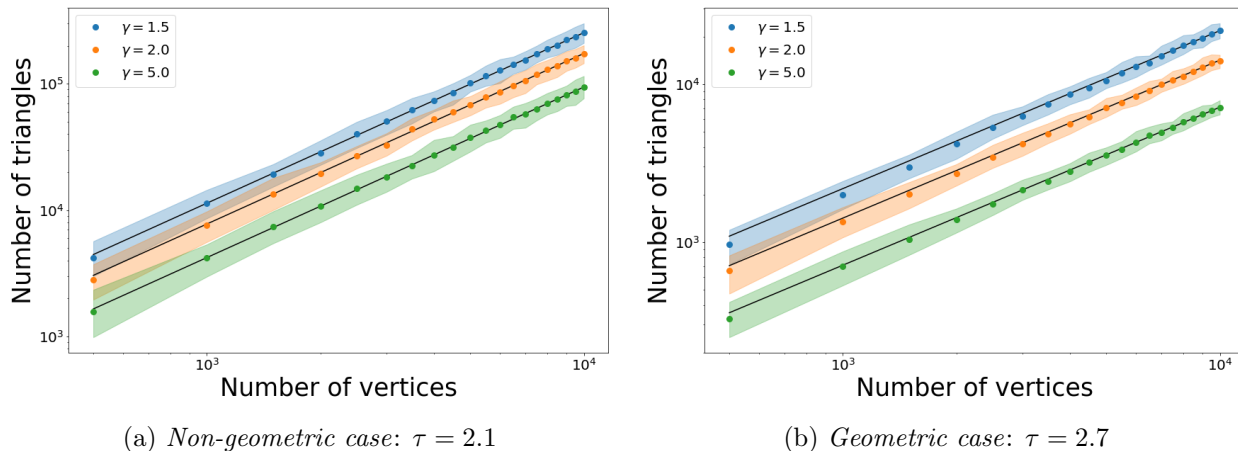


Figure 2: The number of triangles against  $n$  for different values of  $\gamma$  for  $d = 1$ ,  $c = 0.4$ . Black curves show the asymptotic behaviour of  $N(\Delta, M^{(\alpha^*, \beta^*)})$  predicted by Theorem 1, colored dots indicate the average number of triangles over 100 samples of the GIRG. The colored regions contain 80% of all samples.

## 2.2 Discussion

In this paper, we analyze the number of cliques in a general model that incorporates power-law degrees as well as geometry. We also investigate the typical structure of a clique of any given size, and show that this structure depends on the clique size and the power-law exponent. We now discuss some implications of our main results.

**Non-geometry for  $\tau < 7/3$ .** We now analyze the phase-transition of Theorem 1 in more detail. One interesting thing is that when  $\tau < \frac{7}{3}$  the most common  $k$ -cliques are non-geometric, for any  $k \geq 3$ . This can be observed, for instance, in Figure 2. Furthermore, Theorem 1 shows that in this setting, the number of dominating cliques scales as  $n^{k(3-\tau)/2}$  for all  $k$ , which is the same scaling in  $n$  as in many non-geometric scale-free models, such as in the inhomogeneous random graph, the erased configuration model and the uniform random graph [18, 17, 11]. This seems to imply that when  $\tau < 7/3$ , we cannot distinguish geometric and non-geometric scale-free networks by counting the number of cliques, or by studying the clustering coefficient. Thus, in this regime of  $\tau$ , the added geometry of the GIRG model does not add any clustering.

On the other hand, when  $\tau > \frac{7}{3}$ , then small cliques and large cliques behave differently (see Figure 1). Indeed, cliques of small size  $k < 2/(3-\tau)$  are predominantly present on low-degree, close by vertices of distances as low as  $n^{-1/d}$ . Furthermore, the number of such small cliques scales as  $n$ , which is larger than the clique scaling of  $n^{k/2(3-\tau)}$  in the inhomogeneous random graph without geometry [18]. Thus, for  $\tau > 7/3$ , smaller cliques are influenced by geometry, whereas large cliques are not. In this case, it is clearly possible to distinguish between geometric and non-geometric inhomogeneous random graphs through small clique counts. Therefore, studying such statistical tests that distinguish geometric and non-geometric random graphs in more detail would be an interesting avenue for further research.

**Insensitivity to  $\gamma$ .** In the result of Theorem 1, the parameter  $\gamma$  of the GIRG model does not contribute to the asymptotic behaviour of  $N(\mathcal{K}_k)$ , nor to the determination of the phase-transition. This may appear counterintuitive, as the edge probability defined in (3) decreases as  $\gamma$  increases. Hence, for higher values of  $\gamma$ , we should expect fewer edges and therefore fewer cliques to appear. However, by direct computation, it is possible to see that any vertices in the optimal configuration  $(v_1, \dots, v_k) \in M(\alpha^*, \beta^*)$  connect to each other with probability  $\Theta(1)$ , regardless of the value of  $\gamma$ . This is the reason why asymptotically the presence of  $\gamma$  is irrelevant. However, Theorem 1 only computes the asymptotic scaling of the number of cliques in terms of the number of vertices, whereas we do expect the parameter  $\gamma$  to play a role for computing the leading-order constant. Indeed, from Figures 2a and 2b it is clear that if we increase  $\gamma$ , the leading-order constant decreases, and consequently the GIRG contains a lower number of triangles.

**Total number of cliques.** Lastly, we observe that Theorem 1 deals with the number of optimal cliques  $N(\mathcal{K}_k, M(\alpha^*, \beta^*))$ , not with the total number of cliques in the graph  $N(\mathcal{K}_k)$ . Indeed we only determine which kind of cliques appear most frequently, and we obtain the scaling of these dominant cliques. Nevertheless, our simulations suggest that the total number of cliques has the same asymptotic behaviour as the number of optimal cliques. As a consequence, this would imply that if we pick randomly a clique of the GIRG, then its vertices will be in  $M(\alpha^*, \beta^*)$  with high probability (when  $n$  is large). Vice versa, this suggests that if we want to compute the number of  $k$ -cliques in large GIRGs, we just need to check the set of optimal vertices  $(v_1, \dots, v_k)$ , such that

$(v_1, \dots, v_k) \in M^{(\alpha^*, \beta^*)}$ . That is, we believe that the total number of cliques in the GIRG model has the same scaling as the dominant number of cliques in Theorem 1. Proving this however, needs the computation of the integral of the exact clique probability over all  $\alpha, \beta$  in the optimal regime, as was done for a simpler model without geometry in [11], and remains an open direction for further research.

**General heavy-tailed weight distributions.** In (1), the weight of the vertices follows a Pareto distribution with power-law exponent  $\tau \in (2, 3)$ . That is, the weights are independently sampled from a random variable  $w$  with probability density function

$$f_w(w) = \begin{cases} \ell/w^\tau & \text{if } w \geq w_{\min} \\ 0 & \text{if } w < w_{\min} \end{cases} \quad (11)$$

for some  $w_{\min} > 0$ , with  $\ell = (\tau - 1)w_{\min}^{\tau-1}$ . However, since the results summarized in the current section hold asymptotically, it is worth nothing to prove that Theorem 1 still works with more general heavy-tailed weight distributions. Indeed, we can consider a probability density function whose behaviour at infinity is *similar* to the behaviour of a power law function. This is done replacing  $\ell$  in (11) with a bounded slowly varying function  $\ell(w)$ , that is, a bounded measurable function  $\ell : (w_{\min}, \infty) \rightarrow (0, \infty)$  such that  $\lim_{w \rightarrow \infty} \ell(aw)/\ell(w) = 1$ , for all  $a > 0$ , and  $f_w(w)$  is a probability density function. If  $\ell$  in (11) is replaced by a slowly varying function  $\ell(w)$ , then the computations done in Section 3 are unaffected. Indeed, we solve an optimization problem containing powers of  $n$ . The slowly varying function on the other hand, grows slower than any power of  $n$ , and therefore does not affect significantly this optimization problem, nor its solution. However, while the scaling in (7) of the total number of cliques would still have the same polynomial leading order term, it would contain an additional slowly varying factor as well. The relation between the slowly varying function  $\ell$  and the correct asymptotic scaling of the number of cliques remains an open problem for future research.

