Detecting planted structures in random graphs

Detecting planted structures in random graphs

Detecting planted structures in random graphs

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-5211-5

Printed by Gildeprint Drukkerijen, Enschede

Detecting planted structures in random graphs

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof. dr. ir. Frank Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op 17 februari 2021 om 16:00 uur

door

Kay Martin Bogerd

geboren te Amsterdam

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof. dr. J.J. Lukkien
promotor:	prof. dr. R. van der Hofstad
co-promotor:	dr. R.M. Castro
leden:	prof. dr. G. Lugosi (Universitat Pompeu Fabra)
	prof. dr. E. Arias-Castro (UC San Diego)
	prof. dr. F.C.R. Spieksma
	dr. N. Verzelen (INRAE, Montpellier SupAgro)
	prof. dr. A.P. Zwart
	prof. dr. N.V. Litvak

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Acknowledgements

This thesis would not have been possible without the tremendous help and support of my supervisors, colleagues, friends and family. I would like to use this opportunity to thank them.

First and foremost, I would like to express my deepest gratitude to my two supervisors, Remco van der Hofstad and Rui Castro. Thank you for all the guidance and support. You have always been a constant source for new inspiration, yet have also given me the freedom to pursue my own ideas. I have been very fortunate with you as my supervisors. Remco, I have benefited immensely from your deep insight, thoroughness, and passion for research. I have learned a lot from you. Rui, thank you so much for your enthusiasm, kind personality and constant encouragement to go one step further. I thoroughly enjoyed our weekly meetings about work and everything else.

To all the members of the doctoral committee: Gabor Lugosi, Ery Arias-Castro, Frits Spieksma, Nicolas Verzelen, Bert Zwart and Nelly Litvak. Thank you for being part of my defense committee and for taking the time to read this thesis. A special thank you to Nicolas, your invaluable insights helped to significantly improve the contents of Chapter 4.

Many thanks to the stochastics group as a whole for maintaining an ever pleasant and stimulating atmosphere. I wish to thank my former office mates Gianmarco, Jori and Debankur for being most welcoming and making me feel at home when I started four years ago. Gianmarco, it has been a great pleasure to collaborate with you on two of the chapters in this thesis and I very much enjoyed your return visits to Eindhoven. Thanks to Richard, Marta, Ellen, Maliheh for being fantastic office mates and creating the best work environment I could wish for, our many coffee breaks and all the talks. Special thanks to Richard, who was there every morning on platform 18 for the long train commute to Eindhoven. I thoroughly enjoyed all our conversations about work, teaching, and practically everything else. I also want to thank Bart, Joost, Mark, Rik and Youri for all the interesting research discussions and the fun talks during lunch.

To the administrative staff Chantal, Ellen, and Petra, thank you for all the practical support, and to Patty for the wonderful organization of Eurandom that has allowed me to attend many conferences and workshops.

I would also like to show my appreciation to the people I know outside work, and that have supported me over the past years. To Anne, Gijs, Christiaan and Peter, it was a pleasure to start our studies together and all the good times we have enjoyed since. To all my friends from Maarssen: Martyn, Vincent, Folkert, Jouke, Daan, Harmen, Tessa, Bastiaan, Merel, Vincent and Michel. Most of you I already know since high-school and I am glad we are still friends. Thanks for all the parties, holidays, and the good times in general. Philip, I am very grateful for our many years of friendship and thank you for agreeing to be a paranymph during my thesis defense ceremony.

A very special word of thanks goes to my parents, Jan and Yvonne, and my sister Susanne. You are always there for me with help and good care, I would not be who I am today without your continuous love and support. I am also very appreciative of the support from Joey.

Last, but certainly not least, I would like to thank my awesome wife, Tineke. We have traveled the world together and discussed many of the ideas in this thesis. Your support has been paramount to the completion of this work. I am truly grateful to have shared so many unique experiences with you, and I look forward to create many more in the future. Thank you so much, dear.

Contents

Chapter 1

Introduction

Many complex systems can be modeled as a network, which is simply a collection of objects and the relations between them. The objects are typically referred to as *vertices* and the connections between them are called *edges*. Networks have become a useful and flexible representation for a wide variety of systems in different areas, such as social, biological, technological, and communication sciences. For example, social networks consist of people that are connected by friendship or acquaintance, and the world wide web is a network where many webpages are connected by hyperlinks. Furthermore, with advances in computational power it has become increasingly possible to work with huge data sets, and this has subsequently made it feasible to analyze extremely large networks. For instance, the internet, the world wide web, the brain, and many social networks consist of well over a billion vertices each. The need to deal with such a large number of objects is partly what makes the study of networks so interesting.

Networks representing real systems are usually not regular and have many inhomogeneities, with various parts of the network having radically different structure. These *structures* often contain important information about the network and the relation between its various components. For instance, many networks contain a few important vertices, often referred to as hubs, that have many more connections than a typical vertex. Another example is a network that consists of several *communities*. These are densely connected groups of vertices, with significantly more internal connections between vertices in the same community than external connections that connect a vertex inside the community to a vertex outside of it. Communities often arise because vertices in the same community have something in common. For example, communities could be groups of friends or family in a social network, or in the world wide web it could be a collection of webpages dealing with the same topic. Even more generally, a network could contain one or more *components* or *anomalies* that exhibit a different local structure or that contain another connectivity pattern.

Identifying such structures in a network can provide valuable new insights about the network and the objects within. Therefore, we would like to know when this is possible. We formalize this using *random graphs*. These are network models that are commonly used as a baseline when studying real-world networks. Such models make it possible to perform statistical analyses and test whether there is an actual structure present or whether, instead, our observations can also be attributed to random noise present in the baseline model.

In this thesis, we study the detection of structures in networks for several different random graph models and we identify when it is possible to discern communities or anomalies from the natural variability present in these models. Furthermore, we also study random graphs directly, yielding novel insights about their structure.

In this introductory chapter we present the main subjects used throughout this thesis. We start by defining some well-known random graph models and related properties. We then introduce the statistical framework that is commonly used to perform hypothesis testing and explain how this can be applied to network problems. This chapter is concluded with a brief summary of contributions and an outline of this thesis.

1.1 Random graphs

Networks are often represented by a graph, which is simply an ordered pair $G = (V, E)$ consisting of a set of vertices V representing the objects, and a set of edges $E \subseteq \{(i, j) : i, j \in V\}$ defining relationships between pairs of objects or vertices. Sometimes these relations can also have a direction associated with them, in which case the resulting graphs are called directed graphs. However, throughout this thesis we only consider undirected graphs, where edges do not have a direction. In other words, in this thesis edges always indicate a bidirectional relationship. The size of a graph is the number of vertices, which we always denote by $|V| = n$. Furthermore, a *complete* graph is a graph where all possible edges are present. On the other hand, an empty graph is a graph without any edges. The degree of a vertex is the number of edges incident with it. In particular, in a complete graph with n vertices every vertex has degree $n - 1$.

In many cases, one only observes a single instance of a network. To still be able to analyze these networks, one often turns to *random graphs*, which are models that specify a probability distribution over all possible graphs. This makes it possible to see how likely certain properties are to arise, and to test whether the observed graph is, in some sense, special. In this way, random graphs are commonly used as a baseline or null model, making statistical analyses of networks possible. In the models we consider the vertices will often be labeled $V = [n] = \{1, \dots, n\}$ and only the relations between them are random, with each random graph model specifying different probabilities and sometimes dependence among the potential edges.

Below we discuss several common random graph models and explain how they are related. Each of these models is different in order to capture other characteristics observed in real-world networks. This results in models with different properties and varying amount of complexity.

1.1.1 Erdős-Rényi random graph

Probably the most studied and also the simplest random graph model is the Erdős-Rényi random graph [70, 90]. This model is denoted by $\mathbb{G}(n, p)$ and has two parameters: the graph size $n \in \mathbb{N}$ and an edge probability $p \in [0, 1]$. A graph sampled from this model has $|V| = n$ vertices, where each pair of vertices $i, j \in V$ is connected independently with probability p.

Because all edges are added with the same probability, this model is sometimes called *homogeneous* in the sense that the vertex degrees tend to take values in a narrow range. Note that the degree of every vertex is approximately $(n-1)p$ with only little variability. Furthermore, because every pair of vertices is connected independently there is no correlation between edges, and this causes every part of the graph to "look" similar, and with little structure.

The edge probability $p = p_n$ is frequently also allowed to depend on the graph size n. When taking $p = c/n$, for some constant $c > 0$, this leads to what is called the *sparse* regime where the average degree converges to a constant as the graph size n tends to ∞. This is in contrast to the case where p is fixed, in which case one obtains a so-called *dense* graph where the average degree grows linearly with the graph size. Choices of p in between these regimes give rise to graphs of intermediate density.

1.1.2 Inhomogeneous random graph

A natural way to extend the Erdős-Rényi model is to allow for more inhomogeneity of the vertex degrees as is often observed in real-world networks. An obvious approach that provides this inhomogeneity is to allow $p = p_{ij}$ to vary for every pair of vertices $i, j \in V$. However, this model is often too general and more restrictions are needed to obtain meaningful results. To this end, we consider the so-called inhomogeneous random graphs. This model will be denoted by $\mathbb{G}(n; \kappa, \ell_n)$, where $n \in N$ is the graph size, $\kappa(\cdot, \cdot)$ is a symmetric measurable function called the kernel that controls the inhomogeneity, and ℓ_n is a scaling parameter controlling the edge density. To generate a graph from this model, every vertex $i \in V$ is first assigned a weight w_i. Often, these weights are assumed to be in $[0, 1]$. Then, every pair of vertices is connected independently with probability

$$p_{ij} = \min\left(\frac{\kappa(w_i, w_j)}{\ell_n}, 1\right). \tag{1.1}$$

Sometimes the weights w_i are assumed to be random variables instead. In this case, one first samples the weight of every vertex, and then conditionally on these weights the edges are sampled conditionally independently according to (1.1).

When ℓ_n is constant the resulting graphs are dense, with an average degree that grows linearly with the graph size. This leads to the theory of graphons introduced by Lovász and Szegedy [120]. For a detailed description of this theory we refer the reader to Lovász [119]. On the other hand, if $\ell_n = n$ and the kernel is integrable then we are in the so-called sparse case where the resulting graph has bounded average degree. This case was first studied in detail by Bollobás, Janson, and Riordan [31].

A special case of this model is the so-called *rank-1 inhomogeneous random graph* where $\kappa(w_i, w_j) = w_i w_j$. Many well-known random graphs fit in this framework, such as the Erdős-Rényi model when each vertex has the same weight. Furthermore, when $\ell_n = n$ one obtains the Chung-Lu model [53, 54, 55], or closely related variations such as the Norros-Reittu and generalized random graph by replacing the minimum by a soft minimum instead [43, 146]. These models have received much attention over the past decade as they accurately capture some of the inhomogeneity observed in many real-world networks while remaining mathematically tractable.

Another variation of this model is the *stochastic block model* [106], also called planted partition model in computer science. The stochastic block model is obtained from the general inhomogeneous random graph model when $\kappa(\cdot, \cdot)$ can only take on finitely many values. In this case, vertices can be seen as being assigned one of $k \leq n$ groups. Then, the probability of connecting two vertices depends solely on the groups these vertices belong to. Typically, vertices belonging to the same group will have a large probability of connecting, whereas vertices from different groups have a smaller connection probability. In this way, the stochastic block model can generate graphs with a prescribed community structure, with edges being more prevalent between members of the same group.

1.1.3 Random geometric graph

Many real-world networks have, or are believed to have, an underlying spatial structure. This can be the physical locations of the vertices but it can also be more abstract and encode some other form of similarity between the vertices. For instance, friendships are more likely between individuals with similar interests. So, we can imagine that each vertex is endowed with some set of attributes, and edges are more likely to exist between vertices with similar attributes.

To model such networks with spatial structure the geometric random graph was introduced by Gilbert [91], and later studied in detail by Penrose [151]. This model assigns every vertex $i \in V$ to a location x_i in some metric space, typically a d-dimensional cube or sphere. Note that, these locations are latent quantities and are usually not observed. We connect every pair of vertices $i, j \in V$ if their Euclidean distance $d(x_i, x_j)$ is less than some threshold radius parameter r. Similarly to the

previous models, we frequently let $r = r_n$ decrease with the graph size. In this way, the connection radius can be used as a tuning parameter controlling the sparsity of the resulting graphs.

In this model, when two vertices are connected they need to be quite close to each other. Therefore, it is much more likely that two vertices that share a common neighbor are themselves connected. This type of dependence is commonly referred to as *clustering* and can be observed in many real-world networks. Because of this, the random geometric graph has much more structure than, for example, the Erdős-Rényi random graph. However, in many aspects these models are still similar. In particular, the degree of vertices from a geometric random graph still tend to take values in a relatively narrow range, making the resulting graphs rather homogeneous. It turns out that it is the Euclidean geometry that causes the graphs to become homogeneous and a different geometry can result in more inhomogeneity. In particular, the hyperbolic random graph can be highly inhomogeneous and have a power-law degree distribution while retaining the clustering property [29, 94, 117, 156].

1.1.4 Preferential attachment model

The final model we introduce is somewhat different from the previous ones. Until now, we have discussed static models where the vertex set is fixed. However, in the preferential attachment model the graph is dynamically grown by adding one new vertex at a time, and connecting it to the existing network using simple local connection rules [16, 68]. The main success of this model comes from the recognition that these local connection rules are able to explain important macroscopic features observed in real-world networks. For example, many real-world networks are *scale-free* which roughly means that they have a power-law degree sequence, and they are *small-worlds* which means that the typical distance between vertices in these networks is quite small. Both these properties are often observed in real-world networks. For example, the internet [71], the world wide web [4, 44], or scientific collaboration networks [15, 141].

There exist various versions of the preferential attachment model, each following a slightly different convention for adding new vertices. Here we consider the following model. Given two parameters $m \geq 1$ and $\delta \geq -m$, this model generates a sequence of graphs $(G_t)_{t=1}^n$, from which we consider only the final snapshot G_n. The first graph G_1 consists of two vertices v_0 and v_1 connected by m edges. For $2 \leq t \leq n$, the graph G_t is constructed from G_{t-1} by adding a vertex v_t. This vertex has m edges, and these are added one by one and with intermediate updating of degrees. To this end, define $G_{t,0}$ as the graph G_{t-1} together with the vertex v_t without any edges, and let $G_{t,1}, G_{t,2}, \dots, G_{t,m}$ be the intermediate graphs for each of the m edges emanating from v_t. For $1 \leq i \leq m$, the graph $G_{t,i}$ is constructed from $G_{t,i-1}$ by connecting v_t to a randomly selected vertex $v_s \in \{v_0, \dots, v_{t-1}\}$. This is where the parameter δ comes in,

because the probability that the ith edge of v_t connects to v_s is given by

$$\mathbb{P}\big(v_{t,i} \leftrightarrow v_s \,|\, G_{t,i-1}\big) = \frac{\deg_{v_s}(G_{t,i-1}) + \delta}{\sum_{j=0}^{t-1}(\deg_{v_j}(G_{t,i-1}) + \delta)}, \tag{1.2}$$

where $\deg_{v_s}(G_{t,i-1})$ denotes the degree of v_s in $G_{t,i-1}$. After all m edges have been added to the vertex v_t we obtain the graph $G_t = G_{t,m}$.

As can be seen, the newly appearing vertex is more likely to connect to vertices that already have large degrees, thus making these degrees even larger. This behavior is called preferential attachment and generates highly inhomogeneous graphs with a power-law degree distribution [16, 33, 104]. This property can also be obtained in, for example, inhomogeneous random graphs by choosing the appropriate model parameters. However, the preferential attachment model has the appealing characteristic that this follows from a simple local connection rule.

1.2 Cliques and dense subgraphs

An important concept in the study of graphs is that of a *clique*, which is also called a complete subgraph. This is a subset of vertices $C \subseteq V$ such that every pair of vertices $i, j \in C$ is connected. Another way to define a clique is via the notion of induced subgraph. Given a graph $G = (V, E)$, an induced subgraph is another graph, formed from a subset of the vertices $C \subseteq V$ and all of the edges connecting them, that is $\{(i, j) : i, j \in C\} \cap E$. Thus, C is a clique when the subgraph induced by C is a complete graph.

Cliques arise naturally in the context of sociology, where they are commonly used to model communities of people that are all mutually acquainted [6, 121, 122, 133]. For example, these can arise as a model for groups of friends in a social network. However, this definition is sometimes too strict for real-world applications and one would like to consider a more general class of dense subgraphs that allows for a couple of missing edges. This motivated research into relaxations of the notion of a clique. One of the most popular of these relaxations is the quasi-clique [2, 3]. Given $\gamma \in [0, 1]$, a subset of vertices C is called a *quasi-clique* if it contains at least a fraction γ of all possible edges. That is, C is a quasi-clique if the subgraph induced by C contains at least $\gamma\binom{|C|}{2}$ edges. Note that, for $\gamma = 1$, the definition of a quasi-clique coincides with the definition a clique. While cliques are a good model for tight-knit communities, such as groups of friends, the definition of a quasi-clique is more loose and might be better suited as a model for a group of acquaintances.

One is often interested in the largest communities within a given network, or more precisely, one would like to know the size of the largest clique or quasi-clique in a given graph. However, these are well-known to be computationally hard problems in general [101, 115, 149]. Therefore, one usually resorts to studying the size of

the largest clique or quasi-clique in a typical or random graph, using the theory of random graphs. To this end, we define the clique number $\omega(G)$ as the largest clique in a graph G. Similarly, the size of the largest γ-quasi-clique in the graph G is called the γ-quasi-clique number and is denoted by $\omega_\gamma(G)$. It is then common to study the behavior of $\omega(G)$ or $\omega_\gamma(G)$ when the graph G is sampled from one of the random graph models described earlier. It turns out that the clique number $\omega(G)$, as well as the quasi-clique number $\omega_\gamma(G)$, are very well concentrated in many of the popular random graph models: Erdős-Rényi random graph [13, 93, 125, 126], inhomogeneous random graph [26, 27, 66], and random geometric graph [137]. This suggests a more general underlying principle that could be interesting to investigate further. See Chapter 7 for a discussion of this issue.

1.3 Hypothesis testing

In this section we introduce the principal subject of this thesis, the problem of *hypothesis testing* on graphs. As is often the case in statistical inference, one starts with a set of observations. These are the values taken on by some random variable (or in our case random graph) whose distribution $\mathbb{P}_\theta$ is not known, except that we assume that the *unknown* parameter θ lies inside a parameter space Ω. Our goal is then to use these observations to infer some additional information about the distribution $\mathbb{P}_\theta$. In particular, we will focus on hypothesis testing. Here we consider two mutually exclusive classes, called the null hypothesis H_0 and alternative hypothesis H_1, and their corresponding disjoint subsets of the parameter space Ω_{H_0} and Ω_{H_1}, so that $\Omega_{H_0} \cup \Omega_{H_1} = \Omega$ and $\Omega_{H_0} \cap \Omega_{H_1} = \emptyset$. Mathematically, our goal is then to use the observations to decide which of these hypotheses is true, or at least more likely. More specifically, we want to decide whether the unknown parameter θ belongs to Ω_{H_0} or Ω_{H_1}. Below we introduce the formal framework that is used to make hypothesis testing mathematically rigorous.

1.3.1 Testing problems on graphs

Our primary focus will be on combinatorial testing problems involving graphs, meaning that we assume that the graph itself is the quantity we observe. In this case, the null hypothesis H_0 is typically assumed to be a random graph model, for example one of the models described in Section 1.1. The alternative hypothesis H_1 is often largely similar to the null model, with only small parts of the graph being different. Think of a graph possibly containing one or more anomalies, then the null hypothesis and the alternative hypothesis will be largely the same except on small parts of the graph where the vertices could belong to the anomaly. This type of problem has received much attention in the literature over the past decade, see for example [5, 9, 39, 98, 124, 136]. Closely related to this is the planted clique problem, where the anomaly is a clique. That is, one receives a sample from some random graph and has to decide

whether a clique has been added on top of this graph [8, 11, 74].

The main difficulty in all of these problems stems from the fact that one does not know where the anomaly is located. In these problems, the class H_0 typically contains only a single distribution, so that the null hypothesis H_0 is completely specified. We call such a hypothesis *simple*. On the other hand, the class of alternatives H_1 contains several distributions. This is called a *composite* hypothesis. In our case, this class contains many closely related distributions, one for each possible location of the anomaly. Because of this, the class of alternatives will often be indexed by subsets of vertices and thus has a combinatorial structure.

1.3.2 Performance of a test

When performing a hypothesis test one has to make the decision of either accepting or rejecting the null hypothesis. To this end, define a test $T_n(g) \in \{0, 1\}$ as any measurable function taking as input an observed graph g on n vertices. If $T_n(g) = 0$, then there is enough reason to believe that the null hypothesis H_0 is true and we say that the test accepts the null hypothesis; otherwise, when $T_n(g) = 1$, the alternative hypothesis H_1 is deemed correct and we say that the null hypothesis is rejected in favor of the alternative hypothesis.

There are various ways of measuring the performance of a test, and most of these are based on the two types of error one can make [145]. The type-I error is rejecting the null hypothesis when it is true. Symmetrically, the type-II error is accepting the null hypothesis when in fact one of the alternative hypotheses in H_1 is true. A common way of measuring the quality of a test T_n is by using the so-called *worst-case* risk [10, 67, 99]. This is defined as the sum of the maximal probability of type-I and type-II error, that is

$$R(T_n) \coloneqq \mathbb{P}_0(T_n(G) \neq 0) + \sup_{\theta \in \Omega_{H_1}} \mathbb{P}_\theta(T_n(G) \neq 1). \tag{1.3}$$

where $\mathbb{P}_0(\cdot)$ is the distribution of the random graph model under the null hypothesis, and $\mathbb{P}_\theta(\cdot)$ are the distributions under the alternative hypothesis specified by $\theta \in \Omega_{H_1}$.

In most cases, as the notation already suggests, we will analyze the performance of a sequence of tests T_n for a sequence of graphs g_n as the graph size n tends to ∞. However, we will often refer to such a sequence of tests simply as a test when this causes no confusion. We then say that a test is *asymptotically powerful* when it has vanishing risk, that is $R(T_n) \to 0$ as $n \to \infty$. This means that the probability of making an error becomes smaller as the graphs become larger. Thus an asymptotically powerful test is able to distinguish between the null hypothesis and the alternative hypothesis with a high degree of certainty, provided that the observed graph is large enough. On the other hand, we call a test *asymptotically powerless* when $R(T_n) \to 1$ as $n \to \infty$. This means that the test does not perform substantially better than any random guess that would completely ignore the observed graph.

1.3.3 Minimax lower bounds

After one has found an asymptotically powerful test, it is also desirable to know whether there can exist another test that can do significantly better. Specifically, given a test T_n, we would like to show that in any scenario where T_n is not asymptotically powerful there cannot exist any other test that is. To this end, we want to compare the risk of the test T_n with the theoretically best possible risk, also called the *minimax* risk and is given by

$$R^\star \coloneqq \inf_{T_n} R(T_n) = \inf_{T_n} \sup_{\theta \in \Omega_{H_1}} \mathbb{P}_0(T_n(G) \neq 0) + \mathbb{P}_\theta(T_n(G) \neq 1), \tag{1.4}$$

where the infimum is taken over all measurable functions $T_n : G \mapsto \{0, 1\}$. We would then like to know whether $R(T_n)$ is close to the minimax risk $R^\star$, in which case there cannot exist any other test that performs significantly better than T_n. This is a well-posed problem, but often intractable. Therefore one typically tries to find a lower-bound for the minimax risk $R^\star$ instead. The standard approach for this is to first reduce the composite alternative hypothesis to a simple one by considering the average risk [10, 17, 118], given by

$$\begin{aligned} \overline{R}(T_n) &\coloneqq \mathbb{P}_0(T_n(G) \neq 0) + \frac{1}{|\Omega_{H_1}|} \sum_{\theta \in \Omega_{H_1}} \mathbb{P}_\theta(T_n(G) \neq 1) \\ &= \mathbb{P}_0(T_n(G) \neq 0) + \mathbb{P}_1(T_n(G) \neq 1), \end{aligned} \tag{1.5}$$

where $\mathbb{P}_1(\cdot)$ denotes the average over all distributions in the alternative hypothesis H_1, and for simplicity we assumed that Ω_{H_1} is finite. Note that it is not always the best approach to consider the average as in (1.5) and in some cases it might be beneficial to consider another distribution or weighted average over the class of alternatives instead. The important part is that we obtain a lower bound on the worst-case risk, that is $R(T_n) \geq \overline{R}(T_n)$. For many problems the class of alternatives exhibits some kind of symmetry, so this lower bound is often quite sharp.

Since the average risk can be interpreted as a hypothesis test between two simple hypotheses, it follows from the Neyman-Pearson lemma that the test that minimizes the average risk $\overline{R}(T_n)$ is the so-called likelihood ratio test [118, 144], given by

$$T_n^{\mathrm{LR}}(g) = \mathbb{1}_{\{L(g) > 1\}}, \tag{1.6}$$

where $L(g) = \mathbb{P}_1(G = g)/\mathbb{P}_0(G = g)$ is the likelihood ratio of the two simple hypotheses from (1.5). Following this approach, we have shown that the worst-case risk of any test T_n is greater than or equal to the average risk of the likelihood ratio test T_n^{LR}. That is, for any test T_n we have $R(T_n) \geq \overline{R}(T_n^{\mathrm{LR}})$. In particular, this means that the minimax risk is bounded by $R^\star \geq \overline{R}(T_n^{\mathrm{LR}})$. Therefore, to show that no test can be asymptotically powerful we do not have to analyze the risk of every possible test, but

instead we can simply analyze the average risk of the likelihood ratio test. Although the likelihood ratio test is generally complicated, there exists analytical machinery that allows us to effectively analyze its performance. It is then enough to show that $\overline{R}(T_n^{\mathrm{LR}})$ remains bounded away from 0, because then no test can be asymptotically powerful. Furthermore, if one can show the stronger result that $\overline{R}(T_n^{\mathrm{LR}}) \to 1$ then this implies that all tests are asymptotically powerless.

1.4 Overview of results

The work in this thesis centers around the analysis of community detection methods for inhomogeneous networks, as well the detection of other types of anomalies such as testing for the presence of a botnet.

We study existing community detection methods that have so far predominantly been used in the homogeneous setting, and show how these can be extended to a setting of inhomogeneous random graphs. This leads to new insights about properties of the inhomogeneous random graphs we study, and we show how a certain community detection method can be extended to the inhomogeneous setting in an optimal manner. The insights obtained from this also sparked the interest to consider a related project about the detection of botnets. Lastly, we also consider dynamically growing graphs using the preferential attachment model, where we investigate when it is possible to detect a change in the attachment dynamics of the graph. Each of these projects is described in more detail below.

1.4.1 Detecting a planted clique or dense subgraph

The aim of this project was to answer the question: "What is the smallest community that can theoretically be detected in an already inhomogeneous graph?". Although this problem has been resolved for Erdős-Rényi random graphs [11, 12], we were the first to address this question in the inhomogeneous setting.

The initial work on this project resulted in novel insights about inhomogeneous random graphs and, in particular, about the behavior of cliques in these graphs. Remarkably, the size of the largest clique always takes on one of two consecutive integers with high probability, and this two-point concentration remains true even when the graph is highly inhomogeneous. This is presented in Chapter 2. Some of the ideas from that chapter could also be leveraged to extend these results to quasi-cliques. This is what we do in Chapter 3, where we show that also the size of the largest quasi-clique is well concentrated in a large class of dense inhomogeneous random graphs.

In Chapter 4 we answer our initial question and identify the smallest community that can be detected in an inhomogeneous random graph. Surprisingly, inhomogeneity can be make the detection problem easier than in the homogeneous setting. As a result of the inhomogeneity, some parts of the graph are more informative than

others, and this can be used to create a test that is more powerful than it would have been in a homogeneous graph with the same edge density. Furthermore, we proved that our test is optimal in the sense that it is impossible for any other test to detect significantly smaller communities.

1.4.2 Detecting a planted botnet

Communities are typically modeled as more densely connected subgraphs within a graph, but one is sometimes also interested in subgraphs that are different in other ways. For example, one could consider an anomaly, such as a botnet, that tries to mask its presence by not making too many connections. However, the connectivity structure or underlying geometry of the botnet could still be rather different than that of normal people or normal vertices, which can be used to detect its presence. In Chapter 5 we formalized this idea as a testing problem with the goal of detecting the presence of a homogeneous component in an otherwise random geometric graph. Here we introduce two different tests that can both detect such a botnet, even when the botnet is very small. Furthermore, we show that our tests are asymptotically optimal.

Whereas community detection has already received quite some attention in the literature, this type of problem has not. Because of this, there are still many variations and open problems that could be studied related to this project. We discuss several of these in Chapter 7.

1.4.3 Changepoint detection in the preferential attachment model

Many networks are dynamic and change over time, with some rules concerning the evolution or growth of the graph. We consider the preferential attachment model and study what happens when the attachment function changes after some time. In particular, we investigate whether it is possible to detect a single change in the attachment function. When this change happens early in the graph generation process then it was already known that such a change can be detected [14, 21]. In Chapter 6 we focus on the case where the attachment function changes at a very late time. We show that, also in this case, it is possible to detect that the attachment function has changed, even if the change happens when there are only approximately $\sqrt{n}$ vertices missing from the final network.

1.4.4 Summary

In this thesis we create several methods for the detection of communities and other types of anomalies in networks. To formalize these questions we design new models and these lead to novel and interesting statistics. Furthermore, we develop new insights in the community structure and behavior of cliques in inhomogeneous random graphs.

1.5 General assumptions and notation

Throughout this thesis we are interested in an asymptotic characterization of random graphs and their properties as the number of vertices n increases. When limits are unspecified, they are assumed to be taken as the number of vertices n tends to ∞.

We use standard asymptotic notation: $a_n = O(b_n)$ when $|a_n/b_n|$ is bounded, $a_n = \Omega(b_n)$ when $b_n = O(a_n)$, $a_n = \Theta(b_n)$ when $a_n = O(b_n)$ and $a_n = \Omega(b_n)$, $a_n = o(b_n)$ when $a_n/b_n \to 0$, and $a_n \asymp b_n$ when $a_n = (1 + o(1))b_n$. We also use the probabilistic versions of these: $a_n = O_{\mathbb{P}}(b_n)$ when $|a_n/b_n|$ is stochastically bounded, $a_n = \Omega_{\mathbb{P}}(b_n)$ when $b_n = O_{\mathbb{P}}(a_n)$, $a_n = \Theta_{\mathbb{P}}(b_n)$ when $a_n = O_{\mathbb{P}}(b_n)$ and $a_n = \Omega_{\mathbb{P}}(b_n)$, and $a_n = o_{\mathbb{P}}(b_n)$ when a_n/b_n converges to 0 in probability. In addition, we say that a sequence of events holds with high probability if they holds with probability tending to 1.

Finally, for two numbers $a, b \in \mathbb{R}$, we write $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, and $[a]_+ = \max\{a, 0\}$. Finally, the integral part or integer part of $a \in \mathbb{R}$ is denoted by $\lfloor a \rfloor$.

1.6 Thesis outline

The remaining chapters of this thesis are based on separate journal publications. The contents of these chapters is mostly identical to the published version, except for Chapter 6 which is currently in preparation. In Chapter 7 we conclude this thesis by discussing several open problems and possible extensions to this work.

Chapter 2

Cliques in rank-1 inhomogeneous random graphs

Based on:
Cliques in rank-1 random graphs: the role of inhomogeneity,
K. Bogerd, R. M. Castro, and R. van der Hofstad,
Bernoulli 26.1 (2020), pp. 253–285.

We study the asymptotic behavior of the clique number in rank-1 inhomogeneous random graphs, where edge probabilities between vertices are roughly proportional to the product of their vertex weights. We show that, in many regimes, the clique number is concentrated on at most two consecutive integers, for which we provide an expression. Interestingly, the order of the clique number is primarily determined by the overall edge density, with the inhomogeneity only affecting multiplicative constants or adding at most a $\log\log(n)$ multiplicative factor. For sparse enough graphs the clique number is always bounded and the effect of inhomogeneity completely vanishes.