**Relation to hyperbolic random graphs.** In the past decade, hyperbolic random graphs have been studied widely, as random graph models that include both geometry and scale-free vertex degrees. The downside of analyzing hyperbolic random graphs is that they come with hyperbolic sine and cosine functions, which are typically difficult to work with. However, hyperbolic random graphs can also be seen as a special case of GIRGs, when the dimension is  $d = 1$  and the temperature is  $1/\gamma = 0$  ([6], section 4). In this threshold case,  $\gamma = \infty$ , the connection probability (3) of the GIRG model becomes

$$p_{uv} = \begin{cases} 1, & \text{if } \|x_u - x_v\| \leq \left(\frac{w_u w_v}{W}\right)^{1/d}, \\ 0, & \text{if } \|x_u - x_v\| > \left(\frac{w_u w_v}{W}\right)^{1/d}. \end{cases} \quad (12)$$

As these connection probabilities are easy to determine when the weights of the vertices and their positions are known, we believe that in the case  $\gamma = \infty$  it is still possible to estimate asymptotically the number of cliques, in a similar spirit as Theorem 1, by solving a slightly different optimization problem. In turn, the methodology provided in this paper would work for estimating the number of cliques in hyperbolic random graphs as well. In fact, it is interesting to observe that the result shown by Bläsius, Friedrich, and Khromer in [4] for the expected number of cliques in hyperbolic random graphs is very similar to Theorem 1 presented here. Indeed, they were able to prove that there exists two different regimes for the number of cliques, depending on the size  $k$ , where the transition



point between the different regimes agrees with the one we here obtain for the more general GIRG model.

### 3 Optimization problem

We now describe how Theorem 1 can be proven through solving an optimization problem. First we estimate asymptotically the mean value of  $N(\mathcal{K}_k, M^{(\alpha, \beta)})$  as follows:

$$\mathbb{E}_n[N(\mathcal{K}_k, M^{(\alpha, \beta)})] = n|M^{(\alpha, \beta)}| \cdot \mathbb{P}\left((v_1, \dots, v_k) \in M^{(\alpha, \beta)} \text{ form a } k\text{-clique}\right). \quad (13)$$

In order to prove Theorem 1, we will solve the maximization problem

$$\max_{\alpha, \beta} \mathbb{E}_n[N(\mathcal{K}_k, M^{(\alpha, \beta)})]. \quad (14)$$

We will then show that the solution of (14) determines the features of the most predominant cliques in the graph, as shown in (7) of Theorem 1.

Using (13), we can split the mean value of  $N(\mathcal{K}_k, M^{(\alpha, \beta)})$  into the product of two terms. The first term,  $|M^{(\alpha, \beta)}|$  defined in (6), is the number of vertices  $(v_1, \dots, v_k) \in M^{(\alpha, \beta)}$ , which depends on  $w_1, \dots, w_n$  and of  $x_2, \dots, x_n$ . Note that  $|M^{(\alpha, \beta)}|$  does not depend on  $x_1$ , as we fixed the origin to the position of vertex 1. To adjust for this, we added an extra factor  $n$  to (13) as any of the  $n$  vertices can be defined as vertex 1. For  $\alpha \geq 0, \beta = [\beta^{(1)}, \dots, \beta^{(d)}] \leq 0$  fixed, observe that the number of vertices  $v \in V$  such that  $w_v \in I_\varepsilon(n^\alpha), |x_v - x_{v_1}| \in I_\varepsilon(n^\beta)$  is Binomial( $n, p$ ), with  $p = \mathbb{P}(w_v \in I_\varepsilon(n^\alpha), |x_v - x_{v_1}| \in I_\varepsilon(n^\beta))$ , as all positions and weights are i.i.d.. In particular,  $p = \Theta(n^{(1-\tau)\alpha + \beta^{(1)} + \dots + \beta^{(d)}})$ , so that the number of vertices with prescribed weight and position  $w_v \in I_\varepsilon(n^\alpha), |x_v - x_{v_1}| \in I_\varepsilon(n^\beta)$  is  $\Theta_{\mathbb{P}}(n^{1+(1-\tau)\alpha + \beta^{(1)} + \dots + \beta^{(d)}})$ . Consequently,

$$\begin{aligned} |M^{(\alpha, \beta)}| &= n^{(1-\tau)\alpha_1} \prod_{i=2}^k \Theta_{\mathbb{P}}(n^{1+(1-\tau)\alpha_i + \beta_i^{(1)} + \dots + \beta_i^{(d)}}) \\ &= \Theta_{\mathbb{P}}(n^{k-1+(1-\tau)\sum_i \alpha_i + \sum_{i>1, j} \beta_i^{(j)}}) \end{aligned} \quad (15)$$

Now we focus on the second term in (13), which describes the probability that  $k$  randomly chosen vertices of given weights and position from a clique. The vertices  $v_1, \dots, v_k$  form a  $k$ -clique if and only if every possible pair of vertices is connected. The connection probabilities of different vertices are independent, conditionally on their weights and positions. Then, for any  $(v_1, \dots, v_k) \in M^{(\alpha, \beta)}$  we have

$$\begin{aligned} \mathbb{P}((v_1, \dots, v_k) \text{ form a } k\text{-clique}) &= \prod_{i<j} \mathbb{P}((v_i, v_j) \in E) \\ &= \prod_{i<j} \min \left\{ \left( \frac{1}{\|x_{v_i} - x_{v_j}\|^d} \frac{w_{v_i} w_{v_j}}{W} \right)^\gamma, 1 \right\}. \end{aligned} \quad (16)$$

Let  $i, j$  be fixed, and consider the probability of  $v_i$  and  $v_j$  to connect. First note that  $\frac{w_{v_i} w_{v_j}}{W} = \Theta_{\mathbb{P}}(n^{\alpha_i + \alpha_j - 1})$ , because  $W = \Theta_{\mathbb{P}}(n)$ . Moreover, the distance between  $v_i$  and  $v_j$  can be rewritten as  $\|x_{v_i} - x_{v_j}\| = \max_{1 \leq h \leq d} \{ |x_{v_i}^{(h)} - x_{v_j}^{(h)}|_C \}$ , as we are considering the metric induced by the  $\infty$ -norm. For each  $h \in [d]$  there are two possibilities.

- If  $\beta_i^{(h)} \neq \beta_j^{(h)}$ , then

$$|\mathbf{x}_{v_i}^{(h)} - \mathbf{x}_{v_j}^{(h)}|_C = \left| I_\varepsilon \left( n^{\beta_i^{(h)}} \right) - I_\varepsilon \left( n^{\beta_j^{(h)}} \right) \right| = \Theta \left( n^{\max\{\beta_i^{(h)}, \beta_j^{(h)}\}} \right). \quad (17)$$

- If  $\beta_i^{(h)} = \beta_j^{(h)} =: \beta^{(h)}$ , the component-wise distance is more difficult to obtain, since

$$|\mathbf{x}_{v_i}^{(h)} - \mathbf{x}_{v_j}^{(h)}|_C = \left| I_\varepsilon \left( n^{\beta_i^{(h)}} \right) - I_\varepsilon \left( n^{\beta_j^{(h)}} \right) \right| = \Theta(n^{\delta^{(h)}}) \quad (18)$$

for some  $\delta^{(h)} \leq \beta^{(h)}$ . However, we now show that  $\delta^{(h)} = \beta^{(h)}$  provides the dominant contribution to (13). For simplicity we consider the 1-dimensional case  $d = 1$ , but by using the same argument and introducing some additional notation, it is possible to prove the result for any dimension.

Let  $d = 1$ , and suppose  $\beta_i = \beta_j = \beta$ , that is  $|\mathbf{x}_{v_i} - \mathbf{x}_{v_1}| \in I_\varepsilon(n^\beta)$  and  $|\mathbf{x}_{v_j} - \mathbf{x}_{v_1}| \in I_\varepsilon(n^\beta)$ . As  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are sampled uniformly within the interval  $\mathbf{x}_{v_1} \pm [\varepsilon n^\beta, n^\beta/\varepsilon]$ , the probability that the distance between  $v_i$  and  $v_j$  is proportional to  $n^\delta$ , with  $\delta \leq \beta$ , is

$$\mathbb{P}(\|\mathbf{x}_{v_i} - \mathbf{x}_{v_j}\| = \Theta(n^\delta)) = \Theta(n^{\delta-\beta}).$$

The probability that  $v_i$  and  $v_j$  connect, given that their distance is proportional to  $n^\delta$ , is

$$\mathbb{P}\left((v_i, v_j) \in E \mid \|\mathbf{x}_{v_i} - \mathbf{x}_{v_j}\| = \Theta(n^\delta)\right) = \Theta_{\mathbb{P}}(n^{\min\{\gamma(\alpha_i + \alpha_j - 1 - \delta), 0\}}),$$

by (3). Thus, we have

$$\begin{aligned} \mathbb{P}\left((v_i, v_j) \in E, \|\mathbf{x}_{v_i} - \mathbf{x}_{v_j}\| = \Theta_{\mathbb{P}}(n^\delta)\right) &= \Theta_{\mathbb{P}}(n^{\delta - \beta + \min\{\gamma(\alpha_i + \alpha_j - 1 - \delta), 0\}}) \\ &= \Theta_{\mathbb{P}}(\min\{n^{\delta(1-\gamma) - \beta + \gamma(\alpha_i + \alpha_j - 1 - \delta)}, n^{\delta - \beta}\}). \end{aligned} \quad (19)$$

Since our aim is to solve the maximization problem in (14), we just need to determine for which  $\delta \leq \beta$  the contribution of (19) to (13) is maximized. As  $\gamma > 1$ , (19) is optimized for  $\delta = \beta$ . This proves that, we may assume that if the position of two vertices  $v_i, v_j$  are both in a neighborhood of  $v_1$  of size  $\approx n^\beta$ , then also their distance is  $\approx n^\beta$  in (14).

Then, (16) yields

$$\begin{aligned} \mathbb{P}((v_1, \dots, v_k) \text{ form a } k\text{-clique}) &= \prod_{i < j} \min \left\{ \frac{\Theta_{\mathbb{P}}(n^{\gamma(\alpha_i + \alpha_j - 1)})}{\max_h \left[ \Theta_{\mathbb{P}} \left( \max\{n^{\beta_i^{(h)}}, n^{\beta_j^{(h)}}\} \right)^{d\gamma} \right]}, 1 \right\} \\ &= \Theta_{\mathbb{P}} \left( n^{\sum_{i < j} \gamma \min\{\alpha_i + \alpha_j - 1 - d \max_h(\max\{\beta_i^{(h)}, \beta_j^{(h)}\}), 0\}} \right). \end{aligned} \quad (20)$$

Combining this with (13) and (15) shows that the mean number of cliques with vertices in  $M(\boldsymbol{\alpha}, \boldsymbol{\beta})$  satisfies

$$\mathbb{E}_n[N(\mathcal{K}_k, M(\boldsymbol{\alpha}, \boldsymbol{\beta}))] = \Theta_{\mathbb{P}} \left( n^{k + (1-\tau) \sum_i \alpha_i + \sum_{i > 1, j} \beta_i^{(j)} + \sum_{i < j} \gamma \min\{\alpha_i + \alpha_j - 1 - d \max_h(\max\{\beta_i^{(h)}, \beta_j^{(h)}\}), 0\}} \right). \quad (21)$$

For simplicity, we denote the exponent in the right hand side of (21) as

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) := k + (1 - \tau) \sum_i \alpha_i + \sum_{i>1, j} \beta_i^{(j)} + \sum_{i<j} \gamma \min\{\alpha_i + \alpha_j - 1 - d \max_h(\max\{\beta_i^{(h)}, \beta_j^{(h)}\}), 0\}. \quad (22)$$

At this point, we observe that the optima for  $\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} f(\boldsymbol{\alpha}, \boldsymbol{\beta})$  are the same as those of the maximization problem for  $\mathbb{E}_n[N(\mathcal{K}_k, M^{(\boldsymbol{\alpha}, \boldsymbol{\beta})})]$  defined in (14). In the statement of the next proposition we characterize its solution.

**Proposition 2.** *Let  $f$  be defined as in (22), and consider the problem*

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} f(\boldsymbol{\alpha}, \boldsymbol{\beta}), \quad (23)$$

*with the constraints  $\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta} \leq 0$ . The solution of the maximum problem is attained by one of the following two sets of parameters for  $(\alpha_i)_{i \in [k]}$  and  $(\beta_i)_{i \in [k]}$ :*

$$\alpha_i = \frac{1}{2} \quad \forall i \in [k], \quad \beta_i^{(j)} = 0 \quad \forall i \in [k] \setminus 1, j \in [d], \quad (24)$$

$$\alpha_i = 0 \quad \forall i \in [k], \quad \beta_i^{(j)} = -\frac{1}{d} \quad \forall i \in [k] \setminus 1, j \in [d]. \quad (25)$$

Thus, to obtain the optimal  $\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$  described in Theorem 1, we only need to investigate which of the two candidates listed in Proposition 2 attains the maximal value of  $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$ , which we will do in Section 5. There, we will also prove Theorem 1 by showing that the number of cliques formed on the set of vertices in  $M^{(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}$  converges to its mean value:

$$N(\mathcal{K}_k, M^{(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}) = \mathbb{E}_n[N(\mathcal{K}_k, M^{(\boldsymbol{\alpha}, \boldsymbol{\beta})})](1 + o(1)). \quad (26)$$

## 4 Proof of Proposition 2

To prove Proposition 2, we need first some technical lemmas. The first lemma enables to simplify the exponent  $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$  in (22).

**Lemma 3.** *In optimal solutions of (23)  $\beta_i^{(j_1)} = \beta_i^{(j_2)}$  for all  $j_1, j_2 \in [d], i \in [k]$ .*

*Proof.* By contradiction. Suppose that  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  is an optimizer of (23) and that there exists  $\bar{i} \in [k]$  and  $j_1, j_2 \in [d]$  such that  $\beta_{\bar{i}}^{(j_1)} \neq \beta_{\bar{i}}^{(j_2)}$ . We can assume without loss of generality that  $\beta_{\bar{i}}^{(j_2)} > \beta_{\bar{i}}^{(j_1)}$ .