2.1 Introduction

The clique number of a graph G is the size of the largest clique (i.e. the largest complete subgraph) in G. In an Erdős-Rényi random graph, edges between vertices are present with the same probability independently of one another. This is sometimes also called the *homogeneous* setting because all edges have the same probability of being included. In this setting, it is well known that the clique number is highly concentrated when the graph has a large number of vertices, meaning that with high probability the clique number takes values in a small interval [93, 125, 126]. In fact, Matula [125] shows that the clique number converges to one of two consecutive in-

tegers, and provides an explicit formula for the asymptotic clique size.

In this work, we are interested in understanding the behavior of the clique number in *inhomogeneous* random graphs, where edges have different occupation probabilities. In such random graphs, the properties of different vertices (e.g. their expected degree) can be radically different and can take a wide range of values. This is in contrast, for instance, with Erdős-Rényi random graphs, where degrees can only take values in a relative narrow range.

Our work is in part motivated by the statistical problem of community detection. Formally, this amounts to testing whether a given graph was obtained by "planting" a clique, or dense subgraph, inside a random graph. Arias-Castro and Verzelen [11, 12] have recently considered this problem with an Erdős-Rényi random graph as the underlying model. To extend these results to the inhomogeneous setting, one needs a better understanding of cliques in the corresponding null model; thus, studying the clique number in inhomogeneous random graphs is a natural starting point.

Related work. Inhomogeneous random graphs have received much attention over the past decade because they more accurately model the network structure observed in many real-world networks. The literature on this subject can be divided into sparse and dense graphs.

In the sparse setting, the edge probabilities decrease with the graph size such that the resulting graph has bounded average degree. This setting was first studied in substantial detail by Bollobás, Janson, and Riordan [31] in which the critical value for the existence of the giant component was established, as well as several related fundamental properties of such graphs were derived. This has sparked great interest in this model, see also [104, 105] and the references therein for an overview of recent results.

The dense setting (when the average degree is unbounded) leads to the theory of graphons developed by Lovász and Szegedy [120]. Recently, first order results for the clique number were also obtained for this case by Doležal, Hladký, and Máthé [66], and further studied by McKinley [128].

Inhomogeneous random graphs with an intermediate density have received less attention, although recently results about connectivity have been obtained by Devroye and Fraiman [64], and the diameter was considered by Fraiman and Mitsche [80].

A special class of the inhomogeneous random graphs above are the so-called rank-1 random graphs. Here each vertex receives a weight and, conditionally on these weights, edges are present independently with probability equal to the product of their vertex weights. Many well-known random graphs fit this model, such as the Erdős-Rényi random graph by giving each vertex the same weight or scale-free graphs such as the Chung-Lu, Norros-Reittu, and Generalized random graphs by taking the weights from a power-law [43, 53, 54, 55, 146].

Our contribution. In this chapter we show that the clique number of rank-1 inhomogeneous random graphs is concentrated on at most two consecutive integers, provided that all vertex weights are bounded away from 1. We provide a single expression for the order of the clique number that is valid for every edge density, bringing together results of both the sparse and dense regimes.

To derive our results we essentially make use of the same methodology as Matula [125], namely using the first and second moment methods to obtain, respectively, upper and lower bounds for the clique number. The main contribution here lies in the definition of what we call the typical clique number ω_n, which is the point where the clique number concentrates around. This quantity is defined implicitly, and we show that this is indeed a sound definition. Furthermore, the inhomogeneity of these graphs substantially complicates the derivation of the lower bounds, which now requires significantly more effort than for Erdős-Rényi random graphs.

We find quite different asymptotic behaviors of the clique number depending on the edge density of the graph, although our results are more interesting when the average degree diverges. In sparse graphs, the clique number is always bounded regardless of the "amount of inhomogeneity", and the only parameter that affects the asymptotic clique number is the edge density. In dense graphs, the clique number behaves similarly as in an Erdős-Rényi random graph. Specifically, the clique number is always of order $\log(n)$, with the inhomogeneity only affecting the constants. Interestingly, graphs with intermediate edge density can be rather different, with the inhomogeneity sometimes adding a $\log\log(n)$ multiplicative factor to the clique number.

2.2 Main results

In this chapter, we consider a random graph model denoted by $\mathbb{G}(n; W, \lambda_n)$. This model has three parameters: the number of vertices n, the *weight distribution* W, and the *scaling* λ_n. An element of $\mathbb{G}(n; W, \lambda_n)$ is a simple graph $G = (V, E)$ that has $n \in \mathbb{N}$ vertices with vertex set $V = [n] \coloneqq \{1, \dots, n\}$, and a random edge set E. Each vertex $i \in V$ is assigned a *weight*, which is an independent copy W_i of the non-negative random variable $W \in [0, \infty)$. In other words, W_i are i.i.d. non-negative random variables with the same distribution as W. Conditionally on these weights, the presence of an edge between two vertices $i, j \in V$, with $i \neq j$, is modeled by independent Bernoulli random variables with success probability

$$p_{ij} \coloneqq \mathbb{P}\big((i,j) \in E \,\big|\, (W_k)_{k\in V}\big) = \frac{W_i}{\lambda_n} \cdot \frac{W_j}{\lambda_n} \wedge 1, \tag{2.1}$$

where the scaling $\lambda_n : \mathbb{N} \mapsto \mathbb{R}$ is a deterministic sequence. Note that the weights do not depend on the graph size n. This is why the introduction of the scaling λ_n is useful, as it allows us to naturally control the edge density of the graph. We assume

that the scaling is at most of order $\sqrt{n}$, because otherwise we are in the trivial case where the graph is asymptotically almost or completely empty.

Random graph models like the classical Erdős-Rényi random graphs are *homogeneous* in the sense that for a typical realization the degrees of all vertices tend to take a narrow range of values. Furthermore, all parts of the graph look more or less the same. However, graphs arising in real-world settings do not generally satisfy this property and tend to be *inhomogeneous*, with a relatively wide range of different vertex degrees across the entire graph. In our model the weight distribution W determines the inhomogeneity of the graph, and the heavier the tails of this distribution the more inhomogeneous the graph is. Recall that the weight distribution is not a function of the graph size and without any scaling factor the resulting graphs are dense (i.e. with a number of edges that is quadratic in the graph size n). The parameter λ_n allows us, therefore, to control the edge density. When λ_n is constant, we are in the dense regime. On the other hand, when $\lambda_n \approx \sqrt{n}$, we are in the sparse regime with a number of edges that is linear in n, which corresponds to graphs with finite average degree. Choices of λ_n in between those extremes lead to graphs of intermediate density with a number of edges more than linear but less than quadratic in n.

2.2.1 The clique number

Our main contribution is to show that the clique number of a graph $G \sim \mathbb{G}(n; W, \lambda_n)$, denoted by $\omega(G)$, is concentrated on at most two consecutive integers provided that the following assumptions hold:

Assumption 2.1. There exists $\delta > 0$ such that

$$\mathbb{P}\left(\max_{i \in V} W_i \leq \frac{\lambda_n}{1+\delta}\right) \to 1, \qquad \text{as } n \to \infty. \tag{2.2}$$

This assumption, which seems relatively benign, ensures that all edge probabilities are bounded away from 1 with high probability. Alternatively, it can be regarded as a restriction on the denseness of the graph, requiring that λ_n grows fast enough, which causes the resulting graphs not to become too dense. Our second assumption strengthens the above for large λ_n.

Assumption 2.2. If $\liminf_{n\to\infty} \log(\lambda_n)/\log(n) > 0$, then for every $\eta > 0$

$$\mathbb{P}\left(\max_{i \in V} W_i \leq \frac{\lambda_n}{1+\eta}\right) \to 1, \qquad \text{as } n \to \infty. \tag{2.3}$$

Note that this assumption is only a restriction when the scaling is large (i.e. when λ_n is a positive power of n). Moreover, in many cases, Assumption 2.2 is a direct consequence of Assumption 2.1 (e.g. when all or enough moments of W are finite, or when W has a regularly-varying distribution). We only need Assumption 2.2 to eliminate some pathological cases. This issue is discussed in more detail in Section 2.4.

The clique number in graphs from our model depends both on the amount of inhomogeneity (captured by W) and the average edge density (which is close to $(\mathbb{E}[W]/\lambda_n)^2$). Under Assumptions 2.1 and 2.2, it turns out that the relation between the conditional moments of the weights fully characterizes the asymptotic clique number. To simplify notation define the *truncated weight* $\widetilde{W}$ as the random variable with distribution

$$\mathbb{P}\left(\widetilde{W} \leq x\right) = \mathbb{P}\left(W \leq x \,\middle|\, W \leq \frac{\lambda_n}{1+\delta}\right), \qquad \text{for all } x \in \mathbb{R}, \tag{2.4}$$

where $\delta > 0$ comes from Assumption 2.1. In other words, the distribution of $\widetilde{W}$ is the conditional distribution of W given $W \leq \lambda_n/(1+\delta)$. The relative truncated moments (abbreviated to relative moments in the rest of this chapter) are defined as follows:

Definition 2.1 (Relative moments)**.** Given a weight W and scaling λ_n, the r-th relative moment is defined by

$$c_{n,r} = \frac{\mathbb{E}\left[W^r \,\middle|\, W \leq \frac{\lambda_n}{1+\delta}\right]}{\mathbb{E}\left[W \,\middle|\, W \leq \frac{\lambda_n}{1+\delta}\right]^r} = \frac{\mathbb{E}[\widetilde{W}^r]}{\mathbb{E}[\widetilde{W}]^r}. \tag{2.5}$$

Note that the relative moments $c_{n,r}$ depend on the graph size n, but only through the scaling λ_n. To avoid notational clutter, we often omit the explicit dependence of $c_{n,r}$ on δ.

Building towards our result stating that the clique number is highly concentrated, we define next the *typical clique number*. Unfortunately, the typical clique number depends in a cumbersome way on the relative moments $c_{n,r-1}$. Therefore, it is only possible to give an implicit characterization in the general setting.

Definition 2.2 (Typical clique number)**.** Let $\omega_n \in [1, \infty)$ denote the solution in $r \geq 1$ of

$$r = \frac{\log(n) - \log(r) + \log(c_{n,r-1}) + 1}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} + 1\,. \tag{2.6}$$

We call ω_n the typical clique number of $\mathbb{G}(n; W, \lambda_n)$.

Note that the typical clique number ω_n needs not be an integer. Also, it is not immediately obvious that ω_n is well defined because there could either be no solution or (2.6) might have multiple solutions. However, the following lemma shows that the typical clique number ω_n is well defined:

Lemma 2.1. *Under Assumption 2.1 the typical clique number ω_n from Definition 2.2 exists and is unique.*

The following theorem is our main result and shows that asymptotically almost all graphs generated by our model have a clique number that differs at most one from the typical clique number ω_n:

Theorem 2.1. *Let $\varepsilon > 0$ be arbitrary. Under Assumptions 2.1 and 2.2 the clique number $\omega(G_n)$ of a random graph $G_n \sim \mathbb{G}(n; W, \lambda_n)$ satisfies*

$$\lim_{n\to\infty} \mathbb{P}\left(\omega(G_n) \in \left[\lfloor \omega_n - \varepsilon \rfloor, \lfloor \omega_n + \varepsilon \rfloor\right]\right) = 1, \tag{2.7}$$

where ω_n is the typical clique number from Definition 2.2.

It is important to note that the typical clique number ω_n depends on the δ from Assumption 2.1 through the behavior of the truncated weights $\widetilde{W}$. This might give the impression that the clique number $\omega(G_n)$ of a graph G_n must also depends on δ, which it obviously does not. However, provided δ is small enough to ensure that Assumption 2.1 holds, the dependence of the typical clique number ω_n on δ vanishes. This is further discussed in Section 2.4 below.

Theorem 2.1 shows that the clique number converges to at most one of two possible values with high probability, provided we take $\varepsilon < 1/2$. This shows two-point concentration of the clique number for rank-1 inhomogeneous random graphs. To find the explicit values of these two points, we need to find an explicit solution of (2.6), which is generally difficult, see Section 2.6 for the details. To facilitate this, we give two alternative asymptotic characterizations of the typical clique number ω_n.

Lemma 2.2. *Under Assumption 2.1 the typical clique number ω_n is equal to the solution in r of*

$$r = \log_b(nc_{n,r-1}) - \log_b\log_b(nc_{n,r-1}) + \log_b(\mathrm{e}) + 1 + o(1)\,, \tag{2.8}$$

where we abbreviate $b = \lambda_n/\mathbb{E}[\widetilde{W}]$.

Note that when the weight distribution is degenerate (i.e. has probability 1 on a single point) we obtain an Erdős-Rényi random graph, and $c_{n,r-1} = 1$ for all $r \in \mathbb{N}$. Using Lemma 2.2, our result in Theorem 2.1 reduces to the main result in [125]. On the other hand, when we consider inhomogeneous graphs, Lemma 2.2 shows that we essentially have to rescale the number of vertices n by the relative moments $c_{n,r-1}$ to account for the inhomogeneity.

The second characterization pertains the setting where the scaling is such that it gives rise to relatively sparse graphs. In this case, many of the edge probabilities have become so small that the shape of the distribution W stops playing a role, and the typical clique number ω_n converges to a constant independent of the weight distribution:

Lemma 2.3. *Let $\alpha \in (0, 1)$. Under Assumption 2.1 the following are equivalent:*

(i) The scaling satisfies $\lambda_n = n^{\alpha+o(1)}$.

(ii) The typical clique number satisfies $\omega_n = 1 + 1/\alpha + o(1)$.

This result states that when the typical clique number ω_n converges to a constant, the scaling λ_n is essentially a power of n, and the converse is also true.

It is important to note that we required some assumptions to show two-point concentration of the clique number. A natural question to ask is whether these assumptions are strictly speaking necessary. Although we cannot formally make this statement, we can argue that Assumption 2.1 cannot be significantly relaxed: consider for instance a graph $G_n \sim \mathbb{G}(n; W, \lambda_n)$ where the weights have a positive probability $\rho > 0$ of becoming larger than the scaling λ_n, that is $\mathbb{P}(W \geq \lambda_n) = \rho > 0$. Then the vertices belonging to these weights form a clique because the probability of an edge between any of these vertices equals 1. Hence, the clique number will have approximately a binomial distribution, that is $\omega(G_n) \sim \text{Bin}(n, \rho)$, and we cannot expect the clique number $\omega(G)$ to be concentrated on any fixed length interval.

This shows that Assumption 2.1 defines a rather sharp threshold. Below this threshold the clique number $\omega(G)$ is at most logarithmic and highly concentrated, whereas above this threshold the clique number has polynomial size and cannot be concentrated on any fixed length interval.

2.3 Examples

Theorem 2.1 shows that the typical clique number ω_n must be very close to the clique number $\omega(G_n)$ of a graph $G_n \sim \mathbb{G}(n; W, \lambda_n)$. However, Definition 2.2 does not give an explicit expression for ω_n, but rather it gives an implicit definition as the solution of the fixed-point equation (2.6). Nevertheless, we may derive the asymptotic behavior of ω_n for several interesting choices of weights W and scalings λ_n, illustrating the different regimes one might encounter. Note that in most cases these derivations are far from trivial, see Section 2.6 for the details. Interestingly, in all the examples that we consider, the typical clique number ω_n is primarily determined by the scaling, namely $\omega_n \approx k_n \log_{\lambda_n/\mathbb{E}[W]}(n)$ where k_n is typically just a constant but can be as large as $O(\log \log n)$.

The maximum weight in the graph plays a crucial role in Assumption 2.1 and it is directly related to the tail probabilities of the weight distribution by the relation

$$\mathbb{P}\left(\max_{i \in V} W_i \leq x\right) = \left(1 - \mathbb{P}\left(W > x\right)\right)^n. \tag{2.9}$$

Because of this relation we find that the tail behavior of the weight distribution plays a key role in the asymptotic behavior of the clique number, and we identify three main classes of weight distributions based on this.

When the weight distribution has bounded support, the behavior of the clique number is very similar to an Erdős-Rényi random graph. For weights with unbounded support, the behavior of the clique number depends on how heavy the tails are. For weights with heavy-tailed distributions, the scaling must grow roughly as a power of n to ensure that Assumption 2.1 is satisfied. This restriction on the scaling makes the graph highly sparse, which causes the effect of inhomogeneity due to the weight

distribution to disappear. Interestingly, when the weights have a light-tailed distribution, the behavior of the clique number strongly depends on the scaling λ_n with different regimes depending on how λ_n is chosen.

2.3.1 Weights with bounded support

In this section, we consider the clique number $\omega(G_n)$ for graphs $G_n \sim \mathbb{G}(n; W, \lambda_n)$ with weight distributions W that have bounded support. The best-known example in this class is the Erdős-Rényi random graph. In our model, this corresponds to a degenerate weight distribution W, with all the mass at 1 (denoted by Degen(1)). Note that Assumption 2.1 is trivially satisfied by taking $\lambda_n \geq s$ for any constant $s > 1$, and the edge probability is simply $p_n = 1/\lambda_n^2$. In this case, the relative moments are $c_{n,r-1} = 1$ for all $1 \leq r \leq n$, and we immediately see from Lemma 2.2 that

$$\begin{aligned}\omega_n &= \log_{\lambda_n}(n) - \log_{\lambda_n}\log_{\lambda_n}(n) + \log_{\lambda_n}(\mathrm{e}) + 1 + o(1)\\ &= 2\log_{1/p_n}(n) - 2\log_{1/p_n}\log_{1/p_n}(n) + 2\log_{1/p_n}(\mathrm{e}/2) + 1 + o(1). \qquad (2.10)\end{aligned}$$

This result was also obtained by Matula [125].

For other weight distributions, vertices with large weights are more likely to be in the largest clique than vertices with small weights. This idea can be used to show that the first order behavior of the clique number remains as in (2.10) but with $p_n = (w_{\max}/\lambda_n)^2$ and where $w_{\max}$ is the supremum of the support of W. Therefore, for weights with bounded support, the first order behavior of the clique number remains unchanged when we replace the random weights W by the maximum of their support $w_{\max}$. This happens because vertices with small weight have, asymptotically, a negligible probability of being part of the largest clique.

To see this, note that the edge probabilities are bounded by $p_{ij} \leq (w_{\max}/\lambda_n)^2$ for all $i \neq j \in V$. Plugging this into (2.10) gives the following high probability upper bound on the clique number $\omega(G_n)$ of a graph $G_n \sim \mathbb{G}(n; W, \lambda_n)$,

$$\omega(G_n) \leq (1 + o(1))\frac{\log(n)}{\log(\lambda_n/w_{\max})}. \qquad (2.11)$$

To obtain a matching lower bound we can use the following simple heuristic. Instead of considering the whole graph, consider the subgraph induced by the vertices with large enough weights, in particular the subgraph induced by the vertices $U_n = \{i \in V : W_i > t_n\}$ for some t_n. On this subgraph all weights are larger than t_n, and therefore we can bound the edge probability by $p_{ij} \geq (t_n/\lambda_n)^2$ for all $i \neq j \in U_n$. Since $|U_n| \approx n\mathbb{P}(W > t_n)$ we can use (2.10) to obtain the following high probability lower bound on the clique number,

$$\omega(G_n) \geq (1 + o(1))\frac{\log(|U_n|)}{\log(\lambda_n/t_n)} = (1 + o(1))\frac{\log(n) + \log(\mathbb{P}(W > t_n))}{\log(\lambda_n/t_n)}. \qquad (2.12)$$

Note that this lower bound holds for every t_n, so we can find an optimal t_n that maximizes the right-hand side of (2.12). For weights with bounded support, taking $t_n = w_{\max} - o(1)$ suffices, provided that the $o(1)$ term vanishes slowly enough to ensure that $\log \mathbb{P}(W > t_n) = o(\log(n))$. This gives

$$\omega(G_n) = (1 + o(1))\frac{\log(n)}{\log(\lambda_n/w_{\max})}. \tag{2.13}$$

This is precisely the leading order behavior in (2.10), but with $p_n = (w_{\max}/\lambda_n)^2$.

Determining the asymptotics of the clique number $\omega(G_n)$ more accurately requires significantly more effort, because the argument above no longer suffices and one needs to actually solve (2.6) from Definition 2.2. In Table 2.1 the typical clique number is shown for some weight distributions. As explained above, the first order behavior is the same in all these examples. However, it can clearly be seen that weights with less mass around the maximum of their support have a smaller second order term, as expected.

Table 2.1: The asymptotic behavior of the typical clique number ω_n for some weights W with support on $[0, 1]$. Here $\Gamma(\cdot)$ denotes the gamma function. See Section 2.6 for the derivation of these results.

W	Typical clique number ω_n
Degen(1)	$\log_{\lambda_n}(n) - \log_{\lambda_n}\log_{\lambda_n}(n) + \log_{\lambda_n}(\mathrm{e}) + 1 + o(1)$
Ber(q)	$\log_{\lambda_n}(nq) - \log_{\lambda_n}\log_{\lambda_n}(nq) + \log_{\lambda_n}(\mathrm{e}) + 1 + o(1)$
Unif(0, 1)	$\log_{\lambda_n}(n) - 2\log_{\lambda_n}\log_{\lambda_n}(n) + \log_{\lambda_n}(\mathrm{e}) + 1 + o(1)$
Beta(α, β)	$\log_{\lambda_n}(n) - (1+\beta)\log_{\lambda_n}((1+\beta)\log_{\lambda_n}(n)) + \log_{\lambda_n}(\mathrm{e}) + \log_{\lambda_n}(\Gamma(\alpha+\beta)/\Gamma(\alpha)) + 1 + o(1)$

2.3.2 Weights with light tails

In this section, we consider the clique number $\omega(G_n)$ for graphs $G_n \sim \mathbb{G}(n; W, \lambda_n)$ with weight distributions W that have unbounded support but light tails. This is arguably the most interesting setting, and where the effect of inhomogeneity is the most pronounced. For such weight distributions the maximum weight $M_n \coloneqq \max_{i \in V} W_i$ is typically highly concentrated around its expectation $\mathbb{E}[M_n]$. Therefore, we can choose any scaling λ_n slightly larger than $\mathbb{E}[M_n]$ to satisfy Assumption 2.1. For this class of distributions we observe two very distinct behaviors depending on the choice of scaling λ_n.

The slowest scaling that still ensures that Assumption 2.1 is satisfied is $\lambda_n \approx (1 + \varphi)\mathbb{E}[M_n]$, with $\varphi > 0$. For a given weight distribution, this is the densest graph

for which we can apply Theorem 2.1. In this case, we see that the shape of the weight distribution has a real impact on the asymptotic behavior of the typical clique number ω_n, as shown in Table 2.2. In other words, the asymptotic behavior of the clique number depends on the chosen weight distribution, and this amounts to more than constant multiplicative factors in the various terms.

We consider three distributions, namely the half-normal, the Gamma and the log-normal. For both the half-normal and Gamma distribution we see that the typical clique number is of order $\log(n)$, whereas in an Erdős-Rényi random graph with the same edge density the cliques are smaller, of order $\log(n)/\log\log(n)$. For log-normal weights the first order behavior of the typical clique number ω_n is the same in the corresponding Erdős-Rényi random graph although with different constants. Note that the effect of inhomogeneity is much weaker for log-normal weights. Because the log-normal distribution is "nearly" heavy tailed, this is consistent with our findings in the next section, where we show that, for heavy-tailed weights, the specific shape of the distribution is not relevant anymore.

If the scaling is such that Assumption 2.1 is more easily satisfied (and therefore resulting in sparser graphs) then the contribution of the weight distribution becomes far less prominent. In particular, we consider $\lambda_n \approx \mathbb{E}[M_n]^{1+\varphi}$, with $\varphi > 0$. This choice of scaling leads to behavior that is qualitatively similar to that in Section 2.3.1. As soon as $\varphi > 0$, the resulting graphs become so sparse that the shape of the weight distribution has no severe impact on the asymptotic behavior of the typical clique number ω_n, and only multiplicative factors are affected. This can be seen in Table 2.3.

Table 2.2: The asymptotic behavior of the typical clique number ω_n for some light-tailed weights W and scaling $\lambda_n \approx (1+\varphi)\mathbb{E}[M_n]$ with $\varphi > 0$ arbitrary. For comparison we include the clique number of an Erdős-Rényi random graph with the same edge density, that is $p_n = (\mathbb{E}[W]/\lambda_n)^2$. Here $\Gamma(\cdot)$ is the gamma function, and we write $\xi_k(\varphi) = -k/\mathcal{W}_{-1}\big(-1/(\mathrm{e}(1+\varphi)^k)\big) \in (0,1)$, where $\mathcal{W}_{-1}(\cdot)$ is the lower branch of the Lambert-W function, see (2.39). See Section 2.6 for the derivation of these results.

W	λ_n	Typical clique number ω_n
$\lvert\mathrm{N}(0,\sigma)\rvert$	$(1+\varphi)\sqrt{2\sigma^2\log(n)}$	$(1+o(1))\,\xi_2(\varphi)\log(n)$
Comparable Erdős-Rényi graph		$(1+o(1))\,2\log(n)/\log\log(n)$
$\mathrm{Gamma}(\alpha,\beta)$	$(1+\varphi)\log(n)/\beta$	$(1+o(1))\,\xi_1(\varphi)\log(n)$
Comparable Erdős-Rényi graph		$(1+o(1))\,\log(n)/\log\log(n)$
$\mathrm{LN}(0,1)$	$(1+\varphi)\exp(\sqrt{2\log(n)})$	$(1+o(1))\sqrt{2\log(n)}$
Comparable Erdős-Rényi graph		$(1+o(1))\sqrt{(1/2)\log(n)}$

Table 2.3: The asymptotic behavior of the typical clique number ω_n for some light-tailed weights W and scaling $\lambda_n \approx \mathbb{E}[M_n]^{1+\varphi}$ with $\varphi > 0$ arbitrary. For comparison we include the clique number of an Erdős-Rényi random graph with the same edge density, that is $p = (\mathbb{E}[W]/\lambda_n)^2$. See Section 2.6 for the derivation of these results.

W	λ_n	Typical clique number ω_n
$\lvert\mathrm{N}(0,\sigma)\rvert$	$\sqrt{2\sigma^2\log(n)}^{1+\varphi}$	$(1+o(1))\,(2/\varphi)(\log(n)/\log\log(n))$
Comparable Erdős-Rényi graph		$(1+o(1))\,(2/(1+\varphi))(\log(n)/\log\log(n))$
$\mathrm{Gamma}(\alpha,\beta)$	$\log(n)^{1+\varphi}/\beta$	$(1+o(1))\,(1/\varphi)(\log(n)/\log\log(n))$
Comparable Erdős-Rényi graph		$(1+o(1))\,(1/(1+\varphi))(\log(n)/\log\log(n))$
$\mathrm{LN}(0,1)$	$\exp\big(\sqrt{2\log(n)}\big)^{1+\varphi}$	$(1+o(1))\,\big((1+\varphi)-\sqrt{\varphi(2+\varphi)}\big)\sqrt{2\log(n)}$
Comparable Erdős-Rényi graph		$(1+o(1))\,(1/(1+\varphi))\sqrt{(1/2)\log(n)}$

The heuristic to obtain a high probability lower bound on the clique number, as explained in the previous section, also remains valid for light-tailed distributions. Interestingly, also in this case the lower bound seems to be tight. That is, for the t_n that maximises the lower bound in (2.12), we find exactly the same behavior, including the same constants, of the clique number as in Tables 2.2 and 2.3. However, because the weights are no longer bounded from above, we are not aware of a simple method to obtain a matching upper bound. Nevertheless, we strongly suspect that this heuristic also gives the correct first order behavior of the clique number for other light-tailed distributions.

2.3.3 Weights with heavy tails

In this section we consider the clique number $\omega(G_n)$ for graphs $G_n \sim \mathbb{G}(n; W, \lambda_n)$ with weight distributions W that have heavy tails, which we define as distributions whose moments are not all finite. For these distributions, finding the clique number is surprisingly straightforward. To apply Theorem 2.1 we need a scaling λ_n such that Assumptions 2.1 and 2.2 are satisfied, and we necessarily have $\lambda_n \geq n^{\alpha+o(1)}$ for some $\alpha > 0$. This means that for heavy-tailed distributions we can always apply Lemma 2.3, which shows that the typical clique number ω_n is bounded and completely determined by the scaling.

A notable special case of this was treated in [109] and [24, 25], where the clique number in scale-free graphs with a model similar to ours was considered. In those works the weights have a power-law distribution and the scaling is chosen as $\lambda_n = \sqrt{n}$. The authors find that the clique number asymptotically becomes either 2 or 3 when the variance of the weights is finite. Using Lemma 2.3 we first determine that $\omega_n \to 3$, since the scaling is $\lambda_n = \sqrt{n}$. Therefore, it follows from Theorem 2.1 that,

asymptotically, the clique number must be either 2 or 3, precisely the same result. Note that for this scaling, having weights with finite variance and Assumptions 2.1 and 2.2 are equivalent.

In the highly inhomogeneous case, where the weights have infinite variance, we require a scaling of $\lambda_n \geq n^{\alpha+o(1)}$ for some $\alpha > 1/2$ in order to satisfy Assumptions 2.1 and 2.2. However, when the scaling is this large, the resulting graphs are asymptotically almost or completely empty. On the other hand, as explained at the end of Section 2.2, when $\alpha \leq 1/2$ the clique number will approximately have a binomial distribution and thus cannot concentrate on any fixed length interval.

2.4 Discussion and overview

In this section we remark on our results and discuss some possibilities for future work.

Typical clique number. Let us first remark on our main result, Theorem 2.1, that shows that the clique number $\omega(G_n)$ of a graph G_n and the corresponding typical clique number ω_n must be very close. As explained in Section 2.2 the typical clique number ω_n still depends on the δ from Assumption 2.1, whereas the clique number $\omega(G_n)$ of a graph G_n obviously does not. This is certainly not desirable, and it should be possible to define the typical clique number ω_n independently of δ. In all examples that we considered in Section 2.3 this is indeed possible, since in those examples:

$$\frac{\mathbb{E}[\widetilde{W}^r]}{\mathbb{E}[\widetilde{W}]^r} = \frac{\mathbb{E}\Big[W^r \,\Big|\, W \leq \frac{\lambda_n}{1+\delta}\Big]}{\mathbb{E}\Big[W \,\Big|\, W \leq \frac{\lambda_n}{1+\delta}\Big]^r} = (1+o(1))\frac{\mathbb{E}[W^r]}{\mathbb{E}[W]^r}, \qquad \text{for all } r \leq \omega_n. \tag{2.14}$$

We conjecture that a similar statement should hold in general, or at least for a wide class of weights W and scalings λ_n. When proven, this would imply that the truncation in Definitions 2.1 and 2.2 can be ignored. This would solve the issue of the seeming dependence between the typical clique number ω_n and the δ from Assumption 2.1, and at the same time, make explicit computations of the typical clique number ω_n somewhat easier.

Connection between Assumptions 2.1 and 2.2. Most of our results only require Assumption 2.1, but to prove our main result we require the slightly stronger Assumption 2.2. However, in most cases that we checked, Assumption 2.2 is implied from Assumption 2.1, and the choice of weight distribution W and scaling λ_n.

Suppose that $\lambda_n \geq n^{\alpha+o(1)}$ for some $\alpha \in (0,1)$. Then we necessarily need to have $\mathbb{E}[W^{1/\alpha}] < \infty$ in order to satisfy Assumption 2.1. When a slightly larger moment of W is also finite, that is $\mathbb{E}[W^{1/\alpha+\varepsilon}] < \infty$ for some $\varepsilon > 0$, then both assumptions are simultaneously satisfied. To see this, note that by the moment condition we have

$\mathbb{P}(W^{1/\alpha+\varepsilon} > n) \leq o(1/n)$. Now take any $\eta > 0$, then for n large enough

$$\begin{aligned}\mathbb{P}\left(W > \frac{\lambda_n}{1+\eta}\right) &\leq \mathbb{P}\left(W > \frac{n^{\alpha+o(1)}}{1+\eta}\right) \\ &\leq \mathbb{P}\big(W > n^{\alpha/(1+\alpha\varepsilon)}\big) = \mathbb{P}(W^{1/\alpha+\varepsilon} > n) \leq o(1/n). \end{aligned} \tag{2.15}$$

Hence, both Assumptions 2.1 and 2.2 are simultaneously satisfied.