Then, we define the set of parameters  $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_i^{(j)}\}_{i,j}$  as follows:

$$\hat{\beta}_i^{(j)} = \begin{cases} \beta_{\bar{i}}^{(j_2)} & \text{if } i = \bar{i} \text{ and } j = j_1, \\ \beta_{\bar{i}}^{(j)} & \text{otherwise.} \end{cases} \quad (27)$$

Note that

$$\max_h(\max\{\beta_i^{(h)}, \beta_j^{(h)}\}) = \max_h(\max\{\hat{\beta}_i^{(h)}, \hat{\beta}_j^{(h)}\}). \quad (28)$$

Then, using the definition of  $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$  in (22), we observe that

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) - f(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}) = \beta_{\bar{i}}^{(j_1)} - \hat{\beta}_{\bar{i}}^{(j_1)} = \beta_{\bar{i}}^{(j_1)} - \beta_{\bar{i}}^{(j_2)} < 0. \quad (29)$$

Therefore  $f(\boldsymbol{\alpha}, \boldsymbol{\beta}) < f(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}})$ , which contradicts the optimality of  $\boldsymbol{\alpha}, \boldsymbol{\beta}$ .  $\square$

By applying Lemma 3, we can rewrite  $f(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}})$  as

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) = k + (1 - \tau) \sum_i \alpha_i + d \sum_{i \neq 1} \beta_i + \sum_{i < j} \gamma \min\{\alpha_i + \alpha_j - 1 - d \max\{\beta_i, \beta_j\}, 0\}, \quad (30)$$

where we replaced  $\beta_i^{(d)}$  by  $\beta_i$  for all  $d$ . Thus, from now on the parameters  $\beta_1, \dots, \beta_k$  will denote the value of all the components of the  $d$ -dimensional vector, instead of the vector itself.

We now introduce some notation that will simplify the rest of the proof of Proposition 2. Let  $I_1, I_2 \subseteq [k]$ ,  $J_2 \subseteq [d]$  be a selection of indices. Writing  $f(\{\alpha_i\}_{i \in I_1}, \{\beta_i^{(j)}\}_{i \in I_2, j \in J_2})$  we indicate the exponent  $f$  where our attention is limited to the parameters in the argument. That is, the contribution of all the terms in (30) which are independent from the argument will be omitted, and encoded in a constant value  $C$ . For instance, if we focus on the contribution of only  $\alpha_i$  or  $\beta_i$  we write

$$f(\alpha_i) = C + (1 - \tau)\alpha_i + \sum_j \gamma \min\{\alpha_i + \alpha_j - 1 - d \max\{\beta_i, \beta_j\}, 0\} \quad (31)$$

$$f(\beta_i) = C + d\beta_i + \sum_{j: \beta_j \leq \beta_i} \gamma \min\{\alpha_i + \alpha_j - 1 - d\beta_i, 0\}. \quad (32)$$

If we focus on the contribution of a single vertex  $v_i$  we write

$$f(\alpha_i, \beta_i) = C + (1 - \tau)\alpha_i + \beta_i + \sum_{j \neq i} \gamma \min\{\alpha_i + \alpha_j - 1 - d \max\{\beta_i, \beta_j\}, 0\}. \quad (33)$$

Then, the next lemma provides a lower bound for  $\beta_1, \dots, \beta_k$ , and a relation to the optimal value of  $\alpha_i$  for all  $i$  such that  $\beta_i$  attains the lower bound:

**Lemma 4.** *Suppose  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  is an optimal solution of (23). Then:*

- (i) *there is no  $i \in [k] \setminus \{1\}$  such that  $\beta_i < -\frac{1}{d}$ ;*
- (ii) *if  $\beta_i = -\frac{1}{d}$  for some  $i \in [k] \setminus \{1\}$  then  $\alpha_i = 0$ .*

*Proof.* We first prove (i) by contradiction. Let  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  be a maximizer of (23). We sort and rename the coefficients of  $\boldsymbol{\beta}$  in increasing order, so that  $\beta_1 < \beta_2 \leq \dots \leq \beta_k$ . Suppose that  $\beta_h < -\frac{1}{d}$  for some  $h > 1$ , and assume without loss of generality that either  $\beta_h < \beta_{h+1}$  or  $h = k$ . Then we can rewrite (31) as

$$\begin{aligned} f(\beta_h) &= C + d\beta_h + \sum_{j < h} \gamma \min\{\alpha_h + \alpha_j - 1 - d\beta_h, 0\} \\ &= C + d\beta_h \end{aligned} \quad (34)$$

where the constant  $C$  encodes all terms in  $f$  independent from  $\beta_h$ . The last step follows because  $\sum_{j < h} \gamma \min\{\alpha_h + \alpha_j - 1 - d\beta_h, 0\} = 0$ , as  $\beta_h < -\frac{1}{d}$  and the parameters  $\alpha_i$  are non-negative.

Now we define a new set of parameters  $\hat{\beta}_i$  such that  $\hat{\beta}_i = \beta_i$  for all  $i \neq h$  and  $\hat{\beta}_h = \min(\beta_{h+1}, -\frac{1}{d})$ . Since  $\hat{\beta}_j \geq \hat{\beta}_h$ , for all  $j > h$ , we have

$$\begin{aligned} f(\hat{\beta}_h) &= C + d\hat{\beta}_h + \sum_{j < h} \gamma \min\{\alpha_h + \alpha_j - 1 - d\hat{\beta}_h, 0\} \\ &= C + d\hat{\beta}_h, \end{aligned} \quad (35)$$

where again  $\sum_{j < h} \gamma \min\{\alpha_h + \alpha_j - 1 - d\hat{\beta}_h, 0\} = 0$ , as  $\hat{\beta}_h \leq -\frac{1}{d}$ . Therefore

$$f(\hat{\beta}_h) - f(\beta_h) = d\left(\min\left(\beta_{h+1}, -\frac{1}{d}\right) - \beta_h\right) > 0. \quad (36)$$

This contradicts the optimality of  $\beta$ , and therefore proves (i).

We now prove (ii). Suppose that  $\alpha, \beta$  is an optimal solution of (23) and  $\beta_i = -\frac{1}{d}$  for some  $i \in [k] \setminus \{1\}$ . We use an increasing ordering for  $\beta$ , so that  $\beta_2, \dots, \beta_s = -\frac{1}{d}$ ,  $\beta_{s+1} > -\frac{1}{d}$  for some  $s \in [k]$ , and with  $\alpha_2 \leq \dots \leq \alpha_s$ . Now we focus on the contribution given by  $\beta_s$  and  $\alpha_s$  to  $f$  as in (33)

$$f(\alpha_s, \beta_s) = C + (1-\tau)\alpha_s + d\beta_s + \sum_{j < s} \gamma \min\{\alpha_s + \alpha_j - 1 - d\beta_s, 0\} + \sum_{j > s} \gamma \min\{\alpha_s + \alpha_j - 1 - d\beta_j, 0\}. \quad (37)$$

Note that, if  $\alpha_s + \alpha_j > 0$  for all  $j < s$ , then the optimality of the solution is violated. Indeed, in this case, we may define  $\hat{\beta}_s := \beta_s + \varepsilon$ , where  $\varepsilon > 0$  is small enough such that  $\hat{\beta}_s < \beta_{s+1}$  and  $\hat{\beta}_s \leq (\alpha_s + \alpha_j - 1)/d$ , for all  $j < s$ . Then, observe that

$$\sum_{j < s} \gamma \min\{\alpha_s + \alpha_j - 1 - d\beta_s, 0\} = 0 = \sum_{j < s} \gamma \min\{\alpha_s + \alpha_j - 1 - d\hat{\beta}_s, 0\}$$

where the first equality follows by the fact that  $\alpha_s + \alpha_j \geq 0$  and  $\beta_s = \frac{1}{d}$ , whereas the second equality follows from  $\hat{\beta}_s \leq (\alpha_s + \alpha_j - 1)/d$ . Then,

$$f(\alpha_s, \hat{\beta}_s) - f(\alpha_s, \beta_s) = d(\hat{\beta}_s - \beta_s) > 0$$

which contradicts the maximality assumption.