Alternatively, Assumption 2.1 is also sufficient when W is regularly varying of index $\beta < 0$. In this case, for any $\eta > 0$, we have

$$\frac{\mathbb{P}\left(W > \frac{\lambda_n}{1+\eta}\right)}{\mathbb{P}\left(W > \frac{\lambda_n}{1+\delta}\right)} = (1+o(1))\Big(\frac{1+\eta}{1+\delta}\Big)^{-\beta}. \tag{2.16}$$

By Assumption 2.1 we have $\mathbb{P}\big(W > \lambda_n/(1+\delta)\big) = o(1/n)$; therefore, we also have $\mathbb{P}\big(W > \lambda_n/(1+\eta)\big) = o(1/n)$. Hence, Assumption 2.2 is also satisfied.

Different models. In our model, the edge probabilities are $p_{ij} = \min(X_{ij}, 1)$, where $X_{ij} = (W_i/\lambda_n) \cdot (W_j/\lambda_n)$. We require the minimum because otherwise some edge probabilities could exceed 1. To achieve the same effect one has other options; some common examples are $\hat{p}_{ij} = 1 - \exp(-X_{ij})$ or $\tilde{p}_{ij} = X_{ij}/(1+X_{ij})$. Changing the model in this manner does not have a significant influence on the asymptotic clique number, provided Assumption 2.1 holds. To see this, note that we can bound the edge probabilities of these models by $\min(X_{ij}/2, 1) \leq \hat{p}_{ij}, \tilde{p}_{ij} \leq \min(X_{ij}, 1)$ with high probability. Obviously, the clique number is then also bounded by the clique numbers obtained from the models with edge probabilities as given in these bounds. Since these bounds differ only by a constant multiplicative factor, it is easily seen from Definition 2.2 that both lead to the same leading order asymptotics of the clique number when the scaling is diverging. When the scaling is constant, the situation is more subtle and the precise clique number will change by a multiplicative factor that depends on the specific model considered.

Instead of the change in truncation described above, we could also consider different interactions between the weights. We currently only consider so-called rank-1 inhomogeneous random graphs, where the probability of an edge is proportional to the product of the weights of the incident vertices. Instead, we could model different types of interaction by considering an arbitrary symmetric function, called a kernel. It would be interesting to see whether our results can be extended to this more general setting. In particular, whether the two-point concentration of the clique number is specific to rank-1 inhomogeneous random graphs, or whether this remains true for a wider class of kernels.

When weights have bounded support, the heuristic explained in Section 2.3.1 can be extended to obtain first order behavior of the clique number for a large class

of kernels. For these kernels, this gives a simpler approach to finding the asymptotic behavior of the clique number than the method described in [66] which provided a general answer. Moreover, based on the results in Section 2.3.2 it might also be possible to extend these results to unbounded kernels.

Planted clique problem. In the planted clique problem one starts by generating a graph as usual. After generating this graph we select a small number of vertices and connect all of them, so that they form a clique. Given such a graph with a planted clique, the problem is to locate this clique with high probability.

The work on this problem has focussed on two cases. In the first case, the underlying graph is an Erdős-Rényi random graph. In principle, this problem can be solved as soon as the planted clique is of size $O(\log(n))$. However, if one is interested in algorithms that can recover the largest clique in polynomial time, then the best-known algorithms require the planted clique to be of size $O(\sqrt{n})$, see [8, 62, 74]. The second case focusses on the very inhomogeneous case, with graphs that have a power-law degree distribution. Here the largest clique can be recovered in polynomial time, see [82, 109].

Alternatively, one could consider the similar hypothesis testing problem. Here we observe a graph where it is unknown whether a clique was planted, and the problem is to decide whether it was planted or not. Instead of a clique, one could plant a denser subgraph and test whether that was planted or not, see [11, 12]. Using the model from Section 2.2 all these problems can be considered in a single framework. It would be particularly interesting to see what the effects of inhomogeneity and sparsity are on the computational complexity in these problems.

2.5 Proofs

This section is devoted to proving the results in Section 2.2. The proofs of Lemmas 2.1, 2.2, and 2.3 are fairly self explanatory. To prove Theorem 2.1 we use the same approach as Matula [125], using the first and second moment method to obtain an upper and lower bound on the clique number separately.

2.5.1 Proof of Lemma 2.1: Existence and uniqueness of the typical clique number

Lemma 2.1 shows that Assumption 2.1 is sufficient to guarantee the existence and uniqueness of the typical clique number ω_n in Definition 2.2. We first show that there must be at least one solution to (2.6) and then show that this solution is unique.

To simplify notation, let $f_n(r)$ be the right-hand side of (2.6), that is

$$f_n(r) = \frac{\log(n) - \log(r) + \log(c_{n,r-1}) + 1}{\log\left(\lambda_n / \mathbb{E}[\widetilde{W}]\right)} + 1. \tag{2.17}$$

To prove the lemma we must show that the solution set of (2.6), given by $\{r \geq 1 : r = f_n(r)\}$, is non-empty and consists of a single point. First note that $c_{n,r-1}$ is a continuous function in r (since these are relative moments of a *truncated* distribution). This in turn implies that $f_n(r)$ is continuous in r. To ensure that the solution set is non-empty, first note that

$$f_n(1) = \frac{\log(n)+1}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} + 1 \geq 1,$$

and

$$f_n(n) = \frac{\log(c_{n,n-1})+1}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} + 1 \leq (n-1)\frac{\log\left(\frac{\lambda_n}{1+\delta}/\mathbb{E}[\widetilde{W}]\right)}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} + 1 \leq n.$$

Hence, there exists at least one value $r \in [1, n]$ satisfying $r = f_n(r)$. To show the uniqueness of this solution we simply show that the slope of $f_n(r)$ is strictly smaller than 1. Note that

$$\begin{aligned}
\frac{\partial}{\partial r} f_n(r) &= \frac{(c'_{n,r-1}/c_{n,r-1}) - (1/r)}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} \\
&= \frac{\left(\mathbb{E}\left[\log\left(\widetilde{W}/\mathbb{E}[\widetilde{W}]\right)\widetilde{W}^{r-1}\right]/\mathbb{E}\left[\widetilde{W}^{r-1}\right]\right) - (1/r)}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} \\
&\leq \frac{\log\left(\frac{\lambda_n}{1+\delta}/\mathbb{E}[\widetilde{W}]\right) - (1/r)}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} < 1, \qquad (2.18)
\end{aligned}$$

where $c'_{n,r-1}$ denotes the partial derivative of $c_{n,r-1}$ with respect to r. Since the partial derivative of $f_n(r)$ is strictly less than 1, there can be at most a single solution of $r = f_n(r)$. Hence, the typical clique number is well defined. □

2.5.2 Proof of Lemma 2.2: Alternative characterization of the typical clique number

Here we derive an alternative representation for the typical clique number ω_n, as formulated in Lemma 2.2. This is sometimes more convenient than the original in Definition 2.2.

By Lemma 2.1 we know that the typical clique number ω_n exists. Therefore, we can solve (2.6) to see that the typical clique number is also the solution of

$$r = \frac{\mathcal{W}_0\left(n\, c_{n,r-1} \mathrm{e}\, b \log(b)\right)}{\log(b)},$$

where $b = \lambda_n/\mathbb{E}[\widetilde{W}]$ and $\mathcal{W}_0(\cdot)$ denotes the principal branch of the Lambert-W func-

tion, see (2.39). Using the approximation $\mathcal{W}_0(x) = \log(x) - \log\log(x) + o(1)$ as $x \to \infty$, as shown in [56], we obtain

$$\begin{aligned} r &= \log_b(n c_{n,r-1} \mathrm{e}\, b \log(b)) - \log_b \log(n c_{n,r-1} \mathrm{e}\, b \log(b)) + o(1) \\ &= \log_b(n c_{n,r-1}) - \log_b \log_b(n c_{n,r-1}) + \log_b(\mathrm{e}) + 1 + o(1). \end{aligned}$$ □

2.5.3 Proof of Lemma 2.3: Bounded typical clique number

Here we show that the scaling λ_n is a positive power of n if, and only if, the typical clique number ω_n converges to a constant. To this end, we first derive a small lemma:

Lemma 2.4. *Let $\alpha \in (0,1)$. If the scaling satisfies $\lambda_n \geq n^{\alpha + o(1)}$ then*

$$\frac{\log\big(c_{n,1/\alpha+o(1)}\big)}{\log\big(\lambda_n/\mathbb{E}[\widetilde{W}]\big)} = o(1).$$

Proof. By Assumption 2.1,

$$\mathbb{P}\left(\max_{i \in V} W_i \leq \lambda_n\right) = (1 - \mathbb{P}\left(W > \lambda_n\right))^n \to 1.$$

Let $\varepsilon > 0$ be arbitrary, then for n large enough and using the above we obtain

$$\mathbb{P}\left(W^{1/\alpha-\varepsilon} > n\right) \leq \mathbb{P}\left(W^{1/(\alpha+o(1))} > n\right) = \mathbb{P}\left(W > \lambda_n\right) = o\left(\frac{1}{n}\right).$$

Therefore, using the tail formula for expectation,

$$\begin{aligned} \mathbb{E}\left[\widetilde{W}^{1/\alpha-\varepsilon}\right] &\leq (1+o(1))\mathbb{E}\left[W^{1/\alpha-\varepsilon}\mathbb{1}_{\{W^{1/\alpha+o(1)} \leq n\}}\right] \\ &\leq (1+o(1))\mathbb{E}\left[W^{1/\alpha-\varepsilon}\mathbb{1}_{\{W^{1/\alpha-\varepsilon} \leq n\}}\right] \\ &= (1+o(1))\int_0^\infty \mathbb{P}\left(W^{1/\alpha-\varepsilon}\mathbb{1}_{\{W^{1/\alpha-\varepsilon} \leq n\}} > x\right) \mathrm{d}x, \end{aligned}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the usual indicator function. Note that $W^{1/\alpha_n}\,\mathbb{1}_{\{W^{1/\alpha_n} \leq n\}} \leq n$, so we can change the upper integration limit. This gives

$$\begin{aligned} \mathbb{E}\left[\widetilde{W}^{1/\alpha-\varepsilon}\right] &\leq O(1) + (1+o(1))\int_1^n \mathbb{P}\left(W^{1/\alpha-\varepsilon} > x\right) \mathrm{d}x \\ &\leq O(1) + (1+o(1))\int_1^n \frac{1}{x}\,\mathrm{d}x = O(1) + (1+o(1))\log(n). \end{aligned}$$

Based on the above we conclude

$$\begin{aligned}\frac{\log\big(c_{n,1/\alpha+o(1)}\big)}{\log\big(\lambda_n/\mathbb{E}[\widetilde{W}]\big)} &= (1+o(1))\frac{\log\big(\mathbb{E}\big[\widetilde{W}^{1/\alpha+o(1)}\big]\big)}{\log\big(\lambda_n\big)}\\ &\le (1+o(1))\frac{\log\big(\mathbb{E}\big[\widetilde{W}^{1/\alpha-\varepsilon}\big]\lambda_n^{2\varepsilon}\big)}{\log\big(\lambda_n\big)}\\ &= (1+o(1))\frac{\log\log\big(n\big)}{\log\big(n^{\alpha+o(1)}\big)} + 2\varepsilon + o(1) = 2\varepsilon + o(1).\end{aligned}$$

Since $\varepsilon > 0$ can be taken arbitrarily small, this completes the proof. □

We are now ready to prove Lemma 2.3. The main idea is that, as $n \to \infty$, most terms of (2.6) become negligible and the remaining terms no longer involve n.

Proof of Lemma 2.3. First suppose that $\lambda_n = n^{\alpha+o(1)}$ with $\alpha \in (0,1)$. By Definition 2.2 the typical clique number ω_n satisfies

$$\omega_n = \frac{\log(n) - \log(\omega_n) + \log(c_{n,\omega_n-1}) + 1}{\log\big(\lambda_n/\mathbb{E}[\widetilde{W}]\big)} + 1. \tag{2.19}$$

Plugging $\omega_n = 1 + 1/\alpha + o(1)$ into (2.19) and using Lemma 2.4,

$$\begin{aligned}\omega_n &= \frac{\log(n) + \log(c_{n,1/\alpha+o(1)})}{\log\big(\lambda_n/\mathbb{E}[\widetilde{W}]\big)} + 1\\ &= \frac{\log(n)}{\log\big(n^{\alpha+o(1)}/\mathbb{E}[\widetilde{W}]\big)} + 1 + o(1)\\ &= \frac{1}{\alpha} + 1 + o(1).\end{aligned}$$

Hence (2.19) is satisfied for $\omega_n = 1 + 1/\alpha + o(1)$ and by Lemma 2.1 this must be the unique solution.

For the other direction, suppose that $\omega_n = 1 + 1/\alpha + o(1)$ with $\alpha \in (0,1)$. Then, by Definition 2.2 and Lemma 2.4,

$$\begin{aligned}\omega_n - 1 &= \frac{\log(n) - \log(\omega_n) + \log(c_{n,\omega_n-1}) + 1}{\log\big(\lambda_n/\mathbb{E}[\widetilde{W}]\big)}\\ &= \frac{\log(n)}{\log\big(\lambda_n/\mathbb{E}[\widetilde{W}]\big)} + o(1).\end{aligned}$$

Solving for λ_n we see that $\lambda_n = n^{\alpha+o(1)}$ since $\mathbb{E}[\widetilde{W}]$ is uniformly bounded. □

2.5.4 Proof of Theorem 2.1: Concentration of the clique number

In this section we prove Theorem 2.1, our main result. First we derive some useful results characterizing the relative moments. Then the proof itself is split into two parts: the high-probability upper bound on the clique number in Sections 2.5.4.2 and 2.5.4.3 using the first moment method, and the high-probability lower bound on the clique number in Sections 2.5.4.4 and 2.5.4.5, using the second moment method. In both parts we separately consider two cases: $\omega_n \to \infty$ and ω_n is bounded. In fact, a third might be possible, namely $\liminf_{n\to\infty} \omega_n < \limsup_{n\to\infty} \omega_n = \infty$. However, in that case we can apply the reasoning below to a maximal subsequence of ω_n converging to infinity, and control the remaining terms by the argument used when ω_n is bounded.

2.5.4.1 Auxiliary results

Binomial coefficients play an important role in counting the number of cliques. Therefore, it is crucial to have tight bounds on the binomial coefficient, provided by the lemma below. This lemma and the corresponding proof can be found in [155]:

Lemma 2.5. *Suppose that $r = o(\sqrt{n})$, then the binomial coefficient can be approximated by*

$$\binom{n}{r} = (1 + o(1)) \frac{1}{\sqrt{2\pi r}} \left(\frac{ne}{r}\right)^r.$$

Another important ingredient for the proof of Theorem 2.1 is to have sharp bounds on the relative moments from Definition 2.1, which are provided by the following lemma. By definition, the typical clique number ω_n is the solution to (2.6). If we consider the right-hand side and left-hand side of (2.6) separately, then we see that these two functions must intersect at ω_n. Moreover, the right-hand side of (2.6) will always grow more slowly than the left-hand side, as shown in the proof of Lemma 2.1. This means that there exists a straight line in between these two functions, intersecting at ω_n as illustrated in Figure 2.1. Using this line we can then find bounds on the right-hand side of (2.6) which in turn lead to bounds on the relative moments.

Lemma 2.6. *Under Assumption 2.1, the relative moments $c_{n,r-1}$ from Definition 2.1 can be bounded by*

$$c_{n,r-1} \geq \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{\beta_n((r-1)-(\omega_n-1))+(\omega_n-1)} \frac{r}{ne}, \qquad \text{for all } 1 \leq r \leq \omega_n,$$

$$c_{n,r-1} \leq \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{\beta_n((r-1)-(\omega_n-1))+(\omega_n-1)} \frac{r}{ne}, \qquad \text{for all } \omega_n \leq r \leq n,$$

with β_n given by

$$\beta_n = \frac{\log\left(\frac{\lambda_n}{1+\delta}/\mathbb{E}[\widetilde{W}]\right)}{\log\left(\lambda_n/\mathbb{E}[\widetilde{W}]\right)} < 1, \tag{2.20}$$

and where $\delta > 0$ arises from Assumption 2.1.

Proof. Let $f_n(r)$ be as defined in (2.17), then the typical clique number ω_n is the solution in r of $r = f_n(r)$. Consider $g_n(r) = \beta_n(r - \omega_n) + \omega_n$, which is the line through ω_n with slope β_n, as shown in Figure 2.1. We will show that, for all n, the line $g_n(r)$ is a lower bound on $f_n(r)$ when $r \in [1, \omega_n]$, and an upper bound on $f_n(r)$ when $r \in [\omega_n, n]$.

The slope of $f_n(r)$ was derived in (2.18) and is bounded by β_n given in (2.20). Hence, we have $g_n(r) \leq f_n(r)$ when $r \in [1, \omega_n]$ and $g_n(r) \geq f_n(r)$ otherwise.

To finish the proof note that, for all $r \in [1, \omega_n]$,

$$\begin{aligned}\beta_n((r-1) - (\omega_n - 1)) + \omega_n &= g_n(r) \\ &\leq f_n(r) = \frac{\log(n) - \log(r) + \log(c_{n,r-1}) + 1}{\log(\lambda_n/\mathbb{E}[\widetilde{W}])} + 1.\end{aligned}$$

Exponentiating both sides yields

$$\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{\beta_n((r-1)-(\omega_n-1))+(\omega_n-1)} \leq \frac{nc_{n,r-1}\mathrm{e}}{r}.$$

Multiplying both sides by $r/(ne)$ gives the first result. The second result follows similarly. □

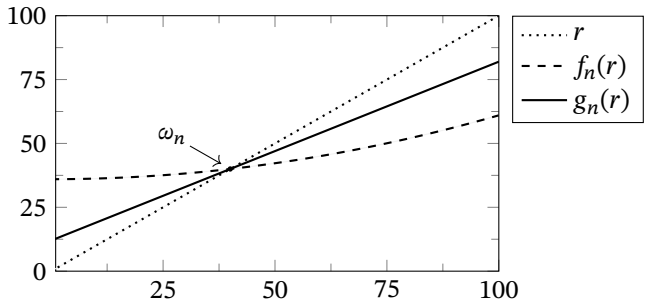


Figure 2.1: Example of the line $g_n(r)$ with slope β_n through ω_n (solid line). Note that this line is a lower bound on $f_n(r)$ for all $r \in [1, \omega_n]$, and an upper bound on $f_n(r)$ for all $r \in [\omega_n, n]$.

2.5.4.2 Upper bound with diverging typical clique number

In this section we prove the upper bound of Theorem 2.1 assuming that $\omega_n \to \infty$. Define the event

$$\mathcal{T}_{n,\delta} = \left\{ \max_{i \in V} W_i \le \frac{\lambda_n}{1+\delta} \right\}. \tag{2.21}$$

Assumption 2.1 enforces that $\mathbb{P}(\mathcal{T}_{n,\delta}) \to 1$ as $n \to \infty$. A trivial, but crucial, observation is that the joint distribution of the weights $(W_1, \dots, W_n)$ conditional on the event $\mathcal{T}_{n,\delta}$ is the same as that of a sequence of i.i.d. truncated weights $(\widetilde{W}_1, \dots, \widetilde{W}_n)$. In other words

$$(W_1, \dots, W_n) \,|\, \mathcal{T}_{n,\delta} \stackrel{d}{=} (\widetilde{W}_1, \dots, \widetilde{W}_n), \tag{2.22}$$

where $\widetilde{W}_i$ are i.i.d. random variables with the same distribution as $\widetilde{W}$. This statement can be checked by an elementary computation.

Let $\omega(G_n)$ be the clique number of the graph G_n and define N_r to be the number of cliques of size r in G_n. Then, by Assumption 2.1 and the first moment method,

$$\begin{aligned} \mathbb{P}(\omega(G_n) \ge r) &= (1+o(1))\mathbb{P}(\omega(G_n) \ge r \,|\, \mathcal{T}_{n,\delta}) \\ &= (1+o(1))\mathbb{P}(N_r \ge 1 \,|\, \mathcal{T}_{n,\delta}) \\ &\le (1+o(1))\mathbb{E}[N_r \,|\, \mathcal{T}_{n,\delta}]. \end{aligned} \tag{2.23}$$

Then by linearity of expectation,

$$\begin{aligned} \mathbb{E}[N_r \,|\, \mathcal{T}_{n,\delta}] &= \sum_{C \subseteq V, |C|=r} \mathbb{P}(C \text{ is a clique in } G_n \,|\, \mathcal{T}_{n,\delta}) \\ &= \sum_{C \subseteq V, |C|=r} \mathbb{E}\left[\prod_{i<j \in C} \frac{W_i}{\lambda_n} \cdot \frac{W_j}{\lambda_n} \wedge 1 \,\middle|\, \mathcal{T}_{n,\delta} \right] \\ &= \sum_{C \subseteq V, |C|=r} \mathbb{E}\left[\prod_{i<j \in C} \frac{\widetilde{W}_i}{\lambda_n} \cdot \frac{\widetilde{W}_j}{\lambda_n} \right] \\ &= \binom{n}{r} \mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n} \right)^{r-1} \right]^r \\ &= \binom{n}{r} \left(c_{n,r-1} \left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n} \right)^{r-1} \right)^r. \end{aligned} \tag{2.24}$$

To prove the upper bound of Theorem 2.1, we need show that $\mathbb{E}[N_r \,|\, \mathcal{T}_{n,\delta}] \to 0$, as $n \to \infty$, when $r > \lfloor \omega_n + \varepsilon \rfloor$, and since r is integer this implies $r \ge \omega_n + \varepsilon$. Using Lemmas 2.5 and 2.6 we can further bound the above expression as

$$\mathbb{E}[N_r \,|\, \mathcal{T}_{n,\delta}] \le (1+o(1)) \frac{1}{\sqrt{2\pi r}} \left(\frac{ne}{r} \right)^r \left(c_{n,r-1} \left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n} \right)^{r-1} \right)^r$$

$$\leq (1+o(1))\frac{1}{\sqrt{2\pi r}}\left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n}\right)^{r(r-1)-r(\beta_n((r-1)-(\omega_n-1))+(\omega_n-1))}$$
$$= (1+o(1))\frac{1}{\sqrt{2\pi r}}\left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n}\right)^{-(1-\beta_n)\cdot r((\omega_n-1)-(r-1))}, \tag{2.25}$$

where β_n comes from Lemma 2.6. Using the definition of β_n we have

$$\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-(1-\beta_n)} = \left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n}\right)\left(\frac{\lambda_n/(1+\delta)}{\mathbb{E}[\widetilde{W}]}\right) = \frac{1}{1+\delta}. \tag{2.26}$$

Combining (2.25) and (2.26) and because $r-\omega_n \geq \varepsilon$ we obtain

$$\mathbb{E}[N_r \mid \mathcal{T}_{n,\delta}] = (1+o(1))\frac{1}{\sqrt{2\pi r}}\left(\frac{1}{1+\delta}\right)^{-r(\omega_n-r)}$$
$$\leq (1+o(1))\frac{1}{\sqrt{2\pi r}}\left(\frac{1}{1+\delta}\right)^{r\varepsilon}. \tag{2.27}$$

Since $\omega_n \to \infty$ it is easily seen from (2.27) that $\mathbb{E}[N_r \mid \mathcal{T}_{n,\delta}] \to 0$ when $r > \lfloor \omega_n + \varepsilon \rfloor$. Hence it follows from (2.23) that $\mathbb{P}(\omega(G_n) > \lfloor \omega_n + \varepsilon \rfloor) \to 0$. □

2.5.4.3 Upper bound with bounded typical clique number

Here we prove the upper bound of Theorem 2.1 assuming that ω_n is bounded. First we consider the case where ω_n converges, in this case there exists an $\alpha > 0$ such that $\omega_n = 1/\alpha + 1 + o(1)$. We want to apply all the steps in Section 2.5.4.2, but instead of conditioning on the event in (2.21) we will condition on the event

$$\mathcal{T}_{n,\eta} = \left\{\max_{i\in V} W_i \leq \frac{\lambda_n}{1+\eta}\right\}, \tag{2.28}$$

where $\eta > 0$ comes from Assumption 2.2.

Since $\omega_n = 1/\alpha + 1 + o(1)$ it follows from Lemma 2.3 that $\lambda_n = n^{\alpha+o(1)}$, and by Assumption 2.2 we have $\mathbb{P}(\mathcal{T}_{n,\eta}) \to 1$. Moreover, by repeating the steps in Lemmas 2.3 and 2.4 it can easily be checked that replacing δ by η in Definitions 2.1 and 2.2 the typical clique number remains equal to $\omega_n = 1/\alpha + 1 + o(1)$. Therefore, we can follow all steps in Section 2.5.4.2 but conditioning on $\mathcal{T}_{n,\eta}$ instead of $\mathcal{T}_{n,\delta}$. This gives

$$\mathbb{P}(\omega(G_n) \geq r) = (1+o(1))\mathbb{P}(\omega(G_n) \geq r \mid \mathcal{T}_{n,\eta})$$
$$\leq (1+o(1))\mathbb{E}[N_r \mid \mathcal{T}_{n,\eta}]$$
$$\leq (1+o(1))\frac{1}{\sqrt{2\pi r}}\left(\frac{1}{1+\eta}\right)^{r\varepsilon}. \tag{2.29}$$

Since $\eta > 0$ is arbitrary and $r = \omega_n + \varepsilon$ is bounded it follows from (2.29) that we can make $\mathbb{P}(\omega(G_n) \geq r)$ arbitrarily small, hence $\mathbb{P}(\omega(G_n) \geq r) \to 0$.

To complete the proof we consider the case when ω_n does not converge. In this case, we know that every subsequence $(n_i)_{i\in\mathbb{N}}$ contains a further subsequence $(n_{i_j})_{j\in\mathbb{N}}$ along which $\omega_{n_{i_j}}$ converges. Applying the arguments above shows that every subsequence $(n_i)_{i\in\mathbb{N}}$ has a further subsequence $(n_{i_j})_{j\in\mathbb{N}}$ along which $\mathbb{P}(\omega(G_{n_{i_j}}) \geq r) \to 0$, and it follows that $\mathbb{P}(\omega(G_n) \geq r) \to 0$. □

2.5.4.4 Lower bound with diverging typical clique number

In this section we prove the lower bound of Theorem 2.1 assuming that $\omega_n \to \infty$. Recall that $\omega(G_n)$ denotes the clique number of the graph G_n and N_r is the number of cliques of size r in G_n. Then, by the second moment method, and using the truncation event $\mathcal{T}_{n,\delta}$ given by (2.21) together with Assumption 2.1,

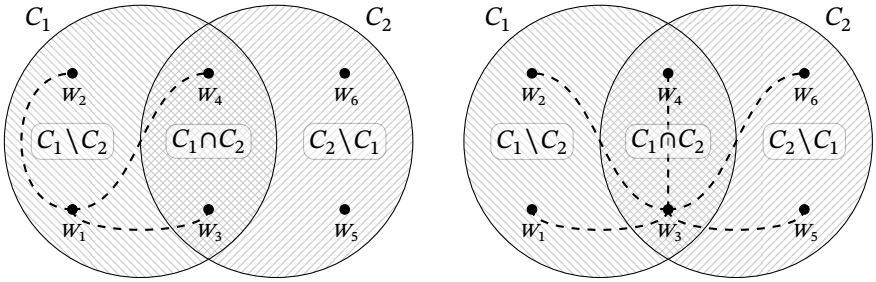
$$\begin{aligned}\mathbb{P}(\omega(G_n) < r) &= (1+o(1))\mathbb{P}(\omega(G_n) < r \mid \mathcal{T}_{n,\delta}) \\ &= (1+o(1))\mathbb{P}(N_r = 0 \mid \mathcal{T}_{n,\delta}) \\ &\leq (1+o(1))\frac{\mathrm{Var}(N_r \mid \mathcal{T}_{n,\delta})}{\mathbb{E}[N_r \mid \mathcal{T}_{n,\delta}]^2} = (1+o(1))\left(\frac{\mathbb{E}[N_r^2 \mid \mathcal{T}_{n,\delta}]}{\mathbb{E}[N_r \mid \mathcal{T}_{n,\delta}]^2} - 1\right). \quad (2.30)\end{aligned}$$

Hence we need to show that $\mathbb{E}[N_r^2 \mid \mathcal{T}_{n,\delta}]/\mathbb{E}[N_r \mid \mathcal{T}_{n,\delta}]^2 \to 1$ as $n \to \infty$, with $r = \lfloor \omega_n - \varepsilon \rfloor$. The first moment of the number of cliques N_r was computed in (2.24), and is given by

$$\mathbb{E}[N_r \mid \mathcal{T}_{n,\delta}] = \binom{n}{r}\mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{r-1}\right]^r. \quad (2.31)$$

Similarly, the second moment of the number of cliques N_r is also found using (2.22) and linearity of expectation as

$$\begin{aligned}\mathbb{E}[N_r^2 \mid \mathcal{T}_{n,\delta}] &= \sum_{|C_1|=r,\,|C_2|=r} \mathbb{P}(C_1 \text{ and } C_2 \text{ are both cliques in } G_n \mid \mathcal{T}_{n,\delta}) \\ &= \sum_{|C_1|=r,\,|C_2|=r} \mathbb{E}\left[\frac{\prod_{i<j\in C_1}\frac{\widetilde{W}_i}{\lambda_n}\cdot\frac{\widetilde{W}_j}{\lambda_n}\prod_{i<j\in C_2}\frac{\widetilde{W}_i}{\lambda_n}\cdot\frac{\widetilde{W}_j}{\lambda_n}}{\prod_{i<j\in C_1\cap C_2}\frac{\widetilde{W}_i}{\lambda_n}\cdot\frac{\widetilde{W}_j}{\lambda_n}}\right] \\ &= \sum_{k=0}^{r}\sum_{\substack{|C_1|=r,\,|C_2|=r,\\ |C_1\cap C_2|=k}} \mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{r-1}\right]^{2(r-k)}\mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{2(r-1)-(k-1)}\right]^k \quad (2.32) \\ &= \sum_{k=0}^{r}\binom{n}{r}\binom{r}{k}\binom{n-r}{r-k}\mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{r-1}\right]^{2(r-k)}\mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{2(r-1)-(k-1)}\right]^k. \quad (2.33)\end{aligned}$$



(a) Example of all the edges connecting to a vertex in $C_1 \setminus C_2$.

(b) Example of all the edges connecting to a vertex in $C_1 \cap C_2$.

Figure 2.2: Example of edges connecting to vertices in different parts of $C_1 \cup C_2$.

The equality in (2.33) comes from counting how many times each vertex is an endpoint of an edge, and thus how many times each weight is present in the product. We count two cases separately:

- Vertices in $C_1 \setminus C_2$ will need edges to each other vertex in C_1. So, each vertex in $C_1 \setminus C_2$ will be $r-1$ times in the product of (2.32) and similarly for vertices in $C_2 \setminus C_1$. Since we have $2(r-k)$ vertices in $C_1 \setminus C_2$ and $C_2 \setminus C_1$ we get the $\mathbb{E}[W^{r-1}]^{2(r-k)}$ term. See Figure 2.2(a).
- Vertices in $C_1 \cap C_2$ will need edges to each vertex in $C_1 \cup C_2$. So, each vertex in $C_1 \cap C_2$ will be $2(r-1)-(k-1)$ times in the product of (2.32) and we have k vertices in $C_1 \cap C_2$. So we get the $\mathbb{E}[W^{2(r-1)-(k-1)}]^k$ term. See Figure 2.2(b).