Consequently, the only possibility is that there exists  $s' < s$  such that  $\alpha_s + \alpha_{s'} = 0$ . In this case, since  $\alpha_s, \alpha_{s'} \geq 0$ , also  $\alpha_s = \alpha_{s'} = 0$ . Then, also  $\alpha_2 = \dots = \alpha_s = 0$ , as from our initial assumption  $\alpha_2 \leq \dots \leq \alpha_s$ .  $\square$

The next lemma shows that in any optimal solution of (23), the values  $\alpha_1, \dots, \alpha_k$  and  $\beta_1, \dots, \beta_k$  can be ordered increasingly jointly, i.e., using the same index permutations:

**Lemma 5.** *Suppose that  $\gamma \neq \tau - 1$ . For any optimal solution of (23), there exists an ordering such that both  $\alpha$  and  $\beta$  are ordered increasingly. That is, without loss of generality we may assume that*

$$\beta_1 < \beta_2 \leq \dots \leq \beta_k, \quad \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_k. \quad (38)$$

Furthermore,  $\alpha_1 = \alpha_2$ .

*Proof.* Choose arbitrarily an index  $j \neq 1$ . Let  $\beta \leq \beta_j$ , and consider the quantity

$$d\beta(1 - \gamma N_{\{k: \beta_k \leq \beta, \alpha_k + \alpha_j - 1 - d\beta < 0\}}). \quad (39)$$

(note that if  $\beta = \beta_j$  then (39) is equal to  $f(\beta_j)$ : the contribution of the term  $\beta_j$  of (31), and for  $\beta < \beta_j$  it denotes the contribution of the term  $\beta_j$  to the function  $f$  when  $\beta_j$  is set to  $\beta$ ). We prove the lemma by contradiction. Suppose there exists a  $\bar{\beta} < \beta_j$  such that  $N_{\{k: \beta_k \leq \bar{\beta}, \alpha_k + \alpha_j - 1 - d\bar{\beta} < 0\}} > 0$ . Since the number  $N_{\{k: \beta_k \leq \beta, \alpha_k + \alpha_j - 1 - d\beta < 0\}}$  is increasing in  $\beta$ , then the quantity  $(1 - \gamma N_{\{k: \beta_k \leq \beta, \alpha_k + \alpha_j - 1 - d\beta < 0\}})$  is negative for any  $\beta \in [\bar{\beta}, \beta_j]$ , as  $\gamma > 1$ . However, this contradicts the fact that  $\beta_j$  is optimal: indeed in this case, decreasing  $\beta_j$  to  $\bar{\beta}$  yields a higher contribution in (39), as  $\beta$  is negative as well.

Thus,  $N_{\{k:\beta_k \leq \beta, \alpha_k + \alpha_j - 1 - d\beta < 0\}} = 0$  for any  $\beta < \beta_j$ . Now, suppose that there exists  $i \neq 1$  such that  $\beta_i < \beta_j$ . By contradiction suppose that  $\alpha_i > \alpha_j$ . Then, also

$$N_{\{k:\beta_k \leq \beta_i, \alpha_k + \alpha_i - 1 - d\beta_i < 0\}} = 0 \quad (40)$$

for any  $\beta < \beta_j$ . In particular,

$$f(\beta_i) = \beta_i(1 - \gamma N_{\{k:\beta_k \leq \beta_i, \alpha_k + \alpha_i - 1 - d\beta_i < 0\}}) = \beta_i. \quad (41)$$

But in this case, increasing  $\beta_i$  would give a higher value of  $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$ , contradicting the hypothesis of optimality.

Therefore, it remains to prove that  $\alpha_1 \leq \alpha_i$  for all  $i \in [k] \setminus \{1\}$ . From the proof above, we can order the parameters  $\{\beta_i\}_{i \neq 1}$  in an increasing order such that also the  $\{\alpha_i\}_{i \neq 1}$  are increasingly ordered. Then we just need to prove that  $\alpha_1 \leq \alpha_2$ .

Note that  $\alpha_1$  and  $\alpha_2$  are symmetric in the optimization problem (23), as their contributions to the exponent  $f$  are

$$f(\alpha_1) = C + (1 - \tau)\alpha_1 + \gamma \min\{\alpha_1 + \alpha_2 - 1 - d\beta_2, 0\} + \sum_{j \geq 3} \gamma \min\{\alpha_1 + \alpha_j - 1 - d\beta_j, 0\}, \quad (42)$$

$$f(\alpha_2) = C + (1 - \tau)\alpha_2 + \gamma \min\{\alpha_2 + \alpha_1 - 1 - d\beta_2, 0\} + \sum_{j \geq 3} \gamma \min\{\alpha_2 + \alpha_j - 1 - d\beta_j, 0\}. \quad (43)$$

Suppose that  $\alpha_1 \neq \alpha_2$ . Then swapping  $\alpha_1$  with  $\alpha_2$  gives the same value for (23). This means that  $\alpha_1$  has multiple optima. However, the contribution of  $\alpha_1$  to  $f$  is given by

$$f(\alpha_1) = \alpha_1(1 - \tau + \gamma N_{\{k:\alpha_k + \alpha_1 - 1 - d\beta_k < 0\}}). \quad (44)$$

This contribution can have multiple optima only when  $\tau - 1 = \gamma$ , as it is linear in  $\alpha_1$ . Therefore,  $\alpha_1 \leq \alpha_2$   $\square$

We now use Lemmas 3-5 to prove Proposition 2:

*Proof of Proposition 2.* In view of Lemma 3 we can solve (23) by maximizing (30). Suppose that  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  is an optimal solution of (23), and sort them in *increasing order*, that is an order as in (38), which is possible by Lemma 5. Moreover, denote  $\beta_{\max} := \max_i \beta_i$ . We split the proof in three parts:

- (a) Proving that either  $\beta_{\max} = 0$ , or  $\beta_{\max} < 0$  and  $\alpha_i = 0$  for all  $i \in [k]$ ;
- (b) Proving that if  $\beta_{\max} < 0$  then  $\beta_i = -\frac{1}{d}$  for all  $i \in [k] \setminus \{1\}$  and  $\alpha_i = 0$  for all  $i \in k$ ;
- (c) Proving that if  $\beta_{\max} = 0$  then  $\beta_i = 0$  for all  $i \in [k] \setminus \{1\}$  and  $\alpha_i = \frac{1}{2}$  for all  $i \in k$ .

**Proof of (a).** The parameters  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  are ordered increasingly, so  $\beta_k = \beta_{\max}$  and  $\alpha_1 \leq \dots \leq \alpha_k$ . Suppose that  $\beta_k < 0$ . We focus on the contribution given by vertex  $v_k$  to  $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$  as defined in (33):

$$f(\alpha_k, \beta_k) = C + (1 - \tau)\alpha_k + d\beta_k + \sum_{j \neq k} \gamma \min\{\alpha_k + \alpha_j - 1 - d\beta_k, 0\}, \quad (45)$$

where  $\max(\beta_k, \beta_j) = \beta_k$  for all  $j$ , as we assumed that  $\beta_k$  is the largest value of  $\beta$ . Now suppose that  $\alpha_k > 0$ . Then, we can increase  $\beta_k$  while decreasing  $\alpha_k$  as follow:  $\hat{\beta}_k := \beta_k + \frac{\varepsilon}{d}$ ,  $\hat{\alpha}_k := \alpha_k - \varepsilon$ , with  $\varepsilon > 0$  small enough such that  $\hat{\beta}_k < 0$  and  $\hat{\alpha}_k \geq 0$ . After this change of parameters we obtain

$$\begin{aligned} f(\hat{\alpha}_k, \hat{\beta}_k) &= C + (1 - \tau)(\alpha_k - \varepsilon) + d(\beta_k + \frac{\varepsilon}{d}) + \sum_{j \neq k} \gamma \min\{\alpha_k - \varepsilon + \alpha_j - 1 - d(\beta_k + \varepsilon/d), 0\} \\ &= C + \tau\varepsilon + (1 - \tau)\alpha_k + d\beta_k + \sum_{j \neq k} \gamma \min\{\alpha_k + \alpha_j - 1 - d\beta_k, 0\}. \end{aligned}$$

Comparing (45) and (4) we deduce that

$$f(\hat{\alpha}, \hat{\beta}) - f(\alpha, \beta) = \tau\varepsilon > 0, \quad (46)$$

which contradicts the optimality of  $\alpha, \beta$ . Therefore, when  $\beta_{\max} < 0$  necessarily  $\alpha_k = 0$ . Then, our assumption on the ordering of  $\alpha_i$  shows that if  $\beta_{\max} < 0$  then  $\alpha_i = 0$  for all  $i \in [k]$ .

**Proof of (b).** Suppose that  $\beta_{\max} < 0$ . Then it follows from (a) that  $\alpha_i = 0$  for all  $i \in [k]$ . Denote the number of indices such that  $\beta_i = \beta_{\max}$  with  $s$ , that is,  $\beta_{k-s+1} = \dots = \beta_k = \beta_{\max}$ . We focus on the contribution to  $f(\alpha, \beta)$  by the term  $\beta_{\max}$ :

$$\begin{aligned} f(\{\beta_i\}_{i>k-s}) &= C + ds\beta_{\max} + \gamma \sum_{i>k-s, j<i} \min\{-1 - d\beta_{\max}, 0\} \\ &= C + ds\beta_{\max} + \gamma s \left( k - s + \frac{s-1}{2} \right) (-1 - d\beta_{\max}) \\ &= \tilde{C} + ds\beta_{\max} \left[ 1 - \gamma \left( k - s + \frac{s-1}{2} \right) \right]. \end{aligned} \quad (47)$$

In the second equality of (47) we exploit the independence of  $\min\{-1 - d\beta_{\max}, 0\}$  from the summation indices  $i$  and  $j$ , that is,  $s(k - s + \frac{s-1}{2})$  is the number of terms of the summation and the fact that  $\beta_{\max} \geq -\frac{1}{d}$  (from Lemma 4). Lastly, observe that in the last equality  $C$  is replaced by  $\tilde{C}$ , which encodes additional constant values.

Since  $s < k$  and  $\gamma > 1$  the quantity inside the square brackets in (47) is always negative. Therefore, decreasing  $\beta_{\max}$  to  $-\frac{1}{d}$  gives a higher contribution to  $f$ , contradicting optimality. Thus, if  $\beta_{\max} < 0$  then  $\beta_i = -\frac{1}{d}$  for all  $i \in [k] \setminus \{1\}$  and  $\alpha_i = 0$  for all  $i \in k$ .

**Proof of (c).** Let  $i$  be such that  $\beta_i = \beta_{\max} = 0$ . Then the contribution of  $\alpha_i$  to (23) is

$$f(\alpha_i) = C + (1 - \tau)\alpha_i + \gamma \sum_{j \neq i} \min\{\alpha_i + \alpha_j - 1, 0\} \quad (48)$$

The parameters  $\alpha$  are increasingly ordered. Thus, observe that:

- if  $\alpha_i > 1 - \alpha_1$ , then  $\alpha_i + \alpha_j - 1 > 0$  for all  $j \neq i$ . Therefore (48) becomes

$$f(\alpha_i) = C + (1 - \tau)\alpha_i. \quad (49)$$

In particular, changing  $\alpha_i$  to  $1 - \alpha_1$  yields a higher contribution: if  $\hat{\alpha}_i = 1 - \alpha_1$ , then

$$f(\hat{\alpha}_i) - f(\alpha_i) = (1 - \tau)(1 - \alpha_1 - \alpha_i) > 0, \quad (50)$$

which violates the hypothesis of optimality for  $\alpha_i$ . Thus,  $\alpha_i \leq 1 - \alpha_1$  when  $\beta_i = 0$ .

- if  $\alpha_i < 1 - \alpha_2$ , then  $\alpha_i + \alpha_j - 1 < 0$  for all  $j \leq m$ , and  $\alpha_i + \alpha_j - 1 \geq 0$  for all  $j > m$  (for some  $m \geq 2$ ). Therefore, (48) becomes

$$f(\alpha_i) = C + (1 - \tau)\alpha_i + \gamma \sum_{j=1}^m (\alpha_i + \alpha_j - 1) \quad (51)$$

Defining  $\hat{\alpha}_i = \alpha_i + \varepsilon$ , where  $\varepsilon$  is small enough such that  $\hat{\alpha}_i < 1 - \alpha_2$ , we have

$$f(\hat{\alpha}_i) - f(\alpha_i) = (1 - \tau + \gamma m)\varepsilon > 0. \quad (52)$$

As  $1 - \tau > -2$  and  $\gamma m > 2$ , this contradicts the hypothesis of optimality. Therefore,  $\alpha_i \geq 1 - \alpha_2$  when  $\beta_i = 0$ .

Consequently, if  $\alpha_i$  is optimal and  $\beta_i = 0$ , then  $1 - \alpha_2 \leq \alpha_i \leq 1 - \alpha_1$ . However, from Lemma 5 we know that  $\alpha_1 = \alpha_2$ . Hence,  $\alpha_i = 1 - \alpha_1$ , and  $\alpha_1 \leq \frac{1}{2}$  (because  $\alpha_j \geq \alpha_1$  for all  $j > 1$ ).

Now let  $M := \{j \in [k] : \beta_j = 0\}$  and suppose  $|M| = t$ . Let  $i \in [k]$  be such that  $\beta_i = \max_{j \in [k] \setminus M} \beta_j$ . Then, the contribution of vertex  $v_i$  to (23) is by (33)

$$\begin{aligned} f(\alpha_i, \beta_i) &= C + (1 - \tau)\alpha_i + d\beta_i + \gamma \sum_{j \in M} \min(\alpha_i + \alpha_j - 1, 0) + \gamma \sum_{j \notin M \cup \{i\}} \min(\alpha_i + \alpha_j - 1 - d\beta_i, 0) \\ &= C + (1 - \tau)\alpha_i + d\beta_i + \gamma t \min\{\alpha_i - \alpha_1, 0\} + \gamma \sum_{j \notin M \cup \{i\}} \min\{\alpha_i + \alpha_j - 1 - d\beta_i, 0\}, \end{aligned}$$

where we have used that  $\alpha_i = 1 - \alpha_1$  for all  $i \in M$ . Denote  $\hat{\beta}_i = \beta_i + \frac{\varepsilon}{d}$  and  $\hat{\alpha}_i = \alpha_i - \varepsilon$ , where  $\varepsilon$  is a quantity small enough for which  $\hat{\beta}_i < 0$ . Then,

$$\begin{aligned} f(\hat{\alpha}_i, \hat{\beta}_i) - f(\alpha_i, \beta_i) &= (\tau - 1)\varepsilon + \varepsilon + \gamma t \min(\alpha_i - \alpha_1 - \varepsilon, 0) - \gamma t \min\{\alpha_i - \alpha_1, 0\} \\ &= \varepsilon(\tau - \gamma t 1_{\{\alpha_i \leq \alpha_1\}}). \end{aligned} \quad (53)$$

If  $\alpha_i > \alpha_1$ , then the quantity in the right hand side of equation (53) is positive, and the optimality hypothesis is violated. Therefore,  $\alpha_i \leq \alpha_1$ , and in particular  $\alpha_i = \alpha_1$  (because by hypothesis we assumed that  $\alpha$  is increasingly ordered). Consequently,

$$\alpha_j = \begin{cases} \alpha_1 & \text{if } j \notin M, \\ 1 - \alpha_1 & \text{if } j \in M. \end{cases} \quad (54)$$

Now, we consider any index  $i \notin M \cup \{1\}$ , and we look at the contribution of  $\beta_i$  to  $f$  using (31),

$$f(\beta_i) = C + d\beta_i + \gamma \sum_{j: \beta_j < \beta_i} \min\{2\alpha_1 - 1 - d\beta_i, 0\}. \quad (55)$$