Combining (2.31) and (2.33) we obtain

$$\begin{aligned}
\frac{\mathbb{E}\left[N_r^2 \mid \mathcal{T}_{n,\delta}\right]}{\mathbb{E}[N_r \mid \mathcal{T}_{n,\delta}]^2} &= \sum_{k=0}^{r} \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \cdot \frac{\mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{r-1}\right]^{2(r-k)} \mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{2(r-1)-(k-1)}\right]^{k}}{\mathbb{E}\left[\left(\frac{\widetilde{W}}{\lambda_n}\right)^{r-1}\right]^{2r}} \\
&= \sum_{k=0}^{r} \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \left(\frac{c_{n,2(r-1)-(k-1)}}{c_{n,r-1}^2}\right)^{k} \left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n}\right)^{-k(k-1)} \\
&\leq 1 + \sum_{k=1}^{r} \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \left(\frac{c_{n,2(r-1)-(k-1)}}{c_{n,r-1}^2}\right)^{k} \left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n}\right)^{-k(k-1)} \\
&\leq 1 + \max_{1\leq k\leq r} \underbrace{r\frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \left(\frac{c_{n,2(r-1)-(k-1)}}{c_{n,r-1}^2}\right)^{k} \left(\frac{\mathbb{E}[\widetilde{W}]}{\lambda_n}\right)^{-k(k-1)}}_{\coloneqq b_k^{n,r}}.
\end{aligned} \tag{2.34}$$

We will show that $\max_{k\in[r]} b_k^{n,r} \to 0$ as $n \to \infty$. To continue we consider two cases: *(i)* $k = r$; and *(ii)* $1 \leq k \leq r-1$.

Case (i): Here $k = r$, so we want to show that $b_r^{n,r} \to 0$ as $n \to \infty$. By definition of $b_r^{n,r}$ and by Lemmas 2.5 and 2.6,

$$\begin{aligned} b_r^{n,r} &= r\frac{1}{\binom{n}{r}}\left(\frac{1}{c_{n,r-1}}\right)^r\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{r(r-1)} \\ &\le (1+o(1))r\sqrt{2\pi r}\left(\frac{r}{n\mathrm{e}}\right)^r\left(\frac{1}{c_{n,r-1}}\right)^r\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{r(r-1)} \\ &\le (1+o(1))r\sqrt{2\pi r}\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{r((r-1)-\beta_n((r-1)-(\omega_n-1))-(\omega_n-1))} \\ &= (1+o(1))r\sqrt{2\pi r}\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-(1-\beta_n)r((\omega_n-1)-(r-1))}. \end{aligned}$$

where β_n is given in Lemma 2.6.

Using (2.26) together with the fact that $r = \lfloor \omega_n - \varepsilon \rfloor \le \omega_n - \varepsilon$ yields the bound

$$\begin{aligned} b_r^{n,r} &\le (1+o(1))r\sqrt{2\pi r}\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-(1-\beta)r((\omega_n-1)-(r-1))} \\ &\le (1+o(1))r\sqrt{2\pi r}\left(\frac{1}{1+\delta}\right)^{\varepsilon r}. \end{aligned} \tag{2.35}$$

Since $\omega_n \to \infty$ it is easily seen from (2.35) that $b_r^{n,r} \to 0$.

Case (ii): Here we must show that $\max_{k\in[r-1]} b_k^{n,r} \to 0$ as $n \to \infty$. First we apply Lemma 2.5 on the binomial coefficients, which gives

$$\begin{aligned} \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} &= (1+o(1))\sqrt{\frac{r}{2\pi k(r-k)}}\left(\frac{r\mathrm{e}}{k}\right)^k\left(\frac{(n-r)\mathrm{e}}{r-k}\right)^{r-k}\left(\frac{n\mathrm{e}}{r}\right)^{-r} \\ &= (1+o(1))\sqrt{\frac{r}{2\pi k(r-k)}}\left(\frac{r\mathrm{e}}{k}\right)^k\left(\frac{n-r}{n}\frac{r}{r-k}\right)^r\left(\frac{r-k}{(n-r)\mathrm{e}}\right)^k \\ &= (1+o(1))\sqrt{\frac{r}{2\pi k(r-k)}}\left(\frac{r\mathrm{e}}{k}\right)^k\left(\frac{r-k}{n-r}\right)^k. \end{aligned} \tag{2.36}$$

Now, for all $1 \le k \le r-1$ we have that $k \le \omega_n - \varepsilon - 1 \le \omega_n - 2\varepsilon$ and therefore $2(r-1)-(k-1) \ge \omega_n - 1$. So, we can apply Lemma 2.6 on both $c_{n,2(r-1)-(k-1)}$ and on $c_{n,r-1}$, yielding

$$\begin{aligned} &\left(\frac{c_{n,2(r-1)-(k-1)}}{c_{n,r-1}^2}\right)^k\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{k(k-1)} \\ &\qquad\le \left(\frac{2(r-1)-(k-1)+1}{n\mathrm{e}}\right)^k\left(\frac{n\mathrm{e}}{r}\right)^{2k}\left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{k(k-1)} \end{aligned}$$

$$\times \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{k(\beta_n(2(r-1)-(k-1)-(\omega_n-1))+(\omega_n-1))}$$
$$\times \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-2k(\beta_n((r-1)-(\omega_n-1))+(\omega_n-1))}$$
$$= \left(\frac{2(r-1)-(k-1)+1}{r^2} n\mathrm{e}\right)^k \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-(1-\beta_n)\cdot k((\omega_n-1)-(k-1))} \tag{2.37}$$

Combining (2.36) and (2.37) we obtain

$$\begin{aligned} b_k^{n,r} &\le (1+o(1))r\sqrt{\frac{r}{2\pi k(r-k)}}\left(\frac{r-k}{k}\frac{2(r-1)-(k-1)+1}{r}\mathrm{e}^2\right)^k \\ &\quad \times \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-(1-\beta_n)\cdot k((\omega_n-1)-(k-1))} \\ &\le (1+o(1))r\sqrt{\frac{r}{2\pi k(r-k)}}\left(\frac{r-k}{k}2\mathrm{e}^2\right)^k \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-(1-\beta_n)\cdot k((\omega_n-1)-(k-1))}. \end{aligned}$$

Here we can use (2.26) again. This gives

$$\begin{aligned} b_k^{n,r} &\le (1+o(1))r\sqrt{\frac{r}{2\pi k(r-k)}}\left(\frac{r-k}{k}2\mathrm{e}^2\right)^k \left(\frac{\lambda_n}{\mathbb{E}[\widetilde{W}]}\right)^{-(1-\beta_n)\cdot k((\omega_n-1)-(k-1))} \\ &\le (1+o(1))r\sqrt{\frac{r}{2\pi k(r-k)}}\left(\frac{r-k}{k}2\mathrm{e}^2\right)^k \left(\frac{1}{1+\delta}\right)^{k((\omega_n-1)-(k-1))} \\ &= (1+o(1))r\sqrt{\frac{r}{2\pi k(r-k)}}\left(2\mathrm{e}^2\left(\frac{r}{k}-1\right)\left(\frac{1}{1+\delta}\right)^{\omega_n-k}\right)^k. \end{aligned} \tag{2.38}$$

Fix $\zeta \in \left(0,(1+2\mathrm{e}^2)^{-1}\right)$ and recall that $\omega_n \to \infty$. Then it can easily be seen from (2.38) that $\max_{1\le k\le(1-\zeta)r} b_k^{n,r} \to 0$ since $(1+\delta)^{-(\omega_n-k)} \to 0$ exponentially, eventually dominating the other terms. Finally, to show that $\max_{(1-\zeta)r\le k\le r} b_k^{n,r} \to 0$, note that $2\mathrm{e}^2(r/k-1) < 1$ and therefore $\left(2\mathrm{e}^2(r/k-1)(1+\delta)^{-(\omega_n-k)}\right)^k \to 0$ exponentially, again dominating the remaining terms.

Hence $\max_{k\in[r]} b_k^{n,r} \to 0$ as $n\to\infty$ and $r = \lfloor \omega_n - \varepsilon \rfloor$. Using (2.34) and (2.30) it follows that $\mathbb{P}(\omega(G_n) < \lfloor \omega_n - \varepsilon \rfloor) \to 0$ as $n\to\infty$. □

2.5.4.5 Lower bound with bounded typical clique number

Here we prove the lower bound of Theorem 2.1 assuming that ω_n is bounded. First we consider the case where ω_n converges, in this case there exists an $\alpha > 0$ such that $\omega_n = 1/\alpha + 1 + o(1)$. We want to apply all the steps in Section 2.5.4.4, but instead of conditioning on the event $\mathcal{T}_{n,\delta}$ given in (2.21) we will condition on the event $\mathcal{T}_{n,\eta}$ given in (2.28). As shown in Section 2.5.4.3, the typical clique number ω_n is unaffected by

this change, and by Assumption 2.2 we also have $\mathbb{P}(\mathcal{T}_{n,\eta}) \to 1$.

Now, following all steps in Section 2.5.4.2 but conditioning on $\mathcal{T}_{n,\eta}$ instead of $\mathcal{T}_{n,\delta}$ we obtain

$$\mathbb{P}(\omega(G_n) < r) = (1 + o(1))\mathbb{P}(\omega(G_n) < r \mid \mathcal{T}_{n,\eta}) \leq (1 + o(1))\frac{\mathbb{E}[N_r^2 \mid \mathcal{T}_{n,\eta}]}{\mathbb{E}[N_r \mid \mathcal{T}_{n,\eta}]^2} - 1.$$

By combining (2.35) and (2.38), and using that $r = \omega_n - \varepsilon$ is bounded we get

$$\mathbb{P}(\omega(G_n) < r) \leq (1 + o(1))\frac{\mathbb{E}[N_r^2 \mid \mathcal{T}_{n,\eta}]}{\mathbb{E}[N_r \mid \mathcal{T}_{n,\eta}]^2} - 1 \leq O(1)\left(\frac{1}{1+\eta}\right)^{\varepsilon}.$$

Since we can make $\eta > 0$ arbitrarily large it follows that $\mathbb{P}(\omega(G_n) < r) \to 0$.

To complete the proof we consider the case when ω_n does not converge. In this case, we know that every subsequence $(n_i)_{i\in\mathbb{N}}$ contains a further subsequence $(n_{i_j})_{j\in\mathbb{N}}$ along which $\omega_{n_{i_j}}$ converges. Applying the arguments above shows that every subsequence $(n_i)_{i\in\mathbb{N}}$ has a further subsequence $(n_{i_j})_{j\in\mathbb{N}}$ along which $\mathbb{P}(\omega(G_{n_{i_j}}) < r) \to 0$, and it follows that $\mathbb{P}(\omega(G_n) < r) \to 0$. □

2.6 Derivation of examples

In this section we derive the asymptotic behavior of the typical clique number ω_n for some given weight distributions W and scalings λ_n. This can be very difficult in general, but for several choices of weights good asymptotic characterizations can be given. An overview of these results can be found in Tables 2.1, 2.2, and 2.3.

Throughout the derivation of the examples below we make use of the Lambert-W functions, which are obtained from the solutions in $y \in \mathbb{R}$ of the equation

$$x = y\mathrm{e}^y, \tag{2.39}$$

When $x \geq 0$ this has a unique real solution, while for $x \in (-1/\mathrm{e}, 0)$ there are two real solutions. This gives rise to two branches: the *principal branch*, denoted by $\mathcal{W}_0 : [-1/\mathrm{e}, \infty) \mapsto [-1, \infty)$ and the *lower branch*, denoted by $\mathcal{W}_{-1} : [-1/\mathrm{e}, 0) \mapsto (-\infty, -1]$. For an overview of this function and its properties see [56].

2.6.1 Bernoulli weights

Let W have a Bernoulli distribution with parameter p, that is $W \sim \mathrm{Ber}(p)$, and take any scaling $\lambda_n \geq c > 1$. In this case, we have an Erdős–Rényi random graph with connection probability λ_n^{-2} on approximately np vertices, with all remaining vertices being isolated. Therefore, by (2.10), we expect the typical clique number ω_n to be

$$\omega_n = \log_{\lambda_n}(np) - \log_{\lambda_n}\log_{\lambda_n}(np) + \log_{\lambda_n}(\mathrm{e}) + 1 + o(1).$$

In this section, we show that the same result is obtained by solving (2.6) from Definition 2.2.

The relative moments from Definition 2.1 are given by

$$c_{n,r-1} = \frac{\mathbb{E}\big[\widetilde{W}^{r-1}\big]}{\mathbb{E}[\widetilde{W}]^{r-1}} = \frac{\mathbb{E}\big[W^{r-1}\big]}{\mathbb{E}[W]^{r-1}} = p^{2-r}.$$

The typical clique number ω_n from Definition 2.2 is given by the solution in r of

$$r = \frac{\log(n) - \log(r) + (2-r)\log(p) + 1}{\log(\lambda_n/p)} + 1.$$

Solving this we obtain

$$\omega_n = \frac{\mathcal{W}_0(np\mathrm{e}\lambda_n \log(\lambda_n))}{\log(\lambda_n)}.$$

where $\mathcal{W}_0$ denotes the principal branch of the Lambert-W function, see (2.39). We can simplify the solution above using the approximation $\mathcal{W}_0(x) = \log(x) - \log\log(x) + o(1)$ as $x \to \infty$ from [56]. This gives

$$\begin{aligned}
\omega_n &= \frac{\log(np\mathrm{e}\lambda_n \log(\lambda_n)) - \log\log(np\mathrm{e}\lambda_n \log(\lambda_n))}{\log(\lambda_n)} + o(1) \\
&= \frac{\log(np\mathrm{e}\lambda_n \log(\lambda_n)) - \log\log(np)}{\log(\lambda_n)} + o(1) \\
&= \log_{\lambda_n}(np\log(\lambda_n)) - \log_{\lambda_n}\log(np) + \log_{\lambda_n}(\mathrm{e}) + 1 + o(1) \\
&= \log_{\lambda_n}(np) - \log_{\lambda_n}\log_{\lambda_n}(np) + \log_{\lambda_n}(\mathrm{e}) + 1 + o(1),
\end{aligned}$$

which is exactly the expected solution.

2.6.2 Beta weights

Let W have a beta distribution with parameters $\alpha > 0$ and $\beta > 0$, that is $W \sim \mathrm{Beta}(\alpha, \beta)$, and take any scaling $\lambda_n \geq c > 1$. Then the relative moments from Definition 2.1 are given by

$$\begin{aligned}
c_{n,r-1} &= \frac{\mathbb{E}\big[\widetilde{W}^{r-1}\big]}{\mathbb{E}[\widetilde{W}]^{r-1}} = \frac{\mathbb{E}\big[W^{r-1}\big]}{\mathbb{E}[W]^{r-1}} = \prod_{r=0}^{r-2} \frac{\alpha+r}{\alpha+\beta+r} \Big/ \left(\frac{\alpha}{\alpha+\beta}\right)^{r-1} \\
&= \frac{\Gamma(\alpha+r-1)}{\Gamma(\alpha+\beta+r-1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \cdot \left(\frac{\alpha}{\alpha+\beta}\right)^{r-1}.
\end{aligned}$$

Using Stirling's approximation, the above can be simplified for large r. This gives

$$\log(c_{n,r-1}) = -\beta\log(r) + \log\big(\Gamma(\alpha+\beta)/\Gamma(\alpha)\big) + (r-1)\log\left(\frac{\alpha}{\alpha+\beta}\right) + o(1).$$

Therefore, the typical clique number ω_n from Definition 2.2 is given by the solution in r of

$$r = \frac{\log(n) - (\beta+1)\log(r) + (r-1)\log\big(\frac{\alpha}{\alpha+\beta}\big) + \log\big(\Gamma(\alpha+\beta)/\Gamma(\alpha)\big) + 1}{\log\Big(\lambda_n/\big(\frac{\alpha}{\alpha+\beta}\big)\Big)} + 1 + o(1).$$

Solving this we obtain

$$\omega_n = \frac{(1+\beta)\mathcal{W}_0\Big(\frac{\left(n\mathrm{e}\lambda_n\Gamma(\alpha+\beta)/\Gamma(\alpha)\right)^{\frac{1}{1+\beta}}\log(\lambda_n)}{1+\beta}\Big)}{\log(\lambda_n)} + o(1),$$

As in the previous example, using the approximation $\mathcal{W}_0(x) = \log(x) - \log\log(x) + o(1)$, we obtain

$$\begin{aligned}
\omega_n &= \frac{(1+\beta)\log\Big(\frac{\left(n\mathrm{e}\lambda_n\Gamma(\alpha+\beta)/\Gamma(\alpha)\right)^{\frac{1}{1+\beta}}\log(\lambda_n)}{1+\beta}\Big)}{\log(\lambda_n)} \\
&\qquad - \frac{(1+\beta)\log\log\Big(\frac{\left(n\mathrm{e}\lambda_n\Gamma(\alpha+\beta)/\Gamma(\alpha)\right)^{\frac{1}{1+\beta}}\log(\lambda_n)}{1+\beta}\Big)}{\log(\lambda_n)} + o(1) \\
&= \frac{\log\big(n\mathrm{e}\lambda_n\Gamma(\alpha+\beta)/\Gamma(\alpha)\big) + (1+\beta)\log\Big(\frac{\log(\lambda_n)}{1+\beta}\Big) - (1+\beta)\log\log(n)}{\log(\lambda_n)} + o(1) \\
&= \log_{\lambda_n}\big(n\mathrm{e}\Gamma(\alpha+\beta)/\Gamma(\alpha)\big) - (1+\beta)\log_{\lambda_n}\big((1+\beta)\log_{\lambda_n}(n)\big) + 1 + o(1) \\
&= \log_{\lambda_n}(n) - (1+\beta)\log_{\lambda_n}\big((1+\beta)\log_{\lambda_n}(n)\big) \\
&\qquad + \log_{\lambda_n}(\mathrm{e}) + \log_{\lambda_n}\big(\Gamma(\alpha+\beta)/\Gamma(\alpha)\big) + 1 + o(1).
\end{aligned}$$

2.6.3 Gamma weights

Let W have a Gamma distribution with shape α and rate β, that is $W \sim \mathrm{Gamma}(\alpha,\beta)$. First we assume that truncating the weight distribution has asymptotically almost no effect on the relative moments from Definition 2.1. We begin by assuming that

$$c_{n,r-1} = \frac{\mathbb{E}\big[\widetilde{W}^{r-1}\big]}{\mathbb{E}[\widetilde{W}]^{r-1}} = (1+o(1))\frac{\mathbb{E}\big[W^{r-1}\big]}{\mathbb{E}[W]^{r-1}}, \tag{2.40}$$

for all $r \le \omega_n$, and use this to find the typical clique number ω_n. After that, we will show that the assumption in (2.40) is valid.

By the assumption in (2.40)

$$c_{n,r-1} = (1+o(1))\frac{\mathbb{E}\big[W^{r-1}\big]}{\mathbb{E}[W]^{r-1}} = (1+o(1))\frac{\Gamma(\alpha+r-1)}{\Gamma(\alpha)\alpha^{r-1}}.$$

To satisfy Assumption 2.1 we must have $\lambda_n \to \infty$, and therefore

$$\frac{\log(c_{n,r-1})}{\log(\lambda_n/\mathbb{E}[\widetilde{W}])} = \frac{\log(\Gamma(\alpha+r-1)) - \log(\Gamma(\alpha)) - (r-1)\log(\alpha)}{\log(\lambda_n/\mathbb{E}[\widetilde{W}])} + o(1).$$

Using Stirling's approximation again relying on the fact that r is large, the typical clique number ω_n is given by the solution in r of

$$\begin{aligned}
r &= \frac{\log(n) - \log(r) + \log(\Gamma(\alpha+r-1)) - \log(\Gamma(\alpha)) - (r-1)\log(\alpha) + 1}{\log\big(\lambda_n\beta/\alpha\big)} + 1 + o(1) \\
&= \frac{\log(n) - \log(r) + \big(\alpha+r-\frac{3}{2}\big)\log(\alpha+r-2) + 2}{\log\big(\lambda_n\beta/\alpha\big)} \\
&\qquad - \frac{(\alpha+r-2) + \log(\Gamma(\alpha)) + (r-1)\log(\alpha) + 2}{\log\big(\lambda_n\beta/\alpha\big)} + 1 + o(1) \\
&= \frac{\log(n) + \big(\alpha+r-\frac{5}{2}\big)\log\big(\alpha+r-\frac{5}{2}\big)}{\log\big(\lambda_n\beta/\alpha\big)} \\
&\qquad - \frac{\big(\alpha+r-\frac{5}{2}\big) + \log(\Gamma(\alpha)) + (r-1)\log(\alpha)}{\log\big(\lambda_n\beta/\alpha\big)} + 1 + o(1).
\end{aligned}$$

Substituting $x = r + \alpha - 5/2$, we get

$$x = \frac{\log(n) + \big(\alpha-x-\frac{3}{2}\big)\log(\alpha) + x\log(x) - x - \log(\Gamma(\alpha)) + 2}{\log\big(\lambda_n\beta/\alpha\big)} - \frac{3}{2} + \alpha + o(1).$$

Solving for x we find $\omega_n + \alpha - 5/2$, and therefore the typical clique number ω_n is given by

$$\omega_n = \frac{2\log(n) - (3-2\alpha)\log(\beta\lambda_n) + 4 - 2\log(\Gamma(\alpha))}{-\mathcal{W}_{-1}\Big(-\frac{2\log(n)-(3-2\alpha)\log(\beta\lambda_n)+4-2\log(\Gamma(\alpha))}{2\mathrm{e}\beta\lambda_n}\Big)} + \frac{5}{2} - \alpha + o(1), \tag{2.41}$$

where $\mathcal{W}_{-1}$ denotes the lower branch of the Lambert-W function, see (2.39).

2.6.3.1 First scaling: $\lambda_n = (1+\varphi)\log(n)/\beta$

Let $\lambda_n = (1+\varphi)\log(n)/\beta$, with $\varphi > 0$. We will show that in this case (2.41) simplifies to the result in Table 2.2. This gives

$$\begin{aligned}\omega_n &= \frac{\log(n) - (3/2-\alpha)\log((1+\varphi)\log(n)) + 2 - \log(\Gamma(\alpha))}{-\mathcal{W}_{-1}\Big(-\frac{\log(n)-(3/2-\alpha)\log((1+\varphi)\log(n))+2-\log(\Gamma(\alpha))}{e(1+\varphi)\log(n)}\Big)} + \frac{5}{2} - \alpha + o(1)\\ &= \frac{\log(n) - (3/2-\alpha)\log((1+\varphi)\log(n)) + 2 - \log(\Gamma(\alpha))}{-\mathcal{W}_{-1}\Big(-\frac{1}{\mathrm{e}(1+\varphi)}\Big) + o(1)} + \frac{5}{2} - \alpha + o(1)\\ &= (1+o(1))\frac{\log(n)}{-\mathcal{W}_{-1}\Big(-\frac{1}{\mathrm{e}(1+\varphi)}\Big)}.\end{aligned}$$

It remains to show that our assumption from (2.40) holds. We will do this in two parts: *(i)* where we show $\mathbb{E}\big[\widetilde{W}^{r-1}\big]/\mathbb{E}\big[W^{r-1}\big] \to 1$ for any $r \le \omega_n$; and *(ii)* where we show $(\mathbb{E}[\widetilde{W}]/\mathbb{E}[W])^{r-1} \to 1$ for any $r \le \omega_n$. First, observe that for any $k \ge 1$ we have

$$\begin{aligned}\frac{\mathbb{E}\big[\widetilde{W}^k\big]}{\mathbb{E}\big[W^k\big]} &= \frac{\mathbb{E}\big[W^k \,\big|\, W \le \frac{\lambda_n}{1+\delta}\big]}{\mathbb{E}\big[W^k\big]}\\ &= \frac{1}{\mathbb{P}\big(W \le \frac{\lambda_n}{1+\delta}\big)}\frac{\int_0^{\frac{\lambda_n}{1+\delta}} x^k f_W(x)\,\mathrm{d}x}{\int_0^{\infty} x^k f_W(x)\,\mathrm{d}x} = \frac{\mathbb{P}\big(Z_k \le \frac{\lambda_n}{1+\delta}\big)}{\mathbb{P}\big(W \le \frac{\lambda_n}{1+\delta}\big)},\end{aligned} \tag{2.42}$$

where $Z_k \sim \mathrm{Gamma}(\alpha+k,\beta)$.

Part (i): To simplify notation, let $a := (1+\varphi)/(1+\delta)$, $b := -1/\mathcal{W}_{-1}(-1/(\mathrm{e}(1+\varphi)))$, and $z_n := \omega_n + \alpha - 1 = (1+o(1))\,b\log(n) + \alpha - 1$. Note that, because $\varphi > \delta > 0$ we have

$$a = \frac{1+\varphi}{1+\delta} > 1 > \left(-\mathcal{W}_{-1}\left(-\frac{1}{\mathrm{e}(1+\varphi)}\right)\right)^{-1} = b.$$

Finally, let $X_i \sim \mathrm{Exp}(\beta)$. Then using (2.42) and Assumption 2.1 we have

$$\begin{aligned}\frac{\mathbb{E}\big[\widetilde{W}^{r-1}\big]}{\mathbb{E}\big[W^{r-1}\big]} &= (1+o(1))\mathbb{P}\Big(Z_{r-1} \le \frac{\lambda_n}{1+\delta}\Big)\\ &\ge (1+o(1))\mathbb{P}\Big(\sum_{i=1}^{\lceil z_n\rceil} X_i \le \frac{a}{\beta}\log(n)\Big)\\ &= (1+o(1))\Big(1 - \mathbb{P}\Big(\frac{1}{\lceil z_n\rceil}\sum_{i=1}^{\lceil z_n\rceil} X_i > (1+o(1))\frac{1}{\beta}\frac{a}{b}\Big)\Big)\\ &\ge (1+o(1))\Big(1 - \exp\Big(-\lceil z_n\rceil I\Big((1+o(1))\frac{1}{\beta}\frac{a}{b}\Big)\Big)\Big),\end{aligned} \tag{2.43}$$

where $I(x) := x\beta - 1 - \log(x\beta)$ is the rate function of an exponential distribution with rate β. Hence, for n large enough and because $a/b > 1$ we have

$$I\Big((1+o(1))\frac{1}{\beta}\frac{a}{b}\Big) = (1+o(1))I\Big(\frac{1}{\beta}\frac{a}{b}\Big) = (1+o(1))\Big((a/b) - 1 - \log(a/b)\Big) > 0. \quad (2.44)$$

Combining (2.43) and (2.44) we see that $\mathbb{E}\big[\widetilde{W}^{r-1}\big]/\mathbb{E}\big[W^{r-1}\big] \to 1$.

Part (ii): From (2.42) and integration by parts we obtain

$$\begin{aligned}\frac{\mathbb{E}\big[\widetilde{W}\big]}{\mathbb{E}\big[W\big]} &= \frac{\mathbb{P}\big(Z_1 \le \frac{\lambda_n}{1+\delta}\big)}{\mathbb{P}\big(W \le \frac{\lambda_n}{1+\delta}\big)} = \frac{\gamma(1+\alpha, \beta\lambda_n/(1+\delta))}{\alpha\gamma(\alpha, \beta\lambda_n/(1+\delta))} \\ &= 1 - \frac{(\beta\lambda_n/(1+\delta))^\alpha \exp(-\beta\lambda_n/(1+\delta))}{\alpha\gamma(\alpha, \beta\lambda_n/(1+\delta))} \\ &= 1 - O(1)\frac{\log(n)^\alpha}{n^a},\end{aligned}$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function and we recall that $a = (1+\varphi)/(1+\delta) > 1$. Hence, we have $(\mathbb{E}[\widetilde{W}]/\mathbb{E}[W])^{r-1} \to 1$ for any $r \le n$.

From parts *(i)* and *(ii)* we see that our assumption in (2.40) indeed holds.

2.6.3.2 Second scaling: $\lambda_n = \log(n)^{1+\varphi}/\beta$

Let $\lambda_n = \log(n)^{1+\varphi}/\beta$, with $\varphi > 0$. We will show that in this case (2.41) simplifies to the result in Table 2.3. This gives

$$\begin{aligned}\omega_n &= \frac{\log(n) - (3/2-\alpha)(1+\varphi)\log\log(n) + 2 - \log(\Gamma(\alpha))}{-\mathcal{W}_{-1}\Big(-\frac{\log(n)-(3/2-\alpha)(1+\varphi)\log\log(n)+2-\log(\Gamma(\alpha))}{\mathrm{e}\log(n)^{1+\varphi}}\Big)} + \frac{5}{2} - \alpha + o(1) \\ &= \frac{\log(n) - (3/2-\alpha)(1+\varphi)\log\log(n) + 2 - \log(\Gamma(\alpha))}{-\mathcal{W}_{-1}\Big(-\frac{1}{\mathrm{e}\log(n)^{\varphi}}\Big) + o(1)} + \frac{5}{2} - \alpha + o(1) \\ &= (1+o(1))\frac{\log(n)}{\log\left(\mathrm{e}\log(n)^{\varphi}\right)} = (1+o(1))\frac{1}{\varphi}\frac{\log(n)}{\log\log(n)}.\end{aligned}$$

Compared to Section 2.6.3.1 the scaling λ_n is larger and the typical clique number ω_n is smaller. Therefore, it is evident that our assumption from (2.40) is also valid in this case.

2.6.4 Half-normal weights

Let W have a half-normal distribution with parameters $\mu = 0$ and $\sigma > 0$, that is $W \sim |X|$, where $X \sim \mathrm{N}(0,\sigma)$. We proceed in the exact same way as for the Gamma

distribution, and assume first that

$$c_{n,r-1} = \frac{\mathbb{E}\big[\widetilde{W}^{r-1}\big]}{\mathbb{E}[\widetilde{W}]^{r-1}} = (1+o(1))\frac{\mathbb{E}\big[W^{r-1}\big]}{\mathbb{E}[W]^{r-1}} = (1+o(1))\pi^{\frac{r}{2}-1}\Gamma(r/2), \tag{2.45}$$

for all $r \le \omega_n$. To satisfy Assumption 2.1 we must have $\lambda_n \to \infty$, and therefore

$$\frac{\log(c_{n,r-1})}{\log(\lambda_n/\mathbb{E}[\widetilde{W}])} = \frac{\left(\frac{r}{2}-1\right)\log(\pi) + \log(\Gamma(r/2))}{\log(\lambda_n/\mathbb{E}[\widetilde{W}])} + o(1).$$

Using Stirling's approximation and the fact that the typical clique number ω_n grows with n, the typical clique number ω_n is given by the solution in r of

$$\begin{aligned} r &= \frac{\log(n) - \log(r) + \left(\frac{r}{2}-1\right)\log(\pi) + \log\left(\Gamma\left(\frac{r}{2}\right)\right) + 1}{\log\big(\lambda_n/\sqrt{2\sigma^2/\pi}\big)} + 1 + o(1) \\ &= \frac{\log(n) - \log(r) + \left(\frac{r}{2}-1\right)\log(\pi) + \left(\frac{r}{2}-\frac{1}{2}\right)\log\left(\frac{r}{2}-1\right) - \frac{r}{2} + 3}{\log\big(\lambda_n/\sqrt{2\sigma^2/\pi}\big)} + 1 + o(1) \\ &= \frac{\log(n) + \left(\frac{r-3}{2}\right)\log\left(\frac{r-3}{2}\right) + \left(\frac{r-3}{2}\right)\big(\log(\pi)-1\big) + \log\left(\frac{\mathrm{e}^2\sqrt{\pi}}{2}\right)}{\log\big(\lambda_n/\sqrt{2\sigma^2/\pi}\big)} + 1 + o(1). \end{aligned}$$

Substituting $x = (r-3)/2$, we get

$$x = \frac{1}{2}\frac{\log(n) + x\log(x) + x(\log(\pi)-1) + \log\left(\frac{\mathrm{e}^2\sqrt{\pi}}{2}\right)}{\log\big(\lambda_n/\sqrt{2\sigma^2/\pi}\big)} - 1 + o(1).$$

Solving for x we find $(\omega_n - 3)/2$, and therefore the typical clique number ω_n is given by

$$\omega_n = \frac{2\log(n) - 4\log(\lambda_n) + 4 - \log(\pi)}{-\mathcal{W}_{-1}\left(-\frac{2\log(n)-4\log(\lambda_n)+4-\log(\pi)}{\mathrm{e}\lambda_n^2}\right)} + 3 + o(1), \tag{2.46}$$

where $\mathcal{W}_{-1}$ denotes the lower branch of the Lambert-W function, see (2.39).

2.6.4.1 First scaling: $\lambda_n = (1+\varphi)\sigma\sqrt{2\log(n)}$

Let $\lambda_n = (1+\varphi)\sigma\sqrt{2\log(n)}$, with $\varphi > 0$. We will show that in this case (2.46) simplifies to the result in Table 2.2. This gives

$$\begin{aligned}
\omega_n &= \frac{2\log(n) - 4\log((1+\varphi)\sqrt{2\log(n)}) + 4 - \log(\pi)}{-\mathcal{W}_{-1}\left(-\frac{2\log(n)-4\log((1+\varphi)\sqrt{2\log(n)})+4-\log(\pi)}{2\mathrm{e}(1+\varphi)^2\log(n)}\right)} + 3 + o(1) \\
&= \frac{2\log(n) - 2\log\log(n) - 2\log(\sqrt{4\pi}(1+\varphi)^2/\mathrm{e}^2)}{-\mathcal{W}_{-1}\left(-\frac{1}{\mathrm{e}(1+\varphi)^2}\right)} + 3 + o(1) \\
&= (1+o(1))\frac{2\log(n)}{-\mathcal{W}_{-1}\left(-\frac{1}{\mathrm{e}(1+\varphi)^2}\right)}.
\end{aligned}$$

It remains to show that our assumption from (2.45) holds. We will do this in two parts: *(i)* where we show $\mathbb{E}\big[\widetilde{W}^{r-1}\big]/\mathbb{E}\big[W^{r-1}\big] \to 1$ for any $r \le \omega_n$; and *(ii)* where we show $(\mathbb{E}[\widetilde{W}]/\mathbb{E}[W])^{r-1} \to 1$ for any $r \le \omega_n$. First, observe that for any $k \ge 1$ we have

$$\frac{\mathbb{E}\big[\widetilde{W}^k\big]}{\mathbb{E}\big[W^k\big]} = \frac{\mathbb{E}\big[W^k \,\big|\, W \le \frac{\lambda_n}{1+\delta}\big]}{\mathbb{E}\big[W^k\big]} = \frac{1}{\mathbb{P}\big(W \le \frac{\lambda_n}{1+\delta}\big)} \frac{\int_0^{\frac{\lambda_n}{1+\delta}} x^k f_W(x)\,\mathrm{d}x}{\int_0^{\infty} x^k f_W(x)\,\mathrm{d}x} = \frac{\mathbb{P}\big(Z_k \le \frac{\lambda_n}{1+\delta}\big)}{\mathbb{P}\big(W \le \frac{\lambda_n}{1+\delta}\big)}, \tag{2.47}$$

where it can be recognized that $Z_k \sim \mathrm{Gamma}\big(\frac{k+1}{2}, \frac{\lambda_n}{1+\delta}\frac{1}{2\sigma^2}\big)$.

Part (i): First let $a := (1+\varphi)/(1+\delta)$ and $b := -2/\mathcal{W}_{-1}\big(-1/(\mathrm{e}(1+\varphi)^2)\big)$ to simplify notation. Note that, because $\varphi > \delta > 0$ we have

$$2a^2 = 2\left(\frac{1+\varphi}{1+\delta}\right)^2 > 2 > 2\left(-\mathcal{W}_{-1}\left(-\frac{1}{\mathrm{e}(1+\varphi)^2}\right)\right)^{-1} = b.$$

Finally, let $X_i \sim \mathrm{Exp}\big(\frac{\lambda_n}{1+\delta}\frac{1}{2\sigma^2}\big)$. Then using (2.47) and Assumption 2.1 we have

$$\begin{aligned}
\frac{\mathbb{E}\big[\widetilde{W}^{r-1}\big]}{\mathbb{E}\big[W^{r-1}\big]} &\ge (1+o(1))\mathbb{P}\Big(Z_{\lceil\omega_n/2\rceil} \le \frac{\lambda_n}{1+\delta}\Big) \\
&= (1+o(1))\mathbb{P}\Big(\sum_{i=1}^{\lceil\omega_n/2\rceil} X_i \le a\sqrt{2\sigma^2\log(n)}\Big) \\
&= (1+o(1))\left(1 - \mathbb{P}\left(\frac{1}{\lceil\omega_n/2\rceil}\sum_{i=1}^{\lceil\omega_n/2\rceil} X_i > (1+o(1))2\frac{a}{b}\sqrt{\frac{2\sigma^2}{\log(n)}}\right)\right) \\
&\ge (1+o(1))\left(1 - \exp\left(-\lceil\omega_n/2\rceil I\left((1+o(1))2\frac{a}{b}\sqrt{\frac{2\sigma^2}{\log(n)}}\right)\right)\right),
\end{aligned} \tag{2.48}$$

where $I(\cdot)$ denotes the rate function of an exponential distribution with rate $\frac{\lambda_n}{1+\delta}\frac{1}{2\sigma^2} = a\sqrt{\log(n)/(2\sigma^2)}$. Hence, for n large enough and because $2a^2/b > 1$ we have

$$I\Big((1+o(1))2\frac{a}{b}\sqrt{\frac{2\sigma^2}{\log(n)}}\Big) = (1+o(1))\Big(2\frac{a^2}{b} - 1 - \log\Big(2\frac{a^2}{b}\Big)\Big) > 0. \tag{2.49}$$

Combining (2.48) and (2.49) we see that $\mathbb{E}\big[\widetilde{W}^{r-1}\big]/\mathbb{E}\big[W^{r-1}\big] \to 1$.

Part (ii): From (2.47) we obtain

$$\frac{\mathbb{E}[\widetilde{W}]}{\mathbb{E}[W]} = \frac{\mathbb{P}\big(Z_1 \le \frac{\lambda_n}{1+\delta}\big)}{\mathbb{P}\big(W \le \frac{\lambda_n}{1+\delta}\big)} = \frac{1-\exp\Big(-\Big(\frac{\lambda_n}{1+\delta}\Big)^2\frac{1}{2\sigma^2}\Big)}{\operatorname{erf}\Big(\frac{\lambda_n}{1+\delta}\frac{1}{2\sigma^2}\Big)} = \frac{1-\exp\big(-a^2\log(n)\big)}{\operatorname{erf}\big(a\sqrt{\log(n)/(2\sigma^2)}\big)},$$

where $\operatorname{erf}(\cdot)$ is the error function and recall $a = (1+\varphi)/(1+\delta) > 1$. Hence, we have $(\mathbb{E}[\widetilde{W}]/\mathbb{E}[W])^{r-1} \to 1$ for any $r \le n$.

From parts *(i)* and *(ii)* we see that our assumption from (2.45) was indeed valid.

2.6.4.2 Second scaling: $\lambda_n = \sigma\sqrt{2\log(n)}^{1+\varphi}$

Let $\lambda_n = \sigma\sqrt{2\log(n)}^{1+\varphi}$, with $\varphi > 0$. We will show that in this case (2.46) simplifies to the result in Table 2.3. This gives

$$\begin{aligned}\omega_n &= \frac{2\log(n) - 2(1+\varphi)\log(2\log(n)) + 4 - \log(\pi)}{-\mathcal{W}_{-1}\Big(-\frac{2\log(n)-2(1+\varphi)\log(2\log(n))+4-\log(\pi)}{\mathrm{e}(2\log(n))^{1+\varphi}}\Big)} + 3 + o(1)\\ &= \frac{2\log(n) - 2(1+\varphi)\log(2\log(n)) + 4 - \log(\pi)}{-\mathcal{W}_{-1}\Big(-\frac{1}{\mathrm{e}(2\log(n))^{\varphi}}\Big) + o(1)} + 3 + o(1)\\ &= (1+o(1))\frac{2\log(n)}{\log\big(\mathrm{e}(2\log(n))^{\varphi}\big)} = (1+o(1))\frac{2}{\varphi}\frac{\log(n)}{\log\log(n)}.\end{aligned}$$

Compared to Section 2.6.4.1 the scaling λ_n is larger and the typical clique number ω_n is smaller. Therefore, it is evident that our assumption from (2.45) is also valid in this case.

2.6.5 Log-normal weights

Let W have a log-normal distribution with parameters $\mu = 0$ and $\sigma = 1$, that is $W \sim \exp(X)$, where X is standard normal. Then one can show that the relative moments

from Definition 2.1 are given by

$$\frac{\log(c_{n,r-1})}{\log(\lambda_n/\mathbb{E}[\widetilde{W}])} = \frac{\frac{1}{2}(r-1)(r-2)}{\log(\lambda_n/\mathbb{E}[\widetilde{W}])} + o(1). \tag{2.50}$$

provided that $r \leq \omega_n$. For brevity of presentation, we omit the details of this derivation.

Using this in Definition 2.2, the typical clique number ω_n is the solution in r of

$$r = \frac{\log(n) - \log(r) + \frac{1}{2}(r-1)(r-2) + 1}{\log\bigl(\lambda_n/\sqrt{\mathrm{e}}\bigr)} + 1 + o(1).$$

To solve this, we bound the solution with the following two bounds

$$r \leq \frac{\log(n) + \frac{1}{2}(r-1)(r-2) + 1}{\log\bigl(\lambda_n/\sqrt{\mathrm{e}}\bigr)} + 1 + o(1),$$
$$r \geq \frac{\log(n) - r + \frac{1}{2}(r-1)(r-2) + 1}{\log\bigl(\lambda_n/\sqrt{\mathrm{e}}\bigr)} + 1 + o(1).$$

Solving the above gives

$$\omega_n \leq \log(\lambda_n) - \sqrt{\log(\lambda_n)^2 - 2(\log(n)+1)} + 1 + o(1), \tag{2.51}$$
$$\omega_n \geq \log(\lambda_n) - \sqrt{(1+\log(\lambda_n))^2 - 2\log(n)} + 2 + o(1). \tag{2.52}$$

Combining (2.51) and (2.52), and plugging in $\lambda_n = (1+\varphi)\exp\bigl(\sqrt{2\log(n)}\bigr)$, we obtain the result in Table 2.2. Similarly, the result in Table 2.3 is obtained by plugging in $\lambda_n = \exp\bigl(\sqrt{2\log(n)}\bigr)^{1+\varphi}$.

It remains to show that our assumption from (2.50) holds. We will do this in two parts: *(i)* where we show $\mathbb{E}\bigl[\widetilde{W}^{r-1}\bigr]/\mathbb{E}\bigl[W^{r-1}\bigr] \to 1$ for any $r \leq \omega_n$; and *(ii)* where we show $(\mathbb{E}[\widetilde{W}]/\mathbb{E}[W])^{r-1} \to 1$ for any $r \leq \omega_n$. First, observe that for any $k \geq 1$ we have

$$\frac{\mathbb{E}\bigl[\widetilde{W}^k\bigr]}{\mathbb{E}\bigl[W^k\bigr]} = \frac{\mathbb{E}\bigl[W^k \mid W \leq \frac{\lambda_n}{1+\delta}\bigr]}{\mathbb{E}\bigl[W^k\bigr]} = \frac{1}{\mathbb{P}\bigl(W \leq \frac{\lambda_n}{1+\delta}\bigr)} \frac{\int_0^{\frac{\lambda_n}{1+\delta}} x^k f_W(x)\,\mathrm{d}x}{\int_0^{\infty} x^k f_W(x)\,\mathrm{d}x}$$
$$= \frac{1 - \operatorname{erf}\Bigl(\frac{k-\log(\lambda_n)}{\sqrt{2}}\Bigr)}{1 - \operatorname{erf}\Bigl(\frac{-\log(\lambda_n)}{\sqrt{2}}\Bigr)}, \tag{2.53}$$

where $\operatorname{erf}(\cdot)$ denotes the error function.

Part (i): It follows from (2.51) that $\omega_n \le \log(\lambda_n) + 1 + o(1)$, and therefore by (2.53) we have for all $r - 1 \le \omega_n - 1 \le \log(\lambda_n) + o(1)$,

$$\frac{\mathbb{E}\big[\widetilde{W}^{r-1}\big]}{\mathbb{E}\big[W^{r-1}\big]} \ge \frac{1 - \operatorname{erf}\Big(\frac{\omega_n - 1 - \log(\lambda_n)}{\sqrt{2}}\Big)}{1 - \operatorname{erf}\Big(\frac{-\log(\lambda_n)}{\sqrt{2}}\Big)} \ge \frac{1 - \operatorname{erf}\Big(\frac{o(1)}{\sqrt{2}}\Big)}{1 - \operatorname{erf}\Big(\frac{-\log(\lambda_n)}{\sqrt{2}}\Big)} \to 1.$$

Part (ii): From (2.53) we obtain

$$\left(\frac{\mathbb{E}\big[\widetilde{W}\big]}{\mathbb{E}\big[W\big]}\right)^{r-1} = \left(\frac{1 - \operatorname{erf}\Big(\frac{1 - \log(\lambda_n)}{\sqrt{2}}\Big)}{1 - \operatorname{erf}\Big(\frac{-\log(\lambda_n)}{\sqrt{2}}\Big)}\right)^{r-1} = \left(\frac{1 - \operatorname{erf}\Big(O(1) - \sqrt{\log(n)}\Big)}{1 - \operatorname{erf}\Big(O(1) - \sqrt{\log(n)}\Big)}\right)^{r-1}$$

Hence, we have $(\mathbb{E}[\widetilde{W}]/\mathbb{E}[W])^{r-1} \to 1$ for any $r \le n^{1-\varepsilon}$, with $\varepsilon > 0$.

From parts *(i)* and *(ii)* we see that our assumption from (2.50) was indeed valid.

Chapter 3

Quasi-cliques in inhomogeneous random graphs

Based on:
Quasi-cliques in inhomogeneous random graphs,
K. Bogerd,
Submitted.

Given a graph G and a constant $\gamma \in [0,1]$, let $\omega^{(\gamma)}(G)$ be the largest integer r such that there exists an r-vertex subgraph of G containing at least $\gamma\binom{r}{2}$ edges. It was recently shown by Balister, Bollobás, Sahasrabudhe and Veremyev [13] that $\omega^{(\gamma)}(G)$ is highly concentrated when G is an Erdős-Rényi random graph. This chapter provides a simple method to extend that result to a setting of inhomogeneous random graphs, showing that $\omega^{(\gamma)}(G)$ remains concentrated on a small range of values even if G is an inhomogeneous random graph. Furthermore, we give an explicit expression for $\omega^{(\gamma)}(G)$ and show that it depends primarily on the largest edge probability of the graph G.

3.1 Introduction

Let $G = (V, E)$ be a simple graph, with vertex set V and edge set E. Given a subset of vertices $S \subseteq V$, let $G[S]$ denote the *subgraph* of G induced by S. That is, $G[S]$ is a graph with vertex set S and edge set $\{(i,j) : i,j \in S\} \cap E$. A *clique* is a subset of vertices $C \subseteq V$ such that $G[C]$ is a complete graph, meaning that all vertices in $G[C]$ are connected by an edge. Cliques are an important concept in graph theory, and are often used as a model for community structure [6, 121, 133]. In particular, the problem of finding the largest clique or largest community in a given graph has received much interest [62, 63].

However, for many practical applications the definition of a clique can be too restrictive. Often a few missing edges within a community are fine, as long as the community remains sufficiently well connected. To this end, several relaxations have been proposed for the definition of a clique [150]. One of the most successful of these is known as the *γ-quasi-clique*, where γ is a parameter [2]. For $\gamma \in [0,1]$, a γ-quasi-clique is a subset of vertices $S \subseteq V$ such that $G[S]$ contains at least $\gamma\binom{|S|}{2}$ edges. That is, a γ-quasi-clique is a subset of vertices such that a fraction γ of all possible edges between them is present.

Just as for cliques, one would like to know the size of the largest quasi-clique in a given graph [3, 45, 158]. However, it comes as no surprise that finding the largest quasi-clique is a computationally hard problem [37, 148, 149], similar to the problem of finding the largest clique [73, 101, 115]. To circumvent this difficulty, a common approach has been to study the related problem of determining the size of the largest clique or quasi-clique in random graphs. For cliques this approach has been very fruitful, and it turns out that the size of the largest clique is highly concentrated in a variety of random graph models. The first results of this type were obtained for Erdős-Rényi random graphs [30, 125, 126, 127], and later similar results were obtained for random geometric graphs [137], and inhomogeneous random graphs [27, 66].

Recently, the size of the largest quasi-clique was also studied in an Erdős-Rényi random graph, where it was shown that the largest quasi-clique is again highly concentrated [13], see also [78, 79, 114]. The aim of this chapter is to extend that result to the setting of inhomogeneous random graphs. In particular, we formalize a heuristic presented in Section 2.3.1 from the previous chapter, and show how this (together with the result from [13]) can be applied to show that the largest quasi-clique remains concentrated on a narrow range of values even in an inhomogeneous random graph.

3.2 Model and results

We are interested in understanding the behavior of the largest quasi-clique in an inhomogeneous random graph. To this end, define the γ-quasi-clique number $\omega^{(\gamma)}(G)$ of a graph G as the size of the largest subset of vertices $S \subseteq V$ such that the induced subgraph $G[S]$ contains at least $\gamma\binom{|S|}{2}$ edges, where $\gamma \in [0,1]$ is a parameter. Note that $\omega^{(1)}(G)$ is the familiar clique number of G, usually denoted simply by $\omega(G)$.

In this chapter, we study the behavior of $\omega^{(\gamma)}(G)$ when G is distributed according to the random graph model $\mathbb{G}(n,\kappa)$. This model has two parameters: the number of vertices n, and a symmetric measurable function called a *kernel* $\kappa : [0,1]^2 \to (0,1)$. Below we introduce the key concepts of this model, for a more detailed overview we refer the reader to Lovász's book [119]. An element of $\mathbb{G}(n,\kappa)$ is a simple graph $G = (V,E)$ that has $n \in \mathbb{N}$ vertices with vertex set $V = [n] \coloneqq \{1,\dots,n\}$, and a random edge set E. Each vertex $i \in V$ is assigned a *weight* W_i, which is simply a

uniform variable on $[0,1]$, that is $W_i \sim \mathrm{Unif}(0,1)$. Conditionally on these weights, the presence of an edge between two vertices $i, j \in V$, with $i \neq j$, is modeled by independent Bernoulli random variables with success probability

$$p_{ij} := \mathbb{P}\big((i,j) \in E \,|\, (W_k)_{k\in V}\big) = \kappa(W_i, W_j). \tag{3.1}$$

The kernel $\kappa(\cdot,\cdot)$ and the vertex weights W_i are both not allowed to depend on the graph size n, and therefore the edge probabilities p_{ij} are independent of n. This means that the graphs we consider are necessarily dense and have a number of edges that is quadratic in the graph size.

This brings us to the main result of this chapter, which is to show that the γ-quasi-clique number $\omega^{(\gamma)}(G)$ of a graph $G \sim \mathbb{G}(n,\kappa)$ is concentrated on a small range of values. Furthermore, this result shows that the size of the largest quasi-clique depends primarily on the densest part of the graph, where the edge probabilities are close to their maximum value. This is made precise by the following result.

Theorem 3.1. *Let $\kappa(\cdot,\cdot)$ be a kernel that is continuous and attains it maximum value at the point (c,c) for some $c \in [0,1]$, and let $p_{\max} := \kappa(c,c)$. Given $p_{\max} < \gamma \leq 1$, define*

$$\omega_n^{(\gamma)} := \frac{2\log(n)}{D(\gamma, p_{\max})}, \tag{3.2}$$

where $D(\gamma,p)$ is the Kullback-Leibler divergence between the Bernoulli distributions $\mathrm{Bern}(\gamma)$ and $\mathrm{Bern}(p)$, given by

$$D(\gamma,p) := \begin{cases} \gamma\log\left(\frac{\gamma}{p}\right) + (1-\gamma)\log\left(\frac{1-\gamma}{1-p}\right) & \text{if } \gamma < 1, \\ \log\left(\frac{1}{p}\right) & \text{if } \gamma = 1. \end{cases} \tag{3.3}$$

Then, for every $\varepsilon > 0$,

$$\mathbb{P}\Big(\omega^{(\gamma)}(G) \in \big[(1-\varepsilon)\omega_n^{(\gamma)}, (1+\varepsilon)\omega_n^{(\gamma)}\big]\Big) \to 1, \qquad \text{as } n \to \infty. \tag{3.4}$$

To display the relevance of the above result, we show that it can be applied to many well-known random graph models. The simplest example is probably the Erdős-Rényi random graph, which is obtained by setting the kernel $\kappa(x,y)$ to a constant independent of x and y. Another commonly used example are the so-called rank-1 random graphs, where $\kappa(x,y) = \varphi(x)\varphi(y)$ for some function φ. Often the function $\varphi(\cdot) = F_X^{-1}(\cdot)$ is the inverse cumulative distribution function of some distribution X, so that $\varphi(W_i)$ can be interpreted as a sample from that distribution. This results in a model similar to that considered in the previous chapter. The final model that satisfies the conditions in Theorem 3.1 is the stochastic block model [106], also called the planted partition model in computer science. This model is obtained when the kernel $\kappa(\cdot,\cdot)$ is only allowed to take on finitely many different values.

Note that Theorem 3.1 gives the first-order behavior of $\omega_n^{(\gamma)}$ from (3.2). More precise results are known for the clique and quasi-clique number in an Erdős-Rényi random graph [13, 126], or for the clique number in rank-1 random graphs [27]. Specifically, in those cases the quasi-clique number and clique number are concentrated on two consecutive integers. Therefore, it might be reasonable to expect that it is likewise possible to show such a two-point concentration result in the more general model we consider in this chapter. However, this would require a significantly more detailed analysis. The main difficulty here is that the higher order terms of $\omega_n^{(\gamma)}$ will likely depend in a complex way on the whole kernel $\kappa(\cdot,\cdot)$ and not just on the maximum value $\kappa(c,c)$. This was also observed for rank-1 random graphs in Section 2.3.1 from the previous chapter, where several examples are explicitly computed. Thus, the method we use in the proof of Theorem 3.1 will likely not be precise enough to characterize the higher order terms of $\omega_n^{(\gamma)}$ and a different approach would be needed for this.

3.3 Proof

We end this chapter with the proof of Theorem 3.1. This proof is based on the ideas presented in Section 2.3.1 from the previous chapter combined with the results in [13] and [126]. Below we consider the upper and lower bound of (3.4) separately. Furthermore, we use standard asymptotic notation as explained in Section 1.5.

Upper bound: We first define a coupling between the random graph $\mathbb{G}(n;\kappa)$ and the Erdős-Rényi random graph $\mathbb{G}(n;p_{\max})$, where we recall that $p_{\max} = \kappa(c,c)$ is the maximum edge probability. For $i \neq j \in [n]$, let $U_{ij} \sim \mathrm{Unif}(0,1)$ be independent uniform random variables on $[0,1]$. Conditionally on these uniform random variables and the weights W_i, with $i \in [n]$, define

$$\begin{aligned} G = (V,E), &\quad \text{with} \quad V = [n], \quad \text{and} \quad E = \{(i,j) : U_{ij} \leq \kappa(W_i,W_j)\}, \\ G' = (V',E'), &\quad \text{with} \quad V' = [n], \quad \text{and} \quad E' = \{(i,j) : U_{ij} \leq \kappa(c,c)\}. \end{aligned} \tag{3.5}$$

It can easily be seen that G is an inhomogeneous random graph, that is $G \sim \mathbb{G}(n,\kappa)$. Similarly, $G' \sim \mathbb{G}(n,p_{\max})$ is distributed as an Erdős-Rényi random graph with edge probability $p_{\max} = \kappa(c,c)$.

Because the edge probabilities satisfy $p_{ij} = \kappa(W_i,W_j) \leq p_{\max}$ almost surely, for all $i \neq j \in [n]$, the coupling in (3.5) shows that $\omega^{(\gamma)}(G) \leq \omega^{(\gamma)}(G')$ almost surely. Furthermore, by [13, Theorem 1] if $\gamma < 1$ or [126, Theorem 6] if $\gamma = 1$, it follows that

$$\omega^{(\gamma)}(G') \leq \frac{2}{D(\gamma,p_{\max})}\Big(\log(n) - \log\log(n) + \log(\mathrm{e}D(\gamma,p_{\max})/2)\Big) + 1 + \varepsilon,$$

with high probability.

Combining the above, we obtain

$$\begin{aligned}\omega^{(\gamma)}(G) &\le \omega^{(\gamma)}(G') \\ &\le \frac{2}{D(\gamma, p_{\max})}\Big(\log(n) - \log\log(n) + \log(\mathrm{e}D(\gamma, p_{\max})/2)\Big) + 1 + \varepsilon \\ &\le (1+\varepsilon)\frac{2\log(n)}{D(\gamma, p_{\max})} = (1+\varepsilon)\omega_n^{(\gamma)},\end{aligned}$$

with high probability. This shows that $\mathbb{P}\left(\omega^{(\gamma)}(G) \le (1+\varepsilon)\omega_n^{(\gamma)}\right) \to 1$, completing the proof for the upper bound of (3.4).

Lower bound: Let $\delta_n = 1/\log(n)$ and define $S_n := \{i \in V : W_i \in [c-\delta_n, c+\delta_n]\}$ to be the subset of vertices that have vertex weight W_i close to c, where we recall that c is such that the kernel $\kappa(\cdot,\cdot)$ attains it maximal value at the point (c,c). Note that the set S_n is random and by Hoeffding's inequality (see [36, Theorem 2.8]), for any fixed $t > 0$, we have

$$\mathbb{P}\left(|S_n| \le \mathbb{E}[|S_n|] - t\right) \le \exp\left(-2t^2/n\right) \to 0, \tag{3.6}$$

where $\mathbb{E}[|S_n|] = n\mathbb{P}(W \in [c-\delta_n, c+\delta_n]) = n^{1-o(1)}$ by definition of δ_n. Furthermore, define $p_n := \inf_{(x,y)\in[c-\delta_n,c+\delta_n]^2\cap[0,1]^2} \kappa(x,y)$ and observe that $p_n \to p_{\max}$ by continuity of the kernel. Therefore we conclude that $D(\gamma, p_n) \to D(\gamma, p_{\max})$. Using this, together with (3.6) and t fixed, we obtain

$$\begin{aligned}(1-\varepsilon)\frac{2\log(n)}{D(\gamma, p_{\max})} &\le (1-\varepsilon/2)\frac{2\log(\mathbb{E}[|S_n|]-t)}{D(\gamma, p_{\max})} \\ &\le (1-\varepsilon/3)\frac{2\log(|S_n|)}{D(\gamma, p_{\max})} \\ &\le (1-\varepsilon/4)\frac{2\log(|S_n|)}{D(\gamma, p_n)},\end{aligned} \tag{3.7}$$

with high probability.

Similarly to the coupling in (3.5), conditionally on the uniform random variables U_{ij}, for $i \ne j \in [n]$, and the vertex weights W_i, for $i \in [n]$, define

$$G'' = (V'', E''), \quad \text{with} \quad V'' = [n], \quad \text{and} \quad E'' = \{(i,j) : U_{ij} \le p_n\}. \tag{3.8}$$

Note that the graph G'' is distributed as the Erdős-Rényi random graph $\mathbb{G}(n, p_n)$ with edge probability p_n.

Given a graph G, recall that $G[S_n]$ denotes the subgraph induced by the vertices in S_n. Because the kernel is continuous around the point (c,c), there exists an n large enough such that δ_n is small enough to ensure that the edge probabilities satisfy $p_{ij} \ge p_n$ almost surely, for all $i \ne j \in S_n$ (note that, if the kernel is continuous everywhere then this holds for every n). Hence, the coupling in (3.8) shows that

$\omega^{(\gamma)}(G) \geq \omega^{(\gamma)}(G[S_n]) \geq \omega^{(\gamma)}(G''[S_n])$ almost surely, provided n is large enough. Combining this with (3.7) and [13, Theorem 1] if $\gamma < 1$ or [126, Theorem 6] if $\gamma = 1$, we obtain

$$\begin{aligned}\omega^{(\gamma)}(G) &\geq \omega^{(\gamma)}(G[S_n]) \geq \omega^{(\gamma)}(G''[S_n]) \\ &\geq \frac{2}{D(\gamma, p_n)}\Big(\log(|S_n|) - \log\log(|S_n|) + \log(\mathrm{e}D(\gamma, p_n)/2)\Big) - \varepsilon \\ &\geq (1-\varepsilon/4)\frac{2\log(|S_n|)}{D(\gamma, p_n)} \geq (1-\varepsilon)\frac{2\log(n)}{D(\gamma, p_{\max})} = (1-\varepsilon)\omega_n^{(\gamma)},\end{aligned}$$

with high probability. This shows that $\mathbb{P}\left(\omega^{(\gamma)}(G) \geq (1-\varepsilon)\omega_n^{(\gamma)}\right) \to 1$, completing the proof for the lower bound of (3.4). □

Chapter 4

Detecting planted communities in inhomogeneous random graphs

Based on:
Detecting a planted community in an inhomogeneous random graph,
K. Bogerd, R. M. Castro, R. van der Hofstad, and N. Verzelen,
Bernoulli (accepted).

We study the problem of detecting whether an inhomogeneous random graph contains a planted community. Specifically, we observe a single realization of a graph. Under the null hypothesis, this graph is a sample from an inhomogeneous random graph, whereas under the alternative, there exists a small subgraph where the edge probabilities are increased by a multiplicative scaling factor. We present a scan test that is able to detect the presence of such a planted community, even when this community is very small and the underlying graph is inhomogeneous. We also derive an information theoretic lower bound for this problem which shows that in some regimes the scan test is almost asymptotically optimal. We illustrate our results through examples and numerical experiments.

4.1 Introduction

Many complex systems can be described by networks of vertices connected by edges. Usually, these systems can be organized in communities, with certain groups of vertices being more densely connected than others. A central topic in the analysis of these systems is that of community detection where the goal is to find these more densely connected groups. This can often reveal interesting properties of the network with important applications in sociology, biology, computer science, and many other areas of science [76].

Much of the community detection literature is concentrated around methods that extract the communities from a given network, see [92, 142, 143]. These methods typically output an estimate of the community structure regardless of whether it really is present. Therefore, it is important to investigate when an estimated community structure is meaningful and when it simply is an artifact of the algorithm.

To answer this question, it has been highly fruitful to analyze the performance of these methods on random graphs with a known community structure. The stochastic block model is arguably the simplest model that still captures the relevant community structure, and the study of this model has led to many interesting results [1, 35, 51, 124, 135, 136]. However, there are significant drawbacks because of this simplicity: the communities are typically assumed to be very large (i.e., linear in the graph size), and the graph is homogeneous within each community (i.e., vertices within a community are exchangeable and, in particular have the same degree distribution).

To overcome these issues, several suggestions have been made. For example, the degree-corrected block model allows for inhomogeneity of vertices within each community [116]. This allows one to model real-world networks more accurately, while remaining tractable enough to obtain results similar to those obtained for the stochastic block model [85, 95, 96, 111, 112]. However, the degree-corrected block model still assumes that communities are large. To detect small communities, Arias-Castro and Verzelen consider a hypothesis testing problem where the goal is not to find communities, but instead decide whether or not any communities structure is present in an otherwise homogeneous graph [11, 12].

In this chapter, we also focus on the detection of small communities and we investigate when it is possible to detect the presence of a small community in an already inhomogeneous random graph. In particular, we present a scan test and provide conditions under which it is able to detect the presence of a small community. These results are valid under a wide variety of parameter choices, including cases where the underlying graph is inhomogeneous. Furthermore, we show that for some parameter choices the scan test is optimal. Specifically, we identify assumptions that ensure that if the conditions of the scan test are reversed then it is impossible for any test to detect such a community.

4.2 Model and results

We consider the problem of detecting a planted community inside an inhomogeneous random graph. This is formalized as a hypothesis testing problem, where we observe a single instance of a simple undirected random graph $G = (V, E)$, with vertex set V and edge set E. We denote the adjacency matrix of G by A, i.e. $A_{ij} = \mathbb{1}_{\{(i,j)\in E\}}$. That is, $A_{ij} = 1$ if and only if there is an edge between the vertices $i, j \in V$. Because we only consider simple graphs, we have $A_{ii} = 0$ for all $i \in V$.