Suppose that  $d\beta_i > 2\alpha_1 - 1$ . Define  $\hat{\beta}_i = \beta_i - \varepsilon$ , where  $\varepsilon > 0$  is small enough such that still  $d\hat{\beta}_i > 2\alpha_1 - 1$  and  $\beta_j < \hat{\beta}_i$  for all the indices  $j$  such that  $\beta_j < \beta_i$ . Then we have

$$f(\hat{\beta}_i) - f(\beta_i) = -d\varepsilon + \gamma d\varepsilon N_{\{j: \beta_j < \beta_i\}} = d\varepsilon(\gamma N_{j: \beta_j < \beta_i} - 1), \quad (56)$$



where  $N_{j:\beta_j < \beta_i}$  denotes the number of indices  $j$  such that  $\beta_j < \beta_i$ . Note that  $d\varepsilon(\gamma N_{j:\beta_j < \beta_i} - 1) > 0$ , because the term inside the parenthesis is always positive, since  $\gamma > 1$  and  $\beta_1 = -\infty$ . Thus,  $f(\hat{\beta}_i) - f(\beta_i) > 0$ , that is, whenever  $d\beta_i > 2\alpha_1 - 1$  we can always improve the contribution to the exponent  $f$  by decreasing  $\beta_i$ . So having  $d\beta_i > 2\alpha_1 - 1$  is never optimal.

On the other hand, if  $d\beta_i \leq 2\alpha_1 - 1$  the summation in (55) is equal to 0. Hence, in this case the maximum contribution to  $f$  is achieved for the highest possible value of  $\beta_i$ , i.e., for  $\beta_i = \frac{2\alpha_1 - 1}{d}$ .

Summing up, in an optimal solution to (23),

$$\beta_i = \begin{cases} -\infty & \text{if } i = 1, \\ 0 & \text{if } i \in M, \\ \frac{2\alpha_1 - 1}{d} & \text{otherwise.} \end{cases} \quad (57)$$

Then (23) becomes a piecewise linear problem in the lone variable  $\alpha_1$ . Recalling that  $t = |M|$ ,

$$\begin{aligned} f(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= (1 - \tau)((k - t)\alpha_1 + t(1 - \alpha_1)) + (k - t - 1)(2\alpha_1 - 1) \\ &= C + \alpha_1[(k - 2t)(1 - \tau) + 2k - 2t - 2] \\ &= C + \alpha_1(k(3 - \tau) + t(2\tau - 4) - 2) \end{aligned} \quad (58)$$

where  $C$  encodes the constant terms independent from  $\alpha_1$ . Note that the quantity  $k(3 - \tau) + t(2\tau - 4) - 2$  is always positive, as  $k \geq 3$ ,  $t < k$  and  $\tau \in (2, 3)$ . Therefore,  $f(\boldsymbol{\alpha}, \boldsymbol{\beta})$  is maximized by choosing  $\alpha_1$  as large as possible, which is  $\alpha_1 = \frac{1}{2}$ .  $\square$

## 5 Proof of Theorem 1

Before proving Theorem 1, we state one last lemma showing that the standard deviation of the predominant number of cliques is significantly smaller than its mean. In particular, this condition is sufficient to prove that the number of predominant cliques converges to its mean value.

**Lemma 6.** *The number of cliques formed on vertices in  $M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$  is a self-averaging random variable. That is,*

$$\frac{\text{Var}_n(N(\mathcal{K}_k, M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)))}{\mathbb{E}_n[N(\mathcal{K}_k, M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*))]^2} = 0, \quad w.h.p. \quad (59)$$

*Proof.* For any  $\mathbf{v} = (v_1, \dots, v_k)$  we denote by  $\mathcal{E}_{\mathbf{v}}$  the event  $\mathcal{E}_{\mathbf{v}} :=$  "a clique is formed on  $(v_1, \dots, v_k)$ ". We write the variance of  $N(\mathcal{K}_k, M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*))$  as

$$\begin{aligned} \text{Var}_n \left( N(\mathcal{K}_k, M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)) \right) &= \text{Var}_n \left( \sum_{\mathbf{v} \in M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)} 1_{\mathcal{E}_{\mathbf{v}}} \right) \\ &= \sum_{\mathbf{v}, \mathbf{u} \in M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)} \text{Cov}_n(1_{\mathcal{E}_{\mathbf{v}}}, 1_{\mathcal{E}_{\mathbf{u}}}) \\ &= \sum_{\mathbf{v}, \mathbf{u} \in M(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)} \mathbb{P}_n(\mathcal{E}_{\mathbf{v}}, \mathcal{E}_{\mathbf{u}}) - \mathbb{P}_n(\mathcal{E}_{\mathbf{v}})\mathbb{P}_n(\mathcal{E}_{\mathbf{u}}) \end{aligned} \quad (60)$$

If  $\mathbf{v} \cap \mathbf{u} = \emptyset$ , then the events  $\mathcal{E}_{\mathbf{v}}$  and  $\mathcal{E}_{\mathbf{u}}$  are independent. Consequently the covariance between  $1_{\mathcal{E}_{\mathbf{v}}}$  and  $1_{\mathcal{E}_{\mathbf{u}}}$  is 0. Therefore, we can restrict ourselves to the case  $|\mathbf{v} \cap \mathbf{u}| = s \geq 1$ . In this case, we

bound  $\mathbb{P}_n(\mathcal{E}_v, \mathcal{E}_u) \leq 1$ . Then the contribution of the set of vertices such that  $|\mathbf{v} \cap \mathbf{u}| = s \geq 1$  to (60) can be bounded by

$$\sum_{\substack{\mathbf{v}, \mathbf{u} \in M(\alpha^*, \beta^*) \\ |\mathbf{v} \cap \mathbf{u}| = s}} \mathbb{P}_n(\mathcal{E}_v, \mathcal{E}_u) \leq \left| (\mathbf{v}, \mathbf{u}) \in \left( M(\alpha^*, \beta^*) \right)^2 : |\mathbf{v} \cap \mathbf{u}| = s \right|. \quad (61)$$

Suppose that  $\mathbf{v}$  and  $\mathbf{u}$  overlap on  $v_{i_1} = u_{j_1}, \dots, v_{i_s} = u_{j_s}$ , for some collection of indices  $I = \{i_1, \dots, i_s\}$ ,  $J = \{j_1, \dots, j_s\}$ . Then,

$$\sum_{\substack{\mathbf{v}, \mathbf{u} \in M(\alpha^*, \beta^*) \\ |\mathbf{v} \cap \mathbf{u}| = s}} \mathbb{P}_n(\mathcal{E}_v, \mathcal{E}_u) \leq \left| \mathbf{v} \in M(\alpha^*, \beta^*) \right| \cdot \prod_{j \notin J} \left| u_j \in V : w_{u_j} \in I_\varepsilon(n^{\alpha^*}), |x_{u_j} - x_{u_1}| = I_\varepsilon(n^{\beta^*}) \right| \quad (62)$$

where  $(\alpha^*, \beta^*) = (\frac{1}{2}, 0)$ , or  $(0, -\frac{1}{d})$ . We observe that if  $\mathbf{v}$  and  $\mathbf{u}$  overlap, without loss of generality we can assume that  $1 \in J$ , that is,  $u_1 \in (\mathbf{v} \cap \mathbf{u})$ . Indeed, if there exist indices  $i, j$  such that  $v_i = u_j$ , then after fixing  $\mathbf{v} \in M(\alpha^*, \beta^*)$  the vertex  $u_j$  is fixed (and therefore also its position is fixed). Due to the symmetry of the problem, we can permute the indices of the vector  $\mathbf{u}$ , so that the first element of  $\mathbf{u}$  is the vertex  $u_j$ . Since  $\mathbf{u} \in M(\alpha^*, \beta^*)$ , after this permutation all vertices in  $\mathbf{u}$  that are not overlapping with  $\mathbf{v}$  (i.e., whose position is not fixed after choosing  $\mathbf{v} \in M(\alpha^*, \beta^*)$ ) will have their position in a neighborhood of size  $\Theta(n^{\beta^*})$  of  $u_j$ .