Under the null hypothesis, denoted by H_0, the observed graph is an inhomogen-

eous random graph on $|V| = n$ vertices, where an edge between two vertices $i, j \in V$ is present, independently of all other edges, with probability p_{ij}. In other words, the entries of the adjacency matrix A are independent Bernoulli random variables such that $\mathbb{P}_0(A_{ij} = 1) = p_{ij}$. The alternative hypothesis, denoted by H_1, is similar, but within a subset of the vertices the connection probabilities are increased. Formally, there is a subset $C \subseteq V$ of size $|C| = r$, called the planted community, for which the edge probabilities are increased by a multiplicative scaling factor $\rho_C \geq 1$. Concretely, under the alternative hypothesis the edge probabilities are $\mathbb{P}_1(A_{ij} = 1) = \rho_C p_{ij}$ for $i, j \in C$ and $\mathbb{P}_1(A_{ij} = 1) = p_{ij}$ otherwise. Note that the scaling ρ_C is allowed to depend on the location of the planted community $C \subseteq V$. This is necessary because our graphs are inhomogeneous, making the problem difficulty dependent on the location of the planted community $C \subseteq V$. Specifically, on a sparse region of the graph it is relatively difficult to detect a planted community so a strong signal ρ_C is required to ensure a significant difference between the edge probabilities under the null hypothesis $\mathbb{P}_0(A_{ij} = 1) = p_{ij}$ and the edge probabilities under the alternative hypothesis $\mathbb{P}_1(A_{ij} = 1) = \rho_C p_{ij}$. On the other hand, when the community is planted on a dense region the problem is easier and a smaller signal ρ_C could be sufficient. Throughout this chapter, we assume that the location of the planted community $C \subseteq V$ is unknown, but that we do know its size $|C| = r$. In particular, we focus on the setting where $r \to \infty$ and is much smaller than n.

In our analysis we begin by considering the (unrealistic) case where the parameters p_{ij} are all known. This allows us to get a precise characterization of the statistical difficulty of the problem. In Section 4.2.3 we relax this assumption and show that it is possible to adapt to unknown parameters under some conditions on the structure of the edge probabilities p_{ij}. In particular, there we will assume that the random graph is rank-1, so that $p_{ij} = w_i w_j$ for some vertex weights $(w_i)_{i=1}^n$.

To summarize, our goal is to decide whether a given graph contains a planted community, or equivalently to decide between the hypotheses:

H_0: There is no planted community, that is

$$A_{ij} \sim \begin{cases} \text{Bern}(p_{ij}), & \text{if } i \neq j, \\ 0, & \text{otherwise.} \end{cases}$$

H_1: There exists a planted community $C \subseteq V$ of size $|C| = r$, and $\rho_C > 1$, such that

$$A_{ij} \sim \begin{cases} \text{Bern}(\rho_C p_{ij}), & \text{if } i \neq j, \text{ and } i, j \in C, \\ \text{Bern}(p_{ij}), & \text{if } i \neq j, \text{ and } i \notin C \text{ or } j \notin C, \\ 0, & \text{otherwise.} \end{cases}$$

Note that in the above definition we are implicitly assuming that ρ_C is not too large, so that $\rho_C p_{ij} \leq 1$ for all $i, j \in C$.

Given a graph, we want to determine which of the above models gave rise to the observation. A test T_n is any function taking as input a graph g on n vertices,

and that outputs either $T_n(g) = 0$ to claim that there is reason to believe that the null hypothesis is true (i.e., no community is present) or $T_n(g) = 1$ to deem the alternative hypothesis true (i.e., the graph contains a planted community). The worst-case risk of such a test is defined as

$$R_n(T_n) := \mathbb{P}_0(T_n \neq 0) + \max_{C \subseteq V, |C|=r} \mathbb{P}_C(T_n \neq 1), \tag{4.1}$$

where $\mathbb{P}_0(\cdot)$ denotes the distribution under the null hypothesis, and $\mathbb{P}_C(\cdot)$ denotes the distribution under the alternative hypothesis when $C \subseteq V$ is the planted community. A sequence of tests $(T_n)_{n=1}^{\infty}$ is called asymptotically powerful when it has vanishing risk, that is $R_n(T_n) \to 0$, and asymptotically powerless when it has risk tending to 1, that is $R_n(T_n) \to 1$.

Notation. Our primary goal is to characterize the asymptotic distinguishability between the null and alternative hypothesis as the graph size n increases. Throughout this chapter, when limits are unspecified they are taken as the graph size satisfies $n \to \infty$. The other parameters p_{ij}, ρ_C, and r are allowed to depend on n, although this dependence is left implicit to avoid notational clutter. Furthermore, we use standard asymptotic notation as described in Section 1.5.

We write $e(C) := \sum_{i,j \in C} A_{ij}$ for the number of edges in the subgraph induced by $C \subseteq V$, and $e(C, -C) := \sum_{i \in C, j \notin C} A_{ij}$ for the number of edges between C and its complement $-C = V \setminus C$. Finally, define the entropy function

$$h(x) := (x+1)\log(x+1) - x. \tag{4.2}$$

This function plays a prominent role in most of the results in this chapter.

4.2.1 Information theoretic lower bound

We start with a result highlighting conditions under which all tests are asymptotically powerless. Here we assume that the edge probabilities p_{ij}, the scaling parameters ρ_C, and the size of the planted community $|C| = r$ are all known. When some of these parameters are unknown, the problem of detecting a planted community might become more difficult, hence any test that is asymptotically powerless when these parameters are known remains asymptotically powerless when they are unknown.

We prove a lower bound under two different sets of assumptions. To state these assumptions we define the *average edge probability* as $\overline{p}_D = \mathbb{E}_0[e(D)]/\binom{|D|}{2}$ for any $D \subseteq V$. Our assumptions correspond to different regimes of the problem in terms of planted community size r. For large communities we need to restrict, in a moderate way, the amount of inhomogeneity in the underlying graph, with larger communities requiring stronger restrictions on the amount of inhomogeneity. This results in the following assumption:

Assumption 1.1. There exists $\delta \in (0, 1/2)$ such that the following conditions hold:

(i) The planted community cannot be too large, that is $r = O(n^{1/2-\delta})$.

(ii) On subgraphs D much smaller than the planted community C, the relative edge density $\overline{p}_D/\overline{p}_C$ cannot be too large. That is, there exists $0 < \gamma_n = o(1)$ such that

$$\max_{C\subseteq V,|C|=r} \max_{\substack{D\subseteq C,\\ |D|<r/(n/r)^{\gamma_n}}} \frac{|D|\,\overline{p}_D}{|C|\,\overline{p}_C} \leq \delta. \tag{4.3}$$

(iii) Every potential community C must be dense enough. Specifically,

$$\max_{C\subseteq V,|C|=r} \frac{1}{\overline{p}_C} = o\left(\frac{r}{\log(n/r)}\right). \tag{4.4}$$

Note that the inhomogeneity restriction in Assumption 1.1 (ii) only applies to small subsets $D \subseteq C$. In particular, we have $|D|/|C| < (r/n)^{\gamma_n}$ in (4.3), and thus if the edge probabilities differ by at most a multiplicative factor of $O(\log(n)^k)$, for some fixed constant $k > 0$, then (4.3) can always be satisfied by choosing a sequence γ_n that converges to zero slowly enough. For example, in the homogeneous setting where the graph is an Erdős-Rényi random graph we know that all edge probabilities are equal and therefore (4.3) is easily satisfied for any fixed $\delta \in (0, 1/2)$.

If the planted community size r is much smaller than allowed by Assumption 1.1, then it is not needed to have a restriction on the inhomogeneity, provided that the graph is dense enough. This gives the following assumption:

Assumption 1.2. We assume that the following two conditions hold:

(i) The planted community is small enough. We require that $r = n^{o(1)}$.

(ii) Every potential community C must be dense enough. Specifically,

$$\max_{C\subseteq V,|C|=r} \log\left(\frac{1}{\overline{p}_C}\right) = o\left(\frac{\log(n/r)}{\log(r)}\right).$$

Note that we only need one of the two assumptions above to hold in order to prove the lower bound in this section. The difference between these two assumptions is that Assumption 1.1 works best when the planted community is large, whereas Assumption 1.2 is more easily satisfied if the planted community is small. Furthermore, we need that the underlying graph is not too dense. This is made precise in the following assumption:

Assumption 2. We require that $\max_{C\subseteq V,|C|=r} \max_{i,j\in C} \rho_C^2 p_{ij} \to 0$ as $n \to \infty$.

This assumption accomplishes two goals. First, since $\rho_C > 1$ it forces $p_{ij} \to 0$ for every $i, j \in V$. This ensures that the number of edges in subsets of the vertices is in essence a sufficient statistic for the testing problem. Secondly, at a more technical level, $p_{ij} \to 0$ is necessary for the Poisson approximations we use and it ensures

that the differences in edge probabilities p_{ij} are not magnified too much under the alternative. We note that Assumption 2 is not needed when the underlying graph is homogeneous (i.e., when the null hypothesis corresponds to an Erdős-Rényi random graph), see [11].

We further discuss Assumptions 1.1, 1.2, and 2 in more detail in Section 4.3. In that section we give several examples of random graphs that satisfy these assumptions.

This brings us to the main result of this section, providing conditions under which all tests are asymptotically powerless by deriving a minimax lower bound:

Theorem 4.1. *Suppose that Assumption 2 and either Assumption 1.1 or 1.2 holds. Let $0 < \varepsilon < 1$ be fixed. Then all tests are asymptotically powerless if, for all $C \subseteq V$ of size $|C| = r$,*

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)]h(\rho_C - 1)}{|D| \log(n/|D|)} \leq 1 - \varepsilon. \tag{4.5}$$

Condition (4.5) has its counterpart in the work by Arias-Castro and Verzelen [11, see (9)], who derive a similar result when the underlying graph is an Erdős-Rényi random graph. However, because of the inhomogeneity in our graphs, the maximum in (4.5) is not necessarily attained at the planted community $C \subseteq V$ of size $|C| = r$, but it could be attained at any of its smaller subgraphs $D \subseteq C$. This is why our condition is more complex.

The result in Theorem 4.1 happens to be tight, even in some scenarios where the edge probabilities p_{ij} are unknown, as we construct a scan test that is powerful when the inequality in (4.5) is, roughly speaking, reversed. This is described in the next sections.

Finally, the proof of Theorem 4.1 is given in Section 4.5.5 and follows a common methodology in these cases, by first reducing the composite alternative hypothesis to a simple alternative hypothesis and then characterizing the optimal likelihood ratio test. This is done via a second-moment method, but it requires a highly careful truncation argument to attain the sharp characterization above.

4.2.2 Scan test for known edge probabilities

In this section we present a scan test that is asymptotically powerful. We first consider the case where all edge probabilities p_{ij} and the community size $|C| = r$ are known. Although this case is unrealistic in practice, it allows us to understand the fundamental statistical limits of detection. In a sense, knowing the edge probabilities p_{ij} is the most optimistic scenario, and so the focus is primarily on whether or not it is possible to detect a planted community. In the subsequent section we relax this assumption by showing how the scan test can be extended when the edge probabilities p_{ij} are unknown.

Our test statistic is inspired by Bennett's inequality (see [36, Theorem 2.9]), which ensures that, for any $t > 0$,

$$\mathbb{P}_0(e(D) - \mathbb{E}_0[e(D)] \geq t) \leq \exp\left(-\mathbb{E}_0[e(D)]h\left(\frac{t}{\mathbb{E}_0[e(D)]}\right)\right), \tag{4.6}$$

where we recall that $h(x) = (x+1)\log(x+1) - x$. Note that this inequality is also valid when we are under the alternative hypothesis (by simply changing the subscripts 0 to C). Plugging in $t = \mathbb{E}_0[e(D)]h^{-1}(s/\mathbb{E}_0[e(D)])$ yields the bound

$$\mathbb{P}_0\left(\mathbb{E}_0[e(D)]h\left(\left[\frac{e(D)}{\mathbb{E}_0[e(D)]} - 1\right]_+\right) \geq s\right) \leq \mathrm{e}^{-s}. \tag{4.7}$$

This result motivates the use of the statistic

$$\tau_D^{\mathsf{k}} \coloneqq \frac{\mathbb{E}_0[e(D)]h\left([e(D)/\mathbb{E}_0[e(D)] - 1]_+\right)}{|D|\log(n/|D|)}, \tag{4.8}$$

where the superscript k is used to differentiate between the setting with *known* edge probabilities, and the setting with *unknown* edge probabilities in the next section. Note that the statistic τ_D^{k} can be computed because $\mathbb{E}_0[e(D)]$ is a function of the known edge probabilities p_{ij}.

To construct our test, we simply scan over the whole graph, rejecting the null hypothesis when there exists a subgraph $D \subseteq V$ of size $|D| \leq r$ with an unusually high value for τ_D^{k}. To be precise, fix $\varepsilon > 0$, then the scan test rejects the null hypothesis when

$$\tau^{\mathsf{k}} \coloneqq \max_{D \subseteq V, |D| \leq r} \tau_D^{\mathsf{k}} \geq 1 + \frac{\varepsilon}{2}. \tag{4.9}$$

This test is essentially based on the number of edges $e(D)$ in subsets $D \subseteq V$ of size $1 \leq |D| \leq r$; rejecting the null hypothesis when there exists a subset $D \subseteq V$ for which the number of edges $e(D)$ becomes substantially larger than its expectation $\mathbb{E}_0[e(D)]$. So we are essentially looking for an *overly* dense subset. Furthermore, the reason we need to scan over subsets smaller than r is because of the possible inhomogeneity in our model; some edges carry little information and therefore it can be beneficial to ignore these edges and simply scan over a smaller subgraph instead.

Note that the proposed test is not computationally practical due to the very large number of sets one must consider in the scan (unless r is very small). However, in this chapter we are primarily interested in characterizing the statistical limits of possible tests, apart from computational considerations. See also the discussion in Section 4.4.

In order for the scan test to be powerful under the alternative we need $\mathbb{E}_C[e(D)] \to \infty$ for the most informative subgraph $D \subseteq C$, because otherwise there is a non-vanishing probability that $e(D)$ contains no edges under the alternative (by a standard

Poisson approximation), making it impossible for the scan test to detect the planted community. This subgraph is characterized in the following definition:

Definition 4.1. For every subgraph C of size $|C| = r$, the most informative subgraph is

$$D^\star := \arg\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)]}{|D| \log(n/|D|)}. \tag{4.10}$$

The subgraph $D^\star$ in the definition above is essentially the densest subgraph under the null hypothesis. Using the above we can state the main result of this section, which provides conditions under which the scan test in (4.9) is asymptotically powerful:

Theorem 4.2. *Suppose that all edge probabilities p_{ij} and the community size r are known. Then the scan test* (4.9) *is asymptotically powerful when $r = o(n)$, $\mathbb{E}_C[e(D^\star)] \to \infty$ for all $C \subseteq V$ of size $|C| = r$, and*

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)] h\big(\rho_C - 1\big)}{|D| \log(n/|D|)} \geq 1 + \varepsilon, \tag{4.11}$$

where $\varepsilon > 0$ comes from the definition of the scan test in (4.9).

This result is more widely applicable than the lower bound from Theorem 4.1. The condition $\mathbb{E}_C[e(D^\star)] \to \infty$ is less stringent than either Assumption 1.1 or 1.2. Also, there is no need for a condition like Assumption 2. This is because we can use the upper bound from Bennett's inequality and therefore do not need the Poisson approximations necessary in deriving the lower bounds. To make this precise and to make the result in Theorem 4.2 directly comparable to Theorem 4.1 we provide the following corollary:

Corollary 4.1. *Suppose that all edge probabilities p_{ij} and the community size r are known, and that either Assumption 1.1 or 1.2 holds. Then the scan test in* (4.9) *is asymptotically powerful when for all $C \subseteq V$ of size $|C| = r$,*

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)] h\big(\rho_C - 1\big)}{|D| \log(n/|D|)} \geq 1 + \varepsilon, \tag{4.12}$$

where $\varepsilon > 0$ comes from the definition of the scan test in (4.9).

To show that Theorem 4.2 applies in a broader setting than the lower bound from Theorem 4.1 we also provide the following corollary. This shows that the scan test (4.9) is able to detect large communities (of size larger than $\sqrt{n}$), even when the edge probabilities are very small and highly inhomogeneous:

Corollary 4.2. *Suppose that all edge probabilities p_{ij} and the community size r are known. Define $p_{\max} := \max_{i,j\in V} p_{ij}$ and $p_{\min} := \min_{i\neq j\in V} p_{ij}$. If $r \geq n^a$, $p_{\min} \geq n^{-2b}$, and $p_{\max}/p_{\min} = o(n^{a-b})$ for $0 < b < a < 1$, then the scan test in* (4.9) *is asymptotically powerful when for all $C \subseteq V$ of size $|C| = r$,*

$$\max_{D\subseteq C} \frac{\mathbb{E}_0[e(D)]h\big(\rho_C - 1\big)}{|D|\log(n/|D|)} \geq 1 + \varepsilon, \tag{4.13}$$

where $\varepsilon > 0$ comes from the definition of the scan test in (4.9)*.*

In the corollary above, both a and b above may depend on the graph size n. In particular, if $p_{\max}/p_{\min} = O(1)$ then it is possible that $a - b = o(1)$, provided that $(a - b)\log(n) \to \infty$. For instance, it is necessary to have $a - b = o(1)$ in order to satisfy Assumption 1.1 (ii).

A downside of the scan test presented in this section is that it requires knowledge of all edge probabilities p_{ij}. In practice, these are often unavailable to a statistician. The next section is devoted to extending the scan test to cope with unknown edge probabilities, assuming that the edge probabilities have a rank-1 structure.

4.2.3 Scan test for unknown rank-1 edge probabilities

In this section we show how the scan test from the previous section can be extended to the setting where the edge probabilities p_{ij} are unknown. We do still assume that the community size $|C| = r$ is known. As can be seen in (4.8), the scan statistic depends on the edge probabilities p_{ij} only through $\mathbb{E}_0[e(D)] = \sum_{i<j\in D} p_{ij}$. Therefore, a natural way to approach the situation where the edge probabilities p_{ij} are unknown is to devise a good surrogate for $\mathbb{E}_0[e(D)]$ that can be computed solely based on the observed graph (which could be a sample from either the null hypothesis or alternative hypothesis). Clearly, this is not possible in full generality, but if the edge probabilities have some additional structure then this become possible.

Here we consider the scenario where, under the null hypothesis, the edge probabilities p_{ij} have a so-called rank-1 structure. The resulting model is sometimes also called a hidden-variable model. That is, we assume that each vertex $i \in V$ is assigned a weight $w_i \in (0, 1)$ and that the edge probabilities are given by $p_{ij} = w_i w_j$. This is probably one of the simplest models for inhomogeneous random graphs possible. Note that this model is very similar to the degree corrected stochastic block model [85, 95, 116], except that our focus is on the detection of small communities, whereas the literature on stochastic block models is typically concerned with the detection of much larger communities. Further, there are strong connections between this model and the configuration model [43, 104].

To make it possible to estimate $\mathbb{E}_0[e(D)]$ we need to assume that the graph is not too inhomogeneous and not too sparse, as formulated in the following assumption:

Assumption 3. Let $w_{\max} = \max_{i\in V} w_i$ and $w_{\min} = \min_{i\in V} w_i$, then the maximum allowed inhomogeneity is

$$\left(\frac{w_{\max}}{w_{\min}}\right)^2 = o\left(r^{2/3} \wedge \frac{n}{r} w_{\min}^2\right). \tag{4.14}$$

Using the above assumption, we will show that it is possible to estimate $\mathbb{E}_0[e(D)]$ by using the observed edges going from D to the rest of the graph $-D = V \setminus D$. Note that the exponent 2/3 in Assumption 3 is not an arbitrary choice, but as we explain below, it is actually the best possible exponent that still ensures that our estimator works.

When $C \subseteq V$ is the planted community, and we estimate $\mathbb{E}_0[e(D)]$ for a large enough subgraph $D \subseteq C$ using this approach, we will obtain an almost unbiased estimate both under H_0 as well as under H_1. This is because enough of the edges used in this estimate have the same distribution under the null and alternative hypothesis. Our estimator is based on the identity

$$\mathbb{E}_0[e(D)] = \frac{\left(\sqrt{\mathbb{E}_0[e(V)] + \frac{1}{2}\sum_{i\in V} w_i^2} - \sqrt{\mathbb{E}_0[e(V)] + \frac{1}{2}\sum_{i\in V} w_i^2 - 2\mathbb{E}_0[e(D,-D)]}\right)^2}{4} - \frac{1}{2}\sum_{i\in D} w_i^2. \tag{4.15}$$

This identity is explained in more detail in Section 4.5.6.5, and it is valid when Assumption 3 holds and n is large enough. Note that both $\mathbb{E}_0[e(V)]$ and $\mathbb{E}_0[e(D,-D)]$ are the sum of a large number of edge probabilities $p_{ij} = w_i w_j$, and most of these remain unaffected under the alternative hypothesis. Because of this, and since $\sum_{i\in V} w_i^2$ will generally be negligible, we will estimate $\mathbb{E}_0[e(D)]$ by

$$\widehat{e(D)} := \frac{\left(\sqrt{e(V)} - \sqrt{e(V) - 2e(D,-D)}\right)^2}{4}. \tag{4.16}$$

Here we have used that $(w_{\max}/w_{\min})^2 \leq r^{2/3}$ by Assumption 3, which ensures that the term $\sum_{i\in D} w_i^2/2$ in (4.15) becomes negligible, and therefore that our estimator $\widehat{e(C)}$ is a good surrogate for $\mathbb{E}_0[e(D)]$. This also explains the exponent 2/3 appearing in Assumption 3, as this is the largest exponent that still guarantees that the term $\sum_{i\in D} w_i^2/2$ is negligible. This is discussed in more detail in Section 4.5.2.

In most cases, the estimator in (4.16) can essentially be used as a plugin for the scan test of the previous section. However, this estimator might not concentrate very well when $\mathbb{E}_0[e(D)]$ becomes too small. To remedy this, we use a thresholded version of the estimator given by

$$\widehat{e(D)}^{\vee} := \left(\widehat{e(D)} \vee \frac{|D|^2}{n}\log^4(n/|D|)\right). \tag{4.17}$$

Using the thresholded estimator in (4.17), we can consider the same scan test as in the previous section but with $\mathbb{E}_0[e(D)]$ replaced by the estimator $\widehat{e(D)}^{\vee}$. This leads to the definition of the scan test for unknown edge probabilities as

$$\tau_D^{\mathsf{u}} := \frac{\widehat{e(D)}^{\vee} h\Big(\big[e(D)/\widehat{e(D)}^{\vee} - 1\big]_+\Big)}{|D| \log(n/|D|)}, \tag{4.18}$$

where the superscript u is used to indicate that we consider the setting with *unknown* rank-1 edge probabilities.

As in the previous section, we scan over subgraphs and reject the null hypothesis when τ_D^{u} becomes too large. However, as explained above, when scanning over subgraphs $D \subseteq V$ whose size $|D|$ is much smaller than $|C| = r$ we run into a problem because of the bias in $\widehat{e(D)}^{\vee}$. Luckily this is not a problem because Assumption 3 ensures that asymptotically the maximum of τ_D^{u} will always be attained at a subgraph of size $|D| \geq r^{1/3}$, see the proof of Lemma 4.1 in Section 4.5.2. Therefore, for $\varepsilon > 0$ fixed, the scan test for unknown edge probabilities rejects the null hypothesis when

$$\tau^{\mathsf{u}} := \max_{D \subseteq V,\, r^{1/3} \leq |D| \leq r} \tau_D^{\mathsf{u}} \geq 1 + \frac{\varepsilon}{3}. \tag{4.19}$$

This brings us to the main result of this section, which provides conditions for the scan test in (4.19) to be asymptotically powerful:

Theorem 4.3. *Suppose that the community size r is known and that Assumption 3 holds. Then the scan test* (4.19) *is asymptotically powerful when $r = o(n)$, $\mathbb{E}_C[e(D^\star)] \to \infty$ for all $C \subseteq V$ of size $|C| = r$, and*

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)] h\big(\rho_C - 1\big)}{|D| \log(n/|D|)} \geq 1 + \varepsilon, \tag{4.20}$$

where $\varepsilon > 0$ comes from the definition of the scan test in (4.19).

Comparing this result with Theorem 4.2, we see that for rank-1 random graphs, Assumption 3 is the only extra condition necessary when the edge probabilities are unknown. Furthermore, by the same argument as in the previous section it can be shown that either Assumption 1.1 or 1.2 is sufficient to ensure that $\mathbb{E}_C[e(D^\star)] \to \infty$. Therefore, to make the result in Theorem 4.3 directly comparable to Theorem 4.1 we provide the following corollary:

Corollary 4.3. *Suppose that the community size r is known and that Assumption 3, and either Assumption 1.1 or 1.2 holds. Then the scan test* (4.9) *is asymptotically powerful when, for all $C \subseteq V$ of size $|C| = r$,*

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)] h\big(\rho_C - 1\big)}{|D| \log(n/|D|)} \geq 1 + \varepsilon, \tag{4.21}$$

where $\varepsilon > 0$ comes from the definition of the scan test in (4.19).

Moreover, a result similar to Corollary 4.2 also applies in the setting with unknown edge probabilities. This leads to the following result:

Corollary 4.4. *Suppose the community size r is known and that Assumption 3 holds. If $r \geq n^a$, $w_{\min} \geq n^{-b}$, and $(w_{\max}/w_{\min})^2 = o(n^{a-b})$ for $0 < b < a < 1$, then the scan test in* (4.9) *is asymptotically powerful when for all $C \subseteq V$ of size $|C| = r$,*

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)]h\big(\rho_C - 1\big)}{|D| \log(n/|D|)} \geq 1 + \varepsilon, \tag{4.22}$$

where $\varepsilon > 0$ comes from the definition of the scan test in (4.19)*.*

4.3 Examples

The results in the previous section provide conditions for when it is possible to detect a planted community $C \subseteq V$. When the scaling ρ_C is large enough it is asymptotically possible to detect a planted community using the scan test, and when the scaling ρ_C is too small it is impossible for any test to detect a planted community. To understand at which scaling ρ_C this change happens, we need to characterize the behavior of

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)]h\big(\rho_C - 1\big)}{|D| \log(n/|D|)} \approx 1. \tag{4.23}$$

The subgraph that attains the maximum above will be denoted by $D^\star = D^\star$ and was defined in Definition 4.1. In this section, we present several examples of different random graph models and illustrate how (4.23) depends on the inhomogeneity structure. For clarity of presentation, the parameters in these examples are chosen such that the scaling ρ_C satisfying (4.23) always converges to a constant.

In the examples below, the lower bound from Theorem 4.1 as well as the upper bound from Theorems 4.2 and 4.3 are applicable because Assumptions 1.2, 2 and 3 are all satisfied[1]. Furthermore, it can be checked that Assumption 1.1 (i) and (ii) are also satisfied. Thus, the only reason why Assumption 1.1 does not hold in the examples below is because the edge density condition from Assumption 1.1 (iii) is not satisfied. The reason for this is that it is not possible to simultaneously satisfy that edge density condition and have the scaling ρ_C from (4.23) converge to a constant larger than 1. This means that a choice had to be made between either selecting examples that satisfy Assumption 1.1 or having $\rho_C - 1$ converge to a positive constant. We choose for the latter option to improve the clarity of presentation.

There are, however, also many interesting examples where Assumption 1.1 does hold. For instance, it is possible to satisfy Assumption 1.1 in any of the examples

[1]The examples in Section 4.3.4 consider randomly sampled vertex weights, and therefore the assumptions in this section hold with high probability. Furthermore, this section also contains some examples where Assumption 1.1 instead of Assumption 1.2 holds.

below by simply increasing the community size r or the edge density (by increasing all vertex weights by the same factor). Thus, in the examples below, it is possible to apply Theorems 4.1, 4.2, and 4.3 because Assumptions 1.2, 2 and 3 hold, and this remains true for larger community sizes or denser graphs but then because of Assumptions 1.1, 2 and 3. This explains how Assumptions 1.1 and 1.2 are nicely complementing each other to make our results applicable in a wide range of scenarios.

4.3.1 Erdős-Rényi random graph

The arguably simplest setting where we can apply our results is that of an Erdős-Rényi random graph, where all edge probabilities $p_{ij} = p$ are equal, so that the graph is completely homogeneous. In this case, the subgraph $D^\star$ that attains the maximum in (4.23) is always the complete planted community $C \subseteq V$. Let $r = o(n)$, $r \to \infty$ and $p \to 0$ be such that $r^2 p \to \infty$. One easily sees that (4.23) becomes

$$\max_{D \subseteq C} \frac{\mathbb{E}_0[e(D)]h(\rho_C - 1)}{|D| \log(n/|D|)} = \frac{\mathbb{E}_0[e(C)]h(\rho_C - 1)}{|C| \log(n/|C|)} \asymp \frac{rH_p(\rho_C p)}{2\log(n/r)}. \tag{4.24}$$

where $H_p(\rho_C p)$ is the Kullback-Leibler divergence between Bern(p) and Bern($\rho_C p$). Note that this is the same condition found by Arias-Castro and Verzelen, who considered the problem of detecting a planted community in an Erdős-Rényi random graph [11, see (9) and (15)].

4.3.2 Rank-1 random graph with 2 weights

A slightly more complex setting is where the underlying graph has a rank-1 structure with two different weights. Some of the vertices have large weight $w_{\max}$, and the remaining vertices have small weight $w_{\min}$. Therefore, there are three different edge probabilities in the underlying graph: $p_{ij} = w_{\max}^2$ when both endpoints have large weight, $p_{ij} = w_{\min}^2$ when both endpoints have small weight, and $p_{ij} = w_{\max} w_{\min}$ when one of the endpoints has large weight and the other small weight.

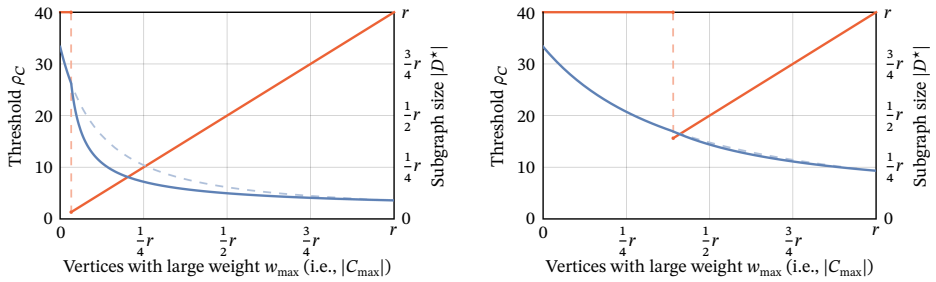
The subgraph $D^\star$ that attains the maximum in (4.23) depends crucially on the amount of inhomogeneity in $C \subseteq V$, and because we only have two different weights this translates to the ratio of vertices with large weight $w_{\max}$ and vertices with small weight $w_{\min}$ in C. Moreover, it can be checked that the maximum in (4.23) is attained either on the whole subgraph C, or on the subgraph $C_{\max} \subseteq C$ consisting of only the large-weight vertices in C. Specifically, assuming $\log(n/|C|) \asymp \log(n)$, the maximum in (4.23) is attained at $C_{\max}$ when

$$|C_{\max}| > (1 + o(1)) \frac{|C| - 1 + (w_{\max}/w_{\min})^2}{(w_{\max}/w_{\min} - 1)^2}, \tag{4.25}$$

and otherwise it is attained at C. Here we can see that the amount of inhomogen-

eity plays an important role in determining the maximum in (4.23), and therefore in determining whether a planted community can be detected or not.

In Figure 4.1 we give two examples of the threshold scaling ρ_C required for the scan test to be asymptotically powerful. When, for every $C \subseteq V$, the scaling ρ_C is chosen above the blue curve then the scan test is asymptotically powerful by Theorems 4.2 and 4.3, and when it is chosen below the blue curve then all tests are asymptotically powerless by Theorem 4.1. Here we can clearly see a sharp bend in the blue curve at the point where $|C_{\max}|$ crosses the threshold in (4.25). This happens because there are many vertices with large weight when $|C_{\max}|$ is large and it is optimal to only use these vertices when trying to detect a planted community. However, there no longer are enough vertices with large weight when $|C_{\max}|$ becomes too small and it becomes more beneficial to also use the vertices with small weight.



(a) $r = \lfloor \log(n)^3 \rfloor$, $w_{\max} = \frac{1}{\log(n)}$, $w_{\min} = \frac{1}{6.5\log(n)}$. (b) $r = \lfloor \log(n)^3 \rfloor$, $w_{\max} = \frac{1}{2.5\log(n)}$, $w_{\min} = \frac{1}{6.5\log(n)}$.

Figure 4.1: Example of the threshold scaling ρ_C required for detecting a planted community using the optimal subgraph $D^\star$ (blue, left axis) and the threshold scaling ρ_C required when using the whole subgraph C instead (dashed blue, left axis), together with the size of the optimal subgraph $|D^\star|$ (red, right axis). The specific numerical values are simply chosen to highlight the different regimes possible; other choices produce similar results.

4.3.3 Rank-1 random graph with 3 weights

Extending the setting in the previous section, we can consider a rank-1 random graph with three different weights. Some vertices have large weight $w_{\max}$, some vertices have medium weight w_{med}, and the remaining vertices have small weight $w_{\min}$. In this setting the situation becomes even more complex, and the subgraph $D^\star$ that attains the maximum in (4.23) depends on the amount of vertices of each type in $C \subseteq V$.

In Figure 4.2 we give an example of the threshold scaling ρ_C required for the scan test to be asymptotically powerful in the setting with three weights. When, for every $C \subseteq V$, the scaling ρ_C is chosen above the surface then the scan test is asymptotically powerful by Theorems 4.2 and 4.3, and when it is chosen below the surface then all tests are asymptotically powerless by Theorem 4.1. We can see that when there are enough vertices with large weight $w_{\max}$ then it is optimal to only use these large-

weight vertices (green region), but as the number of large-weight vertices decreases it becomes beneficial to include also medium-weight vertices (orange region) or even small-weight vertices (blue region). Note that the cross-section with no medium-weight vertices is the same as Figure 4.1(a) and the cross-section with no large-weight vertices is the same as Figure 4.1(b).

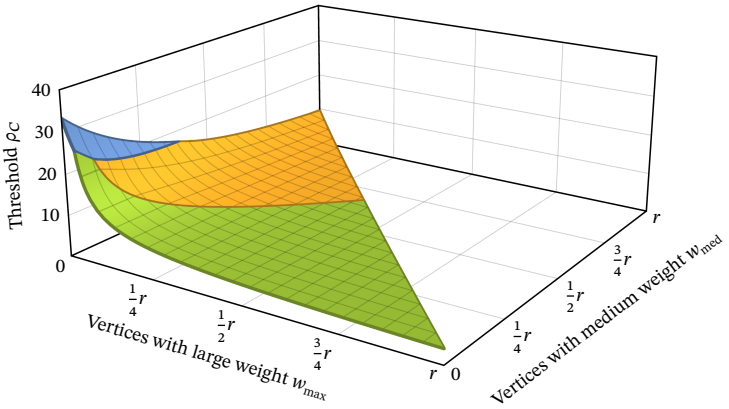


Figure 4.2: Example of the threshold scaling ρ_C required for detecting a planted community when using the optimal subgraph $D^\star$. In the blue region $D^\star$ consists of all vertices, in the orange region $D^\star$ consists of both large and medium-weight vertices, and in the green region $D^\star$ consists only of large-weight vertices. The parameters used are $r = \lfloor \log(n)^3 \rfloor$, $w_{\max} = 1/\log(n)$, $w_{\text{med}} = 1/(2.5\log(n))$, $w_{\min} = 1/(6.5\log(n))$. These values are chosen for ease of comparison with Figure 4.1.

4.3.4 Rank-1 random graph with an arbitrary number of weights

In this section we consider the setting where the graph contains several different vertex weights. In this case it is more difficult to characterize the subgraph $D^\star$ that maximizes (4.23) for a given subgraph $C \subseteq V$, and finding this subgraph becomes optimization problem. This is because, for a given size $|D|$, we only need to consider the subgraph D consisting of the $|D|$ largest weights in C. Using this insight we can approximate (4.23). Let $\hat{F}_C(x)$ be the empirical distribution function of the weights in C, then

$$\begin{aligned}\max_{D\subseteq C} \frac{\mathbb{E}_0[e(D)]h(\rho_C - 1)}{|D|\log(n/|D|)} &\approx \max_{k\in\{1,\dots,r\}} \frac{\binom{k}{2}\left(\frac{r}{k}\int_{\frac{r-k}{r}}^{1} \hat{F}_C^{-1}(y)dy\right)^2 h(\rho_C - 1)}{k\log(n/k)} \\ &\approx \max_{\alpha\in(0,1]} \frac{r}{2\alpha}\frac{\left(\int_{1-\alpha}^{1} \hat{F}_C^{-1}(y)dy\right)^2 h(\rho_C - 1)}{\log(n)},\end{aligned} \tag{4.26}$$

where $\hat{F}_C^{-1}(y) = \inf\{x \in \mathbb{R} : y \leq \hat{F}_C(x)\}$ is the quantile function of $\hat{F}_C(x)$, and we have assumed that $r = n^{o(1)}$ such that $\log(n/r) \asymp \log(n)$ in the second approximation above.

To apply (4.26) we need to know $\hat{F}_C(x)$, which is different for every subgraph $C \subseteq V$. However, instead of characterizing the threshold scaling ρ_C for every subgraph C, we can instead consider a uniformly chosen subgraph C. In this way, if the vertex weights are sampled from a distribution W with distribution function $F(x)$, then we know from the Glivenko-Cantelli theorem that $\hat{F}_C(x)$ will eventually be close to $F(x)$, uniformly in x. With this in mind, we can consider the required threshold scaling ρ_C when C is a uniformly chosen subgraph and the vertex weights are sampled from a distribution W.

In Table 4.1 this is done for a community of size $r = \lfloor \log(n)^4 \rfloor$ and weight distribution $W = (s + X)/\log(n)^{3/2}$, where we consider several different distributions X. We add a small constant s to ensure that none of the vertex weights can become too small and we have normalized the weights by $\log(n)^{3/2}$ to ensure that in each example the maximum weight is less than 1 with high probability. These choices ensure that Assumptions 1.2, 2, and 3 hold with high probability. Furthermore, we have that $r\mathbb{E}[W]^2/\log(n/r) = O(1)$, and by (4.23) this guarantees that $\rho_C = O(1)$, so we obtain a numerical value for ρ_C that is asymptotically independent of n.

Table 4.1: The threshold scaling ρ_C required to detect a planted community C that is planted uniformly at random (based on setting the approximation in (4.26) equal to 1 and then solving for ρ_C). We provide the analytic results together with a numerical example where $\mathbb{E}[X] = 1$. The community size is $r = \lfloor \log(n)^4 \rfloor$.

	W	Threshold ρ_C	$\lvert D^\star \rvert$
$\frac{s+X}{\log(n)^{3/2}}$,	$X \sim \text{Degen}(\delta)$	$h^{-1}\left(\frac{2}{(s+\delta)^2}\right)+1$	r
	$\delta=1$, $s=0.1$	3.311	$1.000 \cdot r$
$\frac{s+tX}{\log(n)^{3/2}}$,	$X \sim \text{Bern}(q)$	$h^{-1}\left(\frac{2}{q(s+t)^2} \wedge \frac{2}{(s+qt)^2}\right)+1$	qr or r
	$q=0.5$, $t=2$, $s=0.1$	2.624	$0.500 \cdot r$
$\frac{s+X}{\log(n)^{3/2}}$,	$X \sim \text{Unif}(a,b)$	$h^{-1}\left(\frac{27}{4}\frac{b-a}{(b+s)^3}\right)+1$	$\frac{2}{3}\frac{b+s}{b-a}r$
	$a=0$, $b=2$, $s=0.1$	3.144	$0.700 \cdot r$
$\frac{s+X}{\log(n)^{3/2}}$,	$X \sim \text{Exp}(\lambda)$	$h^{-1}\left(\frac{\lambda^2}{2e^{s\lambda-1}}\right)+1$	$e^{s\lambda-1}r$
	$\lambda=1$, $s=0.1$	2.939	$0.407 \cdot r$

Table 4.2: The threshold scaling ρ_C required to detect a planted community C that is planted uniformly at random (based on setting the approximation in (4.26) equal to 1 and then solving for ρ_C). We provide the analytic results for community size $r = \lfloor n^{1/4}\log(n)^4 \rfloor$. Note that, the threshold scaling ρ_C is equal to $1 + \Theta(n^{-1/8})$ in these examples because $h(x) \asymp x^2/2$ as $x \to 0$.

W		Threshold ρ_C	$\lvert D^\star \rvert$
$\frac{s+X}{\log(n)^{3/2}}$,	$X \sim \text{Degen}(\delta)$	$h^{-1}\Big(\frac{1}{n^{1/4}}\,\frac{2}{(s+\delta)^2}\Big)+1$	r
$\frac{s+tX}{\log(n)^{3/2}}$,	$X \sim \text{Bern}(q)$	$h^{-1}\Big(\frac{1}{n^{1/4}}\Big(\frac{2}{q(s+t)^2} \wedge \frac{2}{(s+qt)^2}\Big)\Big)+1$	qr or r
$\frac{s+X}{\log(n)^{3/2}}$,	$X \sim \text{Unif}(a,b)$	$h^{-1}\Big(\frac{1}{n^{1/4}}\,\frac{27}{4}\,\frac{b-a}{(b+s)^3}\Big)+1$	$\frac{2}{3}\,\frac{b+s}{b-a}r$
$\frac{s+X}{\log(n)^{3/2}}$,	$X \sim \text{Exp}(\lambda)$	$h^{-1}\Big(\frac{1}{n^{1/4}}\,\frac{\lambda^2}{2\mathrm{e}^{s\lambda-1}}\Big)+1$	$\mathrm{e}^{s\lambda-1}r$

Moreover, in Table 4.2 we consider the same examples as in Table 4.1 but with a larger community size $r = \lfloor n^{1/4}\log(n)^4 \rfloor$. In this case Assumption 1.2 does not hold because the community size r is too large. However, we can now apply Assumption 1.1 instead. To see this, note that Assumption 1.1 (i) and (iii) hold with high probability provided $\delta < 1/4$. Furthermore, Assumption 1.1 (ii) also holds with high probability because the edge probabilities differ by at most a factor $\log(n)^2$ (i.e., $p_{\max}/p_{\min} = O(\log(n)^2)$) with high probability.

This shows that Assumptions 1.1 and 1.2 are nicely complementing each other. For small communities (as in Table 4.1) our results can be applied because Assumptions 1.2, 2, and 3 hold with high probability, and for large communities (as in Table 4.2) our results can still be applied because Assumptions 1.1, 2, and 3 hold with high probability.

4.4 Discussion

In this section we remark on our results and discuss some possibilities for future work.

Unknown community size. When presenting our results, we have always assumed that the size of the planted community is known. In practice, this is often not the case and it would be necessary to estimate the community size before testing. In our case, the scan test can easily be extended to the setting of unknown community size. To see this, note that the scan test can detect any planted community provided that it is not larger than r. Hence, one can simply use the scan test with a large enough value for r and it will detect a planted community of size at most r.

Alternatives to the scan test. When the community size $|C| = r$ becomes much larger than allowed by Assumption 1.1 or 1.2, that is $r \geq \sqrt{n}$, then the scan test is no longer optimal. This was considered by Arias-Castro and Verzelen for an Erdős-Rényi random graph [11], where they show that for large communities, a statistic based on simply counting the total number of edges is optimal. A similar idea can also be applied in the inhomogeneous settings. This suggests that such a test is asymptotically powerful, if for all $C \subseteq V$ of size $|C| = r$,

$$\frac{\mathbb{E}_C[e(C)] - \mathbb{E}_0[e(C)]}{\sqrt{\mathbb{E}_0[e(V)]}} \to \infty. \tag{4.27}$$

Alternatively, when the communities become extremely large such that $r = \Theta(n)$ then our model becomes a version of the degree corrected stochastic block model [116]. In this case, it might be beneficial to consider tests based on spectral methods [35, 95, 124, 136].

Another setting where the scan test is no longer optimal is when the underlying graph is very sparse. In this case, one could consider tests similar to those considered by Arias-Castro and Verzelen [12].

Beyond the rank-1 case. In Section 4.2.3 we consider unknown edge probabilities by additionally assuming a rank-1 structure. This can likely be generalized to edge probabilities that have different structural assumptions, provided Assumption 3 is suitably adjusted. The main difficulty in obtaining a result similar to Theorem 4.3 would then be to find an estimator for $\mathbb{E}_0[e(C)]$ and show a consistency result similar to Lemma 4.2. Such a result will depend heavily on the precise structural assumptions made.

Relaxation of Assumptions 1.1, 1.2, and 2. All assumptions needed to prove the information theoretic lower bound in Section 4.2.1 require that certain conditions hold for all sets $C \subseteq V$ of size $|C| = r$. This can be slightly relaxed because it is only necessary that these conditions hold for most sets $C \subseteq V$. Specifically, there needs to exists a class $\mathcal{C}$ such that the conditions in Assumptions 1.1, 1.2, and 2 hold for all $C \in \mathcal{C}$ and $\bar{\mathbb{P}}(C \in \mathcal{C}) \to 1$, where $\bar{\mathbb{P}}(\cdot)$ denotes probability with respect to a uniformly chosen set $C \subseteq V$ of size $|C| = r$. To see this, one only needs to modify the truncation event in (4.52) to also include all sets $C \notin \mathcal{C}$. That is, one needs to modify the truncation event to $\Gamma'_C = \Gamma_C \cup \{C \notin \mathcal{C}\}$, where Γ_C is the original truncation event from (4.52).

Computational complexity. In general, the computational complexity of scan tests is not polynomial in the graph size n. In the homogeneous settings, it has been conjectured that polynomial time algorithms are not able to achieve the minimax rate [98]. Inhomogeneity in the graphs can make computations easier – for instance in very inhomogeneous cases it is possible to recover the largest clique of a graph in polynomial time [82]. It thus remains an interesting avenue for future work to thoroughly characterize the statistical limits of tests under computational constraints.

4.5 Proofs

In this section we prove our results. We start with the proof of Theorem 4.2 because it is the simplest and it sets the stage for some of the arguments in the proof of Theorem 4.3. We end this section with the proof of Theorem 4.1, which shows that the results obtained in Theorem 4.2 and Theorem 4.3 are, roughly speaking, the best possible.

4.5.1 Proof of Theorem 4.2: Scan test for known edge probabilities is powerful

In this section we prove that the scan test in (4.9) is asymptotically powerful. That is, under the conditions of the theorem, both type-I and type-II errors vanish.

Type-I error. We will show that $\mathbb{P}_0(\tau^{\mathsf{k}} \geq 1 + \varepsilon/2) \to 0$. This is done through a relatively straightforward use of Bennett's inequality and the union bound. Using $\binom{n}{k} \leq \left(\frac{n\,\mathrm{e}}{k}\right)^k$, it follows that

$$\begin{aligned}
\mathbb{P}_0\left(\tau^{\mathsf{k}} \geq 1 + \frac{\varepsilon}{2}\right) &= \mathbb{P}_0\left(\max_{D\subseteq V,\,|D|\leq r} \tau^{\mathsf{k}}_D \geq 1 + \frac{\varepsilon}{2}\right) \\
&= \mathbb{P}_0\left(\max_{1\leq k\leq r}\max_{D\subseteq V,\,|D|=k} \frac{\mathbb{E}_0[e(D)]\, h\left(\left[e(D)/\mathbb{E}_0[e(D)] - 1\right]_+\right)}{k\log(n/k)} \geq 1 + \frac{\varepsilon}{2}\right) \\
&\leq \sum_{1\leq k\leq r}\sum_{D\subseteq V,|D|=k} \mathbb{P}_0\left(\frac{\mathbb{E}_0[e(D)]\, h\left(\left[e(D)/\mathbb{E}_0[e(D)] - 1\right]_+\right)}{k\log(n/k)} \geq 1 + \frac{\varepsilon}{2}\right) \\
&\leq \sum_{1\leq k\leq r}\binom{n}{k}\exp\left(-\left(1+\frac{\varepsilon}{2}\right)k\log\left(\frac{n}{k}\right)\right) \\
&\leq \sum_{1\leq k\leq r}\left(\mathrm{e}\left(\frac{k}{n}\right)^{\varepsilon/2}\right)^k \leq \frac{\mathrm{e}\left(\frac{r}{n}\right)^{\varepsilon/2}}{1-\mathrm{e}\left(\frac{r}{n}\right)^{\varepsilon/2}} \to 0.
\end{aligned}$$

The first and second inequality follow from a simple union bound and Bennett's inequality given in (4.7). The final step relies on the fact that $k/n \leq r/n$ and $r = o(n)$. Therefore we conclude that the scan test (4.9) has vanishing type-I error.

Type-II error. Showing that we have vanishing type-II error starts by realizing that $\mathbb{P}_C(\tau^{\mathsf{k}} \geq 1 + \varepsilon/2) \geq \mathbb{P}_C(\tau^{\mathsf{k}}_{D^\star} \geq 1 + \varepsilon/2)$, for every $C \subseteq V$ of size $|C| = r$, where $D^\star$ was introduced in Definition 4.1. The rest of the proof entails showing that for every $C \subseteq V$ of size $|C| = r$,

$$\tau^{\mathsf{k}}_{D^\star} \geq (1 + o_{\mathbb{P}_C}(1))\,\frac{\mathbb{E}_0[e(D^\star)]\, h(\rho_C - 1)}{|D^\star|\,\log(n/|D^\star|)}. \tag{4.28}$$

Together with (4.11) this implies that, for every C, we have $\mathbb{P}_C(\tau^{\mathsf{k}}_{D^\star} \geq 1 + \varepsilon/2) \to 1$.

Let $C \subseteq V$ be an arbitrary subgraph of size $|C| = r$ and recall $D^\star := D^\star$ from Definition 4.1 (we drop the explicit dependence of $D^\star$ on C to avoid notational clutter). To prove (4.28) it suffices to show that

$$\mathbb{E}_0[e(D^\star)]\, h\Big(\big[e(D^\star)/\mathbb{E}_0[e(D^\star)] - 1\big]_+\Big) \geq (1 + o_{\mathbb{P}_C}(1))\, \mathbb{E}_0[e(D^\star)]\, h\,(\rho_C - 1)\,. \quad (4.29)$$

To see this, note that $x \mapsto h(x-1)$ is convex, with derivative $h'(x-1) = \log(x)$ and therefore $h(x-1) \geq h(y-1) + (x-y)\log(y)$. Using this, together with $x = e(D^\star)/\mathbb{E}_0[e(D^\star)]$ and $y = \mathbb{E}_C[e(D^\star)]/\mathbb{E}_0[e(D^\star)] = \rho_C > 1$, we obtain the lower bound

$$\begin{aligned}
&\mathbb{E}_0[e(D^\star)]\, h\bigg(\bigg[\frac{e(D^\star)}{\mathbb{E}_0[e(D^\star)]} - 1\bigg]_+\bigg) - \mathbb{E}_0[e(D^\star)]\, h\,(\rho_C - 1) \\
&\qquad = \mathbb{E}_0[e(D^\star)]\, h\bigg(\bigg[\frac{e(D^\star)}{\mathbb{E}_0[e(D^\star)]} - 1\bigg]_+\bigg) - \mathbb{E}_0[e(D^\star)]\, h\bigg(\frac{\mathbb{E}_C[e(D^\star)]}{\mathbb{E}_0[e(D^\star)]} - 1\bigg) \\
&\qquad \geq \big(e(D^\star) - \mathbb{E}_C[e(D^\star)]\big) \log\bigg(\frac{\mathbb{E}_C[e(D^\star)]}{\mathbb{E}_0[e(D^\star)]}\bigg) \\
&\qquad = \big(e(D^\star) - \mathbb{E}_C[e(D^\star)]\big) \log\,(\rho_C)\,.
\end{aligned}$$

It follows by Chebyshev's inequality that

$$\big(e(D^\star) - \mathbb{E}_C[e(D^\star)]\big) \log\,(\rho_C) = O_{\mathbb{P}_C}\Big(\sqrt{\mathbb{E}_C[e(D^\star)]}\log(\rho_C)\Big).$$

Therefore, the inequality in (4.29) holds when

$$\frac{\sqrt{\mathbb{E}_C[e(D^\star)]}\log(\rho_C)}{\mathbb{E}_0[e(D^\star)]h\,(\rho_C - 1)} = o(1). \quad (4.30)$$

To show this, we consider three cases depending on the asymptotic behavior of ρ_C. Although these three cases do not cover all possibilities, they suffice, by the argument in Remark 4.1 below.

Case 1 ($\rho_C \to 1$): Using $\sqrt{x}\log(x) \asymp (x-1)$ as $x \to 1$, and $h(x-1) \asymp (x-1)^2/2$ as $x \to 1$ gives

$$\sqrt{\mathbb{E}_C[e(D^\star)]}\log(\rho_C) = (1 + o(1))\sqrt{\mathbb{E}_0[e(D^\star)]}(\rho_C - 1),$$

and

$$\mathbb{E}_0[e(D^\star)]h(\rho_C - 1) = (1 + o(1))\, \mathbb{E}_0[e(D^\star)](\rho_C - 1)^2/2.$$

Hence, by (4.11) we have

$$\mathbb{E}_0[e(D^\star)](\rho_C - 1)^2 \asymp 2\mathbb{E}_0[e(D^\star)]h(\rho_C - 1) > 2|D^\star| \log(n/|D^\star|) \to \infty\,.$$

Combining the above gives

$$\frac{\sqrt{\mathbb{E}_C[e(D^\star)]}\log(\rho_C)}{\mathbb{E}_0[e(D^\star)]h(\rho_C - 1)} = (1 + o(1))\,\frac{2}{\sqrt{\mathbb{E}_0[e(D^\star)](\rho_C - 1)^2}} = o(1).$$

This shows that (4.30) holds when $\rho_C \to 1$.

Case 2 ($\rho_C \to \alpha \in (1, \infty)$): In this case $\sqrt{\rho_C}\log(\rho_C) = O\left(h(\rho_C - 1)\right)$, and by (4.11) we have $\mathbb{E}_0[e(D^\star)]h(\rho_C - 1) \geq |D^\star| \log(n/|D^\star|) \to \infty$. Therefore

$$\sqrt{\mathbb{E}_C[e(D^\star)]}\log(\rho_C) = \sqrt{\mathbb{E}_0[e(D^\star)]}\sqrt{\rho_C}\log(\rho_C) = o(\mathbb{E}_0[e(D^\star)]h(\rho_C - 1)).$$

This shows that (4.30) holds when $\rho_C \to \alpha \in (1, \infty)$.

Case 3 ($\rho_C \to \infty$): Using $h(x-1) \asymp x\log(x)$ as $x \to \infty$ and because $\mathbb{E}_C[e(D^\star)] \to \infty$ we have

$$\frac{\sqrt{\mathbb{E}_C[e(D^\star)]}\log(\rho_C)}{\mathbb{E}_0[e(D^\star)]h(\rho_C - 1)} = \frac{1}{\sqrt{\mathbb{E}_C[e(D^\star)]}} = o(1). \tag{4.31}$$

This shows that (4.30) holds when $\rho_C \to \infty$, and therefore that (4.29) holds in all the three cases.

Remark 4.1 (General ρ_C sequences)**.** Note that ρ_C might not fit one of the above cases, but may rather oscillate between a combination of the three. However, this is not a problem. For every subsequence of ρ_C, there exists a further subsequence along which the scaling ρ_C satisfies one of the three cases. Hence, (4.30) holds along this (further) subsequence, which implies that (4.30) also holds along the full sequence. This type of argument will be used in several more places in the proofs.

The proof of Theorem 4.2 is now easily completed using (4.28) together with (4.11). For every $C \subseteq V$ of size $|C| = r$,

$$\begin{aligned}
\tau^{\mathsf{k}} \geq \tau^{\mathsf{k}}_{D^\star} &= \frac{\mathbb{E}_0[e(D^\star)]\, h\Big(\big[e(D^\star)/\mathbb{E}_0[e(D^\star)] - 1\big]_+\Big)}{|D^\star|\,\log(n/|D^\star|)} \\
&\geq (1 + o_{\mathbb{P}_C}(1))\,\frac{\mathbb{E}_0[e(D^\star)]\, h\big([\rho_C - 1]_+\big)}{|D^\star|\,\log(n/|D^\star|)} \\
&\geq (1 + o_{\mathbb{P}_C}(1))(1 + \varepsilon).
\end{aligned}$$

Hence, $\mathbb{P}_C\left(\tau^{\mathsf{k}} \geq 1 + \varepsilon/2\right) \to 1$. This shows that the type-II error vanishes, completing the proof. □

4.5.2 Proof of Theorem 4.3: Scan test for unknown rank-1 edge probabilities is powerful

In this section we prove that the scan test in (4.19) is asymptotically powerful, but we first derive some auxiliary results. The first of these shows that if a planted community can be detected then it can be detected based on the evidence of the subgraph $D^\star$ from Definition 4.1. Moreover, by Assumption 3 it follows that $D^\star$ must be relatively large. Specifically, we show that $|D^\star| \geq r^{1/3}$. This explains why the scan test in (4.19) is defined to only scan over subgraphs larger than $r^{1/3}$.

Lemma 4.1. *For any $C \subseteq V$ of size $|C| = r$, let $D^\star$ be as given in Definition 4.1. When Assumption 3 holds then $|D^\star| \geq r^{1/3}$.*

Proof. We use a proof by contradiction. For any $D \subseteq V$ of size $|D| \leq r^{1/3}$, it follows by Assumption 3 that

$$\begin{aligned}\frac{\mathbb{E}_0[e(D)]}{|D|\log(n/|D|)} &\leq \frac{|D|-1}{2}\,\frac{w_{\max}^2}{\log(n/|D|)} \\ &\leq \frac{o(r)}{2}\,\frac{w_{\min}^2}{\log(n/r^{1/3})} \\ &< \frac{|C|-1}{2}\,\frac{w_{\min}^2}{\log(n/|C|)} \leq \frac{\mathbb{E}_0[e(C)]}{|C|\log(n/|C|)}.\end{aligned}$$

Hence, a subset $D \subseteq V$ of size $|D| \leq r^{1/3}$ does not maximize the right-hand side of (4.10), and therefore $|D^\star| \geq r^{1/3}$. □

In the second auxiliary result we quantify the deviations of $\widehat{e(D)}$ around $\mathbb{E}_0[e(D)]$. We note that the lemma below remains true when all $(1 + o_{\mathbb{P}_0}(1))$ terms are replaced by $(1 + o_{\mathbb{P}_C}(1))$ terms. So, this results holds under both the null and alternative hypothesis. This crucial property is key to ensure that we can deal with unknown edge probabilities.

Lemma 4.2. *Let $\mathcal{D}$ be a set of subsets of the vertices V, such that $r^{1/3} \leq |D| \leq r$ for all $D \in \mathcal{D}$. Under Assumption 3 and*

$$\begin{aligned} e(V) &= (1 + o_{\mathbb{P}_0}(1))\mathbb{E}_0[e(V)], \\ e(D,-D) &= (1 + o_{\mathbb{P}_0}(1))\mathbb{E}_0[e(D,-D)], \qquad \textit{uniformly over all } D \in \mathcal{D}.\end{aligned}$$

the deviations of $\widehat{e(D)}$ around $\mathbb{E}_0[e(D)]$ satisfy

$$\frac{\widehat{e(D)}}{\mathbb{E}_0[e(D)]} = 1 + o_{\mathbb{P}_0}(1), \qquad \textit{uniformly over all } D \in \mathcal{D}.$$

Additionally, the statement above remains true when all $(1 + o_{\mathbb{P}_0}(1))$ terms are replaced by $(1 + o_{\mathbb{P}_C}(1))$ terms.

Proof. Define $f(x_1, x_2) \coloneqq \left(\sqrt{x_1} - \sqrt{x_1 - 2x_2}\right)^2$ for $x_1 \geq 2x_2$. Then the partial derivatives of $f(x_1, x_2)$ are given by

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = -\frac{\left(\sqrt{x_1} - \sqrt{x_1 - 2x_2}\right)^2}{\sqrt{x_1}\sqrt{x_1 - 2x_2}} = -\frac{f(x_1, x_2)}{\sqrt{x_1}\sqrt{x_1 - 2x_2}},$$
$$\frac{\partial f}{\partial x_2}(x_1, x_2) = 2\,\frac{\sqrt{x_1} - \sqrt{x_1 - 2x_2}}{\sqrt{x_1 - 2x_2}} = \frac{2\,f(x_1, x_2)}{\sqrt{x_1 - 2x_2}\left(\sqrt{x_1} - \sqrt{x_1 - 2x_2}\right)}.$$

We use a Taylor expansion of $f(x_1, x_2)$ around (a_1, a_2) with $a_1 > 2a_2$. Specifically, there exists (ξ_1, ξ_2) with ξ_1 in between x_1 and a_1, and ξ_2 in between x_2 and a_2, such that

$$f(x_1, x_2) = f(a_1, a_2) + \frac{\partial f}{\partial x_1}(\xi_1, \xi_2)\,(x_1 - a_1) + \frac{\partial f}{\partial x_2}(\xi_1, \xi_2)\,(x_2 - a_2). \tag{4.32}$$

To continue we use (4.32) together with $(x_1, x_2) = (e(V), e(D, -D))$ and $(a_1, a_2) = (\mathbb{E}_0[e(V)], \mathbb{E}_0[e(D, -D)])$. Because $e(V) = (1 + o_{\mathbb{P}_0}(1))\mathbb{E}_0[e(V)]$ by assumption, it follows that for any ξ_1 between $e(V)$ and $\mathbb{E}_0[e(V)]$ we have $\xi_1 = (1 + o_{\mathbb{P}_0}(1))\mathbb{E}_0[e(V)]$. Similarly, by assumption we have $e(D, -D) = (1+o_{\mathbb{P}_0}(1))\mathbb{E}_0[e(D, -D)]$ uniformly over all $D \in \mathcal{D}$, and therefore it follows that $\xi_2 = (1 + o_{\mathbb{P}_0}(1))\mathbb{E}_0[e(D, -D)]$. Hence,

$$\begin{aligned}
&\frac{f(e(V), e(D, -D))}{f(\mathbb{E}_0[e(V)], \mathbb{E}_0[e(D, -D)])} \\
&= 1 - \frac{(1 + o_{\mathbb{P}_0}(1))\,(e(V) - \mathbb{E}_0[e(V)])}{\sqrt{\mathbb{E}_0[e(V)]}\sqrt{\mathbb{E}_0[e(V)] - 2\mathbb{E}_0[e(D, -D)]}} \\
&\quad + \frac{(2 + o_{\mathbb{P}_0}(1))\,(e(D, -D) - \mathbb{E}_0[e(D, -D)])}{\sqrt{\mathbb{E}_0[e(V)] - 2\mathbb{E}_0[e(D, -D)]}\left(\sqrt{\mathbb{E}_0[e(V)]} - \sqrt{\mathbb{E}_0[e(V)] - 2\mathbb{E}_0[e(D, -D)]}\right)} \\
&= 1 - (1 + o_{\mathbb{P}_0}(1))\,\frac{e(V) - \mathbb{E}_0[e(V)]}{\mathbb{E}_0[e(V)]} + (2 + o_{\mathbb{P}_0}(1))\,\frac{e(D, -D) - \mathbb{E}_0[e(D, -D)]}{\mathbb{E}_0[e(D, -D)]} \\
&= 1 + o_{\mathbb{P}_0}(1),
\end{aligned} \tag{4.33}$$

where we have used $\mathbb{E}_0[e(D, -D)] = o(\mathbb{E}_0[e(V)])$ and $\mathbb{E}_0[e(V)] \to \infty$ in the second equality above, which is ensured by Assumption 3. To see this, note that $w_{\min}^2 \leq 1$ and therefore $(w_{\max}/w_{\min})^2 \leq o\big(\frac{n}{r} w_{\min}^2\big) \leq o\big(\frac{n}{r}\big)$, hence

$$\frac{\mathbb{E}_0[e(D, -D)]}{\mathbb{E}_0[e(V)]} \leq (1 + o(1))\frac{|D| n w_{\max}^2}{n^2 w_{\min}^2} \leq \frac{|D|}{n} o\left(\frac{n}{r}\right) = o\left(\frac{|D|}{r}\right) = o(1).$$

To continue, we will show that

$$f\left(e(V), e(D,-D)\right) = 4\,\widehat{e(D)}, \tag{4.34}$$

$$\begin{aligned} f\left(\mathbb{E}_0[e(V)], \mathbb{E}_0[e(D,-D)]\right) &= (1+o(1))\left(4\,\mathbb{E}_0[e(D)] + 2\sum_{i\in D} w_i^2\right) \\ &= (1+o(1))\,4\,\mathbb{E}_0[e(D)]. \end{aligned} \tag{4.35}$$

Here (4.34) follows directly from the definition in (4.16). To obtain the first equality in (4.35) we use Assumption 3 to ensure that $\mathbb{E}_0[e(V)] + \frac{1}{2}\sum_{i\in V} w_i^2 = (1+o(1))\mathbb{E}_0[e(V)]$. This is easily shown since

$$\frac{\mathbb{E}_0[e(V)] + \sum_{i\in V} w_i^2}{\mathbb{E}_0[e(V)]} \le 1 + \frac{n w_{\max}^2}{\binom{n}{2} w_{\min}^2} \le 1 + \frac{n r^{2/3} w_{\min}^2}{\binom{n}{2} w_{\min}^2} = 1 + 2\,\frac{r^{2/3}}{n-1} = 1 + o(1).$$

For the second equality in (4.35) we need to show $\sum_{i\in D} w_i^2 / \mathbb{E}_0[e(D)] = o(1)$. To this end, we first show

$$\frac{\sum_{i\in D} w_i^2}{\left(\sum_{i\in D} w_i\right)^2} \le \frac{1}{4|D|}\,\frac{(w_{\min}+w_{\max})^2}{w_{\min}\,w_{\max}}. \tag{4.36}$$

To see this, note that the ratio $\sum_{i\in D} w_i^2 \big/ \left(\sum_{i\in D} w_i\right)^2$ is maximized when a fraction $\alpha = w_{\min}/(w_{\min}+w_{\max})$ of the vertices in D has weight $w_{\max}$ and the remaining $1-\alpha$ fraction of vertices has weight $w_{\min}$. Plugging this in we obtain (4.36). Then, by Assumption 3 it follows that $w_{\max}/w_{\min} = o(r^{1/3})$ and using that $|D| \ge r^{1/3}$ together with (4.36), we obtain

$$\frac{\sum_{i\in D} w_i^2}{\left(\sum_{i\in D} w_i\right)^2} \le \frac{1}{4|D|}\,\frac{(w_{\min}+w_{\max})^2}{w_{\min}\,w_{\max}} \le \frac{1}{|D|}\,\frac{w_{\max}}{w_{\min}} = \frac{o(r^{1/3})}{r^{1/3}} = o(1).$$

Hence, plugging (4.34) and (4.35) into (4.33) gives

$$\frac{\widehat{e(D)}}{\mathbb{E}_0[e(D)]} = (1+o(1))\,\frac{f(e(V), e(D,-D))}{f(\mathbb{E}_0[e(V)], \mathbb{E}_0[e(D,-D)])} = (1 + o_{\mathbb{P}_0}(1)).$$

Finally, it can easily be checked, using the same steps as above, that the lemma remains true under the alternative hypothesis (i.e., when all $(1 + o_{\mathbb{P}_0}(1))$ terms are replaced by $(1 + o_{\mathbb{P}_c}(1))$ terms). □

We are now ready to prove Theorem 4.3, which shows that the scan test in (4.19) is still asymptotically powerful even when the edge probabilities are not known. To this end, we again show that both the type-I and the type-II error vanish, which we do separately below.