Then, we can rewrite (62) as

$$\sum_{\substack{\mathbf{v}, \mathbf{u} \in M(\alpha^*, \beta^*) \\ |\mathbf{v} \cap \mathbf{u}| = s}} \mathbb{P}_n(\mathcal{E}_v, \mathcal{E}_u) \leq |M(\alpha^*, \beta^*)| \cdot \Theta_{\mathbb{P}} \left( n^{1+(1-\tau)\alpha^*+d\beta^*} \right)^{k-s} = \Theta_{\mathbb{P}} \left( n^{1+(1-\tau)\alpha^*+d\beta^*} \right)^{2k-s} \quad (63)$$

Finally, we recall that  $\mathbb{E}_n[N(\mathcal{K}_k, M(\alpha^*, \beta^*))] = \Theta_{\mathbb{P}} \left( n^{k+k(1-\tau)\alpha^*+(k-1)d\beta^*} \right)$ , from which we deduce

$$\text{Var}_n(N(\mathcal{K}_k, M(\alpha^*, \beta^*))) = \mathbb{E}_n[N(\mathcal{K}_k, M(\alpha^*, \beta^*))]^2 o(1). \quad (64)$$

□

We are now ready to prove our main result, Theorem 1:

*Proof of Theorem 1.* Plugging the solution  $\alpha^*, \beta^*$  provided by Proposition 2 into (22) shows that the maximum value of  $f(\alpha, \beta)$  is

$$f(\alpha^*, \beta^*) = \max \left\{ \frac{3-\tau}{2} k, 1 \right\}. \quad (65)$$

For  $k > \frac{2}{3-\tau}$ , the maximum is attained at  $\frac{3-\tau}{2} k$ , and  $(\alpha^*, \beta^*)$  is as in (24). For  $k < \frac{2}{3-\tau}$  on the other hand, it is attained at 1, and  $(\alpha^*, \beta^*)$  is defined as in (25). This provides two cases for  $k$  that distinguish the two possible optima:  $k < \frac{2}{3-\tau}$ , or  $k > \frac{2}{3-\tau}$ . In particular, except for the threshold case  $k = \frac{2}{3-\tau}$ , the solution to (23) is unique.

Consequently, since  $\mathbb{E}_n[N(\mathcal{K}_k, M(\alpha, \beta))] = \Theta_{\mathbb{P}} \left( n^{f(\alpha, \beta)} \right)$ ,

$$\mathbb{E}_n[N(\mathcal{K}_k, M(\alpha, \beta))] = \begin{cases} \Theta_{\mathbb{P}} \left( n^{\frac{3-\tau}{2} k} \right) & \text{if } k > \frac{2}{3-\tau} \\ \Theta_{\mathbb{P}}(n) & \text{if } k < \frac{2}{3-\tau} \end{cases} \quad (66)$$

If  $(\alpha, \beta) \neq (\alpha^*, \beta^*)$ , then  $f(\alpha, \beta) < f(\alpha^*, \beta^*)$ . Moreover,

$$\frac{E_n[N(\mathcal{K}_k, M^{(\alpha, \beta)})]}{E_n[N(\mathcal{K}_k, M^{(\alpha^*, \beta^*)})]} = \frac{\Theta_{\mathbb{P}}(n^{f(\alpha, \beta)})}{\Theta_{\mathbb{P}}(n^{f(\alpha^*, \beta^*)})} = 0 \quad \text{w.h.p.}, \quad (67)$$

proving the first part of Theorem 1, and showing that the mean number of cliques with vertices of weights and positions  $\alpha^*, \beta^*$  is indeed predominant among the others.

Lastly, by Lemma 6 and Chebyshev's inequality, it follows that

$$N(\mathcal{K}_k, M^{(\alpha^*, \beta^*)}) = E_n[N(\mathcal{K}_k, M^{(\alpha^*, \beta^*)})](1 + o(1)), \quad (68)$$

which proves (8). □

**Acknowledgements.** This work is supported by an NWO VENI grant 202.001.

## References

- [1] K. E. Avrachenkov and A. V. Bobu. Cliques in high-dimensional random geometric graphs. *Applied Network Science*, 5:92, 2020.
- [2] G. Bianconi and M. Marsili. Number of cliques in random scale-free network ensembles. *Physica D: Nonlinear Phenomena*, 224(1):1–6, 2006. Dynamics on Complex Networks and Applications.
- [3] T. Bläsius, T. Friedrich, M. Katzmann, U. Meyer, M. Penschuck, and C. Weyand. Efficiently generating geometric inhomogeneous and hyperbolic random graphs. In M. A. Bender, O. Svensson, and G. Herman, editors, *27th Annual European Symposium on Algorithms (ESA 2019)*, volume 144 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [4] T. Bläsius, T. Friedrich, and A. Krohmer. Cliques in hyperbolic random graphs. *Algorithmica*, 80(8):2324–2344, 2018.
- [5] K. Bogerd, R. M. Castro, and R. van der Hofstad. Cliques in rank-1 random graphs: The role of inhomogeneity. *Bernoulli*, 26(1):253 – 285, 2020.
- [6] K. Bringmann, R. Keusch, and J. Lengler. Geometric inhomogeneous random graphs. *Theoretical Computer Science*, 760:35–54, 2019.
- [7] L. Devroye, A. György, G. Lugosi, F. Udina, et al. High-dimensional random geometric graphs and their clique number. *Electronic Journal of Probability*, 16:2481–2508, 2011.
- [8] M. Doležal, J. Hladký, and A. Máthé. Cliques in dense inhomogeneous random graphs. *Random Structures & Algorithms*, 51(2):275–314, 2017.
- [9] T. Gavenčiak. <https://github.com/gavento/girg-sampling>, 2020.
- [10] L. Gugelmann, K. Panagiotou, and U. Peter. Random hyperbolic graphs: degree sequence and clustering. *39th International Colloquium on Automata, Languages, and Programming, ICALP*, pages 573–585, 2012.

- [11] R. van der Hofstad, J. S. H. van Leeuwaarden, and C. Stegehuis. Optimal subgraph structures in scale-free configuration models. *The Annals of Applied Probability*, 31(2):501 – 537, 2021.
- [12] S. Janson, T. Łuczak, and I. Norros. Large cliques in a power-law random graph. *Journal of Applied Probability*, 47(4):1124–1135, 2010.
- [13] A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and S. Shneer. Counting Cliques and Cycles in Scale-Free Inhomogeneous Random Graphs. *Journal of Statistical Physics*, 175:161–184, 2019.
- [14] M. Latapy. Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theoretical Computer Science*, 407(1):458–473, 2008.
- [15] M. Penschuck, C. Weyand, and T. Bläsius. <https://github.com/chistopher/girgs>, 2019.
- [16] T. Schank and D. Wagner. Finding, counting and listing all triangles in large graphs, an experimental study. In S. E. Nikolettseas, editor, *Experimental and Efficient Algorithms*, pages 606–609, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.
- [17] C. Stegehuis. Distinguishing power-law uniform random graphs from inhomogeneous random graphs through small subgraphs. *arXiv:2102.09315*, 2021.
- [18] C. Stegehuis, R. van der Hofstad, and J. S. van Leeuwaarden. Variational principle for scale-free network motifs. *Scientific reports*, 9(1):1–10, 2019.