Type-I error. Here we show $\mathbb{P}_0(\tau^{\mathsf{u}} \geq 1+\varepsilon/3) \to 0$. To this end, we show that using the truncated estimator $\widehat{e(D)}^{\vee}$ from (4.17) is asymptotically as good as using $\mathbb{E}_0[e(D)]$. Specifically, we show that uniformly over all subgraphs $D \subseteq V$ of size $r^{1/3} \leq |D| \leq r$,

$$\max_{D\subseteq V,\, r^{1/3}\leq |D|\leq r} \frac{\mathbb{E}_0[e(D)]}{\widehat{e(D)}^{\vee}} \leq 1 + o_{\mathbb{P}_0}(1). \tag{4.37}$$

To show this, define the random set $\mathcal{D} \coloneqq \{D \subseteq V \,:\, r^{1/3} \leq |D| \leq r,\ \widehat{e(D)}^{\vee} \leq \mathbb{E}_0[e(D)]\}$ and rewrite (4.37) as

$$\max_{D\subseteq V,\, r^{1/3}\leq |D|\leq r} \left(\frac{\mathbb{E}_0[e(D)]}{\widehat{e(D)}^{\vee}} \mathbb{1}_{\{D\in\mathcal{D}\}} + \frac{\mathbb{E}_0[e(D)]}{\widehat{e(D)}^{\vee}} \mathbb{1}_{\{D\notin\mathcal{D}\}} \right). \tag{4.38}$$

In the second term above we have $D \notin \mathcal{D}$, so this term is trivially less than or equal to 1. Therefore we will focus on the first term in (4.38). For any $D \in \mathcal{D}$, it follows by definition of the thresholded estimator $\widehat{e(D)}^{\vee}$ in (4.17) that

$$\frac{|D|^2}{n} \log^4\!\left(\frac{n}{|D|}\right) \leq \widehat{e(D)}^{\vee} \leq \mathbb{E}_0[e(D)] \leq \left(\sum_{i\in D} w_i\right)^2, \quad \text{hence} \quad \frac{|D|}{\sqrt{n}} \log^2\!\left(\frac{n}{|D|}\right) \leq \sum_{i\in D} w_i.$$

Now, by the second part of Assumption 3 we have $1 \leq \left(\frac{w_{\max}}{w_{\min}}\right)^2 \leq \frac{n}{r} w_{\min}^2$, and therefore $w_{\min} \geq \sqrt{r/n} \geq 1/\sqrt{n}$. Using this we obtain

$$\begin{aligned} \mathbb{E}_0[e(D,-D)] = \left(\sum_{i\in D} w_i\right)\left(\sum_{j\notin D} w_j\right) &\geq \frac{|D|}{\sqrt{n}} \log^2\!\left(\frac{n}{|D|}\right) \frac{n-|D|}{\sqrt{n}} \\ &= (1+o(1))|D| \log^2\!\left(\frac{n}{|D|}\right). \end{aligned} \tag{4.39}$$

Recall that Bennett's inequality ensures that, for $t > 0$,

$$\mathbb{P}_0(e(D,-D) - \mathbb{E}_0[e(D,-D)] \leq -t) \ \leq\ \exp\left(-\mathbb{E}_0[e(D,-D)]\, h\left(\frac{t}{\mathbb{E}_0[e(D,-D)]}\right)\right).$$

To get a uniform bound over all subgraphs $D \in \mathcal{D}$, we use a union bound together with (4.39). For any $\delta > 0$ and n large enough, this gives

$$\begin{aligned} &\mathbb{P}\left(\min_{D\in\mathcal{D}} e(D,-D) - \mathbb{E}_0[e(D,-D)] \leq -(1+\delta)\sqrt{2\,\mathbb{E}_0[e(D,-D)]\,|D|\log(n/|D|)}\right) \\ &\quad\leq \sum_{1\leq k\leq r} \sum_{D\subseteq V,\,|D|=k} \mathbb{1}_{\left\{\mathbb{E}_0[e(D,-D)]\geq(1-\delta)|D|\log^2(n/|D|)\right\}} \\ &\qquad\times \mathbb{P}\left(e(D,-D) - \mathbb{E}_0[e(D,-D)] \leq -(1+\delta)\sqrt{2\,\mathbb{E}_0[e(D,-D)]\,|D|\log(n/|D|)}\right) \end{aligned}$$

$$\leq \sum_{1\leq k\leq r} \sum_{D\subseteq V,\,|D|=k} \mathbb{1}\big\{\mathbb{E}_0[e(D,-D)]\geq(1-\delta)|D|\log^2(n/|D|)\big\} \times \exp\left(-\mathbb{E}_0[e(D,-D)]\, h\left((1+\delta)\sqrt{\frac{2\,|D|\log(n/|D|)}{\mathbb{E}_0[e(D,-D)]}}\right)\right)$$

$$\leq \sum_{1\leq k\leq r} \binom{n}{k} \exp\left(-(1+\delta)k\log\left(\frac{n}{k}\right)\right) \tag{4.40}$$

$$\leq \sum_{1\leq k\leq r} \left(\mathrm{e}\left(\frac{k}{n}\right)^{\delta}\right)^{k} \leq \frac{\mathrm{e}\left(\frac{r}{n}\right)^{\delta}}{1-\mathrm{e}\left(\frac{r}{n}\right)^{\delta}} \to 0,$$

For the step in (4.40) we have used the result in (4.39) together with $h(x) \asymp x^2/2$ as $x \to 0$, and the final step relies on the fact that $k/n \leq r/n$ and $r = o(n)$.

Then, using the above together with (4.39), it follows that uniformly over $D \in \mathcal{D}$,

$$\frac{e(D,-D)-\mathbb{E}_0[e(D,-D)]}{\mathbb{E}_0[e(D,-D)]} = O_{\mathbb{P}_0}\left(\sqrt{\frac{|D|\log(n/|D|)}{\mathbb{E}_0[e(D,-D)]}}\right) = o_{\mathbb{P}_0}(1). \tag{4.41}$$

To bound the deviations of $e(V)$ we use Chebyshev's inequality,

$$\frac{e(V)-\mathbb{E}_0[e(V)]}{\mathbb{E}_0[e(V)]} = O_{\mathbb{P}_0}\left(\sqrt{\frac{1}{\mathbb{E}_0[e(V)]}}\right) = o_{\mathbb{P}_0}(1). \tag{4.42}$$

Using (4.41) and (4.42), it follows by Lemma 4.2 that uniformly over $D \in \mathcal{D}$,

$$\frac{\widehat{e(D)}^{\vee}}{\mathbb{E}_0[e(D)]} \geq \frac{\widehat{e(D)}}{\mathbb{E}_0[e(D)]} = 1 + o_{\mathbb{P}_0}(1).$$

This shows that the first term in (4.38) is less than or equal to $1+o_{\mathbb{P}_0}(1)$, and therefore that (4.37) holds.

Then, using (4.37) it becomes relatively straightforward to show that the type-I error vanishes. Indeed, note that $a\,h\big(\big[\frac{x}{a}-1\big]_+\big) \leq b\,h\big(\big[\frac{x}{b}-1\big]_+\big)$ for $a > b$, and therefore

$$\mathbb{P}_0\left(\tau^{\mathrm{u}} \geq 1+\frac{\varepsilon}{3}\right) = \mathbb{P}_0\left(\max_{\substack{D\subseteq V,\\ r^{1/3}\leq|D|\leq r}} \frac{\widehat{e(D)}^{\vee}\, h\left(\left[e(D)/\widehat{e(D)}^{\vee}-1\right]_+\right)}{|D|\log(n/|D|)} \geq 1+\frac{\varepsilon}{3}\right)$$

$$\leq \mathbb{P}_0\left(\max_{\substack{D\subseteq V,\\ r^{1/3}\leq|D|\leq r}} \frac{(1+o_{\mathbb{P}_0}(1))\mathbb{E}_0[e(D)]\, h\left(\left[(1+o_{\mathbb{P}_0}(1))\frac{e(D)}{\mathbb{E}_0[e(D)]}-1\right]_+\right)}{|D|\log(n/|D|)} \geq 1+\frac{\varepsilon}{3}\right).$$

Then using the same reasoning as in the proof of Theorem 4.2, it follows that the type-I error vanishes.

Type-II error. Here we show that $\mathbb{P}_C(\tau^{\mathsf{u}} \geq 1+\varepsilon/3) \geq \mathbb{P}_C(\tau^{\mathsf{u}}_{D^\star} \geq 1+\varepsilon/3) \to 1$, for every $C \subseteq V$ of size $|C| = r$, where $D^\star = D^\star$ is defined as in (4.10). To this end, we start by quantifying the deviation of $\widehat{e(D^\star)}/\mathbb{E}_0[e(D^\star)]$ under the alternative.

By Chebyshev's inequality,

$$\frac{e(D^\star,-D^\star)-\mathbb{E}_C[e(D^\star,-D^\star)]}{\mathbb{E}_C[e(D^\star,-D^\star)]} = O_{\mathbb{P}_C}\left(\sqrt{\frac{1}{\mathbb{E}_C[e(D^\star,-D^\star)]}}\right) = o_{\mathbb{P}_C}(1),$$
$$\frac{e(V)-\mathbb{E}_C[e(V)]}{\mathbb{E}_C[e(V)]} = O_{\mathbb{P}_C}\left(\sqrt{\frac{1}{\mathbb{E}_C[e(V)]}}\right) = o_{\mathbb{P}_C}(1).$$

Moreover, $\rho_C w^2_{\min} \leq 1$ and therefore $\frac{w^2_{\max}}{w^2_{\min}} = o(\frac{n}{r} w^2_{\min}) \leq o(\frac{n}{r}\frac{1}{\rho_C})$ by Assumption 3. Hence $\rho_C \leq o\left(\frac{n}{r}\frac{w^2_{\min}}{w^2_{\max}}\right)$. Therefore

$$1 \leq \frac{\mathbb{E}_C[e(D^\star,-D^\star)]}{\mathbb{E}_0[e(D^\star,-D^\star)]} \leq 1 + \frac{\mathbb{E}_C[e(D^\star, C\setminus D^\star)]}{\mathbb{E}_0[e(D^\star, V\setminus D^\star))]} = 1+\rho_C \frac{\mathbb{E}_0[e(D^\star, C\setminus D^\star)]}{\mathbb{E}_0[e(D^\star, V\setminus D^\star))]}$$
$$\leq 1+\rho_C \frac{|D^\star|(|C|-|D^\star|)}{|D^\star|(|V|-|D^\star|)}\frac{w^2_{\max}}{w^2_{\min}} \leq 1+\rho_C \frac{r}{n}\frac{w^2_{\max}}{w^2_{\min}} \leq 1+o(1).$$

By the above it follows that $\mathbb{E}_C[e(D^\star,-D^\star)] = (1+o(1))\mathbb{E}_0[e(D^\star,-D^\star)]$, and similarly $\mathbb{E}_C[e(V)] = (1+o(1))\mathbb{E}_0[e(V)]$. Therefore

$$e(D^\star,-D^\star) = (1+o_{\mathbb{P}_C}(1))\mathbb{E}_0[e(D^\star,-D^\star)], \quad \text{and} \quad e(V) = (1+o_{\mathbb{P}_C}(1))\mathbb{E}_0[e(V)].$$

Then, applying Lemma 4.2 (using the set $\mathcal{D} = \{D^\star\}$), we obtain

$$\frac{\widehat{e(D^\star)}}{\mathbb{E}_0[e(D^\star)]} = 1+o_{\mathbb{P}_C}(1).$$

Therefore by definition of the thresholded estimator in (4.17),

$$\widehat{e(D^\star)}^\vee = \left(\widehat{e(D^\star)} \bigvee \frac{|D^\star|^2}{n}\log^4\left(\frac{n}{|D^\star|}\right)\right) \tag{4.43}$$
$$= \left((1+o_{\mathbb{P}_C}(1))\mathbb{E}_0[e(D^\star)] \bigvee \frac{|D^\star|^2}{n}\log^4\left(\frac{n}{|D^\star|}\right)\right).$$

We continue by considering the two cases in the maximum of (4.43) separately.

Case 1: Here we have $\widehat{e(D^\star)}^\vee = (1+o_{\mathbb{P}_C}(1))\mathbb{E}_0[e(D^\star)]$. Plugging this into the definition of the test statistic we obtain

$$\begin{aligned}\tau^{\mathrm{u}}_{D^\star} &= \frac{\widehat{e(D^\star)}^\vee h\Big(\Big[\frac{e(D^\star)}{\widehat{e(D^\star)}^\vee}-1\Big]_+\Big)}{|D^\star|\log(n/|D^\star|)} \\ &= \frac{(1+o_{\mathbb{P}_C}(1))\mathbb{E}_0[e(D^\star)]h\Big(\Big[(1+o_{\mathbb{P}_C}(1))\frac{e(D^\star)}{\mathbb{E}_0[e(D^\star)]}-1\Big]_+\Big)}{|D^\star|\log(n/|D^\star|)}.\end{aligned}$$

The proof can then be completed by using the same reasoning as in the proof of Theorem 4.2 from (4.28) to (4.31). Here the additional $o_{\mathbb{P}_C}(1)$ terms do not make any difference.

Case 2: Here we have $\widehat{e(D^\star)}^\vee = (|D^\star|^2/n)\log^4\big(n/|D^\star|\big)$. This corresponds to the case where the underlying graph is very sparse, and therefore a very large signal ρ_C is required to detect a planted community.

We start by deriving a lower bound on ρ_C. Using condition (4.20) and the fact that $\mathbb{E}_0[e(D^\star)] \le (|D^\star|^2/n)\log^4\big(n/|D^\star|\big)$ and $h^{-1}(x) \ge \sqrt{x}$, we obtain

$$\rho_C \ge h^{-1}\bigg(\frac{|D^\star|\log(n/|D^\star|)}{\mathbb{E}_0[e(D^\star)]}\bigg) \ge h^{-1}\bigg(\frac{n}{|D^\star|}\frac{1}{\log^3(n/|D^\star|)}\bigg) \ge (1+o(1))\sqrt{n/|D^\star|}. \quad (4.44)$$

Moreover, by the second part of Assumption 3 we have $1 \le \big(\frac{w_{\max}}{w_{\min}}\big)^2 \le \frac{n}{r}w_{\min}^2$, and therefore $w_{\min} \ge \sqrt{r/n} \ge 1/\sqrt{n}$. Using this together with (4.44) gives

$$\frac{\mathbb{E}_C[e(D^\star)]}{\widehat{e(D^\star)}^\vee} \ge \frac{\rho_C|D^\star|^2 w_{\min}^2}{\frac{|D^\star|^2}{n}\log^4\Big(\frac{n}{|D^\star|}\Big)} \ge \frac{\rho_C}{\log^4\Big(\frac{n}{|D^\star|}\Big)} \to \infty.$$

Then, using that $e(D^\star) = (1+o_{\mathbb{P}_C}(1))\mathbb{E}_C[e(D^\star)]$ by Chebyshev's inequality and $h(x-1) \asymp x\log(x)$ as $x\to\infty$, we obtain

$$\begin{aligned}\widehat{e(D^\star)}^\vee h\bigg(\frac{e(D^\star)}{\widehat{e(D^\star)}^\vee}-1\bigg) &\ge \widehat{e(D^\star)}^\vee h\bigg((1+o_{\mathbb{P}_C}(1))\frac{\mathbb{E}_C[e(D^\star)]}{\widehat{e(D^\star)}^\vee}-1\bigg) \quad (4.45)\\ &= (1+o_{\mathbb{P}_C}(1))\mathbb{E}_C[e(D^\star)]\log\bigg(\frac{\mathbb{E}_C[e(D^\star)]}{\widehat{e(D^\star)}^\vee}\bigg) \\ &= (1+o_{\mathbb{P}_C}(1))\mathbb{E}_C[e(D^\star)]\log\bigg(\rho_C\frac{\mathbb{E}_0[e(D^\star)]}{\widehat{e(D^\star)}^\vee}\bigg).\end{aligned}$$

Now, by the same argument as above we have $w_{\min} \ge 1/\sqrt{n}$. Hence, it follows that $\mathbb{E}_0[e(D^\star)] \ge |D^\star|^2 w_{\min}^2 \ge |D^\star|^2/n$, so that $\mathbb{E}_0[e(D^\star)]/\widehat{e(D^\star)}^\vee \ge \log^{-4}\big(n/|D^\star|\big)$. Then,

using (4.44),

$$\frac{\log(\rho_C\,\mathbb{E}_0[e(D^\star)]/\widehat{e(D^\star)}^\vee)}{\log(\rho_C)} \geq \frac{\log(\rho_C/\log^4(n/|D^\star|))}{\log(\rho_C)} = 1 - 4\frac{\log\log(n/|D^\star|)}{\log(\rho_C)} = 1 + o(1).$$

Plugging this into (4.45), we obtain

$$\begin{aligned}\widehat{e(D^\star)}^\vee h\left(\frac{e(D^\star)}{\widehat{e(D^\star)}^\vee} - 1\right) &= (1 + o_{\mathbb{P}_C}(1))\mathbb{E}_C[e(D^\star)]\log\left(\rho_C\,\frac{\mathbb{E}_0[e(D^\star)]}{\widehat{e(D^\star)}^\vee}\right)\\ &\geq (1 + o_{\mathbb{P}_C}(1))\mathbb{E}_C[e(D^\star)]\log(\rho_C)\\ &= (1 + o_{\mathbb{P}_C}(1))\mathbb{E}_0[e(D^\star)]h(\rho_C - 1)\\ &\geq (1 + o_{\mathbb{P}_C}(1))(1+\varepsilon)|D^\star|\log\left(\frac{n}{|D^\star|}\right),\end{aligned}$$

where the final step follows from (4.20). Therefore, $\tau^{\mathrm{u}}_{D^\star} \geq 1 + \varepsilon/3$ with high probability, completing the proof. □

4.5.3 Proof of Corollaries 4.1 and 4.3

To prove Corollaries 4.1 and 4.3 we need to show that either Assumption 1.1 or 1.2 is sufficient to ensure that $\mathbb{E}_C[e(D^\star)] \to \infty$ for every $C \subseteq V$ of size $|C| = r$. When $\rho_C = O(1)$ this is a direct consequence of conditions (4.12) and (4.21), therefore we will consider the case where $\rho_C \to \infty$.

Using $h(x-1) \asymp x\log(x)$ as $x \to \infty$ together with (4.12) or (4.21), we obtain

$$\mathbb{E}_C[e(D^\star)] = (1 + o(1))\,\frac{\mathbb{E}_0[e(D^\star)]h(\rho_C - 1)}{\log(\rho_C)} \geq (1 + o(1))\,\frac{|D^\star|\log(n/|D^\star|)}{\log(\rho_C)}. \quad (4.46)$$

Below we consider the two cases where Assumption 1.1 or Assumption 1.2 hold separately.

Case 1 (Assumption 1.1 holds): First, by Definition 4.1 and Assumption 1.1, it follows that for every $C \subseteq V$,

$$\frac{\mathbb{E}_0[e(D^\star)]}{|D^\star|\log(n/|D^\star|)} \geq \frac{\mathbb{E}_0[e(C)]}{|C|\log(n/|C|)} = (1 + o(1))\,\frac{r\,\overline{p}_C}{2\log(n/r)} \to \infty,$$

where the final step is a consequence of Assumption 1.1 (iii). In particular, this means that we must have that $|D^\star| \to \infty$.

Then, for every $C \subseteq V$ we have $\rho_C \overline{p}_C \leq 1$, and therefore $\rho_C \leq 1/\overline{p}_C \leq r \leq \sqrt{n}$ for n large enough by Assumption 1.1 (i) and (iii). Therefore, using (4.46) and because $|D^\star| \to \infty$, it follows that

$$\mathbb{E}_C[e(D^\star)] \geq (1+o(1)) \frac{|D^\star| \log(n/|D^\star|)}{\log(\rho_C)} \geq (1+o(1)) |D^\star| \frac{\log(\sqrt{n})}{\log(\sqrt{n})} \to \infty.$$

Case 2 (Assumption 1.2 holds): For every $C \subseteq V$ we have $\rho_C \overline{p}_C \leq 1$, and therefore $\log(\rho_C) \leq \log(1/\overline{p}_C) = o(\log(n))$ by Assumption 1.2 (ii). Hence, using (4.46), we obtain

$$\mathbb{E}_C[e(D^\star)] \geq (1+o(1)) \frac{|D^\star| \log(n/|D^\star|)}{\log(\rho_C)} \geq (1+o(1)) \frac{|D^\star| \log(n)}{o(\log(n))} \to \infty.$$

The above two cases show that either Assumption 1.1 or 1.2 is sufficient to ensure that $\mathbb{E}_C[e(D^\star)] \to \infty$ for every $C \subseteq V$ of size $|C| = r$. □

4.5.4 Proof of Corollaries 4.2 and 4.4

Begin by noting that the conditions in Corollary 4.4 imply the conditions on $p_{\max}$ and $p_{\min}$ that are stated in Corollary 4.2. To prove Corollaries 4.2 and 4.4 we need to show that $\mathbb{E}_C[e(D^\star)] \to \infty$ for every $C \subseteq V$ of size $|C| = r$. Because $p_{\max}/p_{\min} = o(n^{a-b})$, there exists a sequence $x_n \to \infty$ such that $p_{\max}/p_{\min} = n^{a-b}/x_n$. We will first show that $|D^\star| \geq n^b \sqrt{x_n}$, which we will do by a similar argument as in the proof of Lemma 4.1. Suppose $|D^\star| \leq n^b \sqrt{x_n}$, then because $r \geq n^a$,

$$\begin{aligned}
\frac{\mathbb{E}_0[e(D^\star)]}{|D^\star| \log(n/|D^\star|)} &\leq \frac{|D^\star|-1}{2} \frac{p_{\max}}{\log(n/|D^\star|)} \\
&= \frac{|D^\star|-1}{2} \frac{n^{a-b}}{x_n} \frac{p_{\min}}{\log(n/|D^\star|)} \\
&\leq O(1) \frac{n^a}{\sqrt{x_n}} \frac{p_{\min}}{\log(n/r)} \\
&< \frac{r-1}{2} \frac{p_{\min}}{\log(n/r)} \leq \frac{\mathbb{E}_0[e(C)]}{|C| \log(n/|C|)}.
\end{aligned}$$

Hence, $D^\star$ cannot be the maximizer in (4.10) when $|D^\star| \leq n^b \sqrt{x_n}$, and therefore we must have $|D^\star| \geq n^b \sqrt{x_n}$. Therefore,

$$\mathbb{E}_C[e(D^\star)] \geq \mathbb{E}_0[e(D^\star)] \geq \frac{|D^\star|^2}{2} p_{\min} \geq \frac{n^{2b} x_n}{2} n^{-2b} \to \infty.$$

The proof of Corollary 4.2 is then completed by applying Theorem 4.2, and similarly the proof of Corollary 4.4 is completed by applying Theorem 4.3. □

4.5.5 Proof of Theorem 4.1: Information theoretic lower bound

To prove Theorem 4.1 we need to show that $R_n(T_n) \to 1$, where R_n is the worst-case risk given in (4.1) and $T_n \mapsto \{0,1\}$ is any test deciding between the null and alternative hypothesis. The first step is a reduction from the worst-case risk to the average risk

$$\bar{R}_n(T_n) \coloneqq \mathbb{P}_0(T_n(G)=1) + \binom{n}{r}^{-1} \sum_{C\subseteq V,\,|C|=r} \mathbb{P}_C(T_n(G)=0).$$

Note that the average risk is a lower bound for the worst-case risk, that is $R_n(T_n) \geq \bar{R}_n(T_n)$. This average risk corresponds to a hypothesis test between two simple hypotheses, because the alternative hypothesis is now simple. The means that the likelihood ratio test is optimal (by the Neyman-Pearson lemma). In particular, the test $T^{\mathrm{LR}}(G) = \mathbb{1}_{\{L(G)>1\}}$ minimizes the average risk, where $L(G)$ is the likelihood ratio, given in (4.47) below. To avoid overloading the notation we write simply L to denote $L(G)$. The risk of this test is given by

$$\bar{R}_n(T^{\mathrm{LR}}) = \mathbb{P}_0(L>1) + \mathbb{E}_0[L\mathbb{1}_{\{L\leq 1\}}] = 1 - \frac{1}{2}\,\mathbb{E}_0[|L-1|].$$

Therefore, to prove Theorem 4.1, it suffices to show that $\mathbb{E}_0[|L-1|] \to 0$.

Given a graph g, the likelihood ratio $L(g)$ is given by

$$L(g) \coloneqq \binom{n}{r}^{-1} \sum_{C\subseteq V,\,|C|=r} \frac{\mathbb{P}_C(G=g)}{\mathbb{P}_0(G=g)} = \binom{n}{r}^{-1} \sum_{C\subseteq V,\,|C|=r} L_C(g) = \bar{\mathbb{E}}[L_C(g)], \tag{4.47}$$

where $\bar{\mathbb{E}}[\cdot]$ denotes the expectation with respect to a uniformly chosen set $C \subseteq V$ of size $|C| = r$, and

$$L_C(g) \coloneqq \prod_{i<j\in C} \left(\frac{\rho_C p_{ij}}{p_{ij}}\right)^{A_{ij}} \left(\frac{1-\rho_C p_{ij}}{1-p_{ij}}\right)^{1-A_{ij}}. \tag{4.48}$$

To bound $\mathbb{E}_0[|L-1|]$ one generally resorts to the Cauchy-Schwarz inequality to control instead the second moment of L and obtain $\mathbb{E}_0[|L-1|] \leq \mathbb{E}_0[L^2]-1$. However, in our setting this bound is too crude, and the variance of L will be rather large in comparison to the first moment. To see this note that the second moment can be written as

$$\begin{aligned}\mathbb{E}_0[L^2] &= \bar{\mathbb{E}}^{\otimes 2}\left[\mathbb{E}_0\left[L_{C_1}L_{C_2}\right]\right] \\ &= \bar{\mathbb{E}}^{\otimes 2}\left[\mathbb{E}_0\left[\prod_{i<j\in C_1\cap C_2}\left(\frac{\rho_{C_1}p_{ij}\rho_{C_2}p_{ij}}{p_{ij}^2}\right)^{A_{ij}}\left(\frac{(1-\rho_{C_1}p_{ij})(1-\rho_{C_2}p_{ij})}{(1-p_{ij})^2}\right)^{1-A_{ij}}\right]\right],\end{aligned}$$

where $\bar{\mathbb{E}}^{\otimes 2}[\cdot]$ denotes expectation with respect to two independently and uniformly

chosen sets $C_1, C_2 \subseteq V$ of size $|C_1| = |C_2| = r$. This second moment depends crucially on $e(C_1 \cap C_2)$, the number of edges in the intersection of C_1 and C_2. Although this intersection is empty or very small with high probability it can be large with small probability, resulting in a very large second moment if the number of edges inside it is large as well.

To deal with this issue we use a more refined approach suggested by Ingster [108] and later used by Butucea and Ingster [50] and Arias-Castro and Verzelen [11]. This approach relies on a truncation of the likelihood ratio

$$\tilde{L} \coloneqq \binom{n}{r}^{-1} \sum_{C \subseteq V,\, |C|=r} \mathbb{1}_{\Gamma_C} L_C = \bar{\mathbb{E}}[\mathbb{1}_{\Gamma_C} L_C],$$

where L_C is as given by (4.48) and Γ_C is some truncation event. Using $\tilde{L} \leq L$, the triangle inequality, and the Cauchy-Schwarz inequality, we obtain the upper bound

$$\mathbb{E}_0[|L-1|] \leq \mathbb{E}_0[|\tilde{L}-1|] + \mathbb{E}_0[L-\tilde{L}] \leq \sqrt{\mathbb{E}_0[\tilde{L}^2] - 2\mathbb{E}_0[\tilde{L}] + 1} + 1 - \mathbb{E}_0[\tilde{L}].$$

Therefore, $\bar{R}_n(T^{\mathrm{LR}}) \to 1$ when both $\mathbb{E}_0[\tilde{L}] \to 1$ and $\mathbb{E}_0[\tilde{L}^2] \to 1$. So, the ideal truncation event should lower the variance of $\tilde{L}$ while still ensuring that the first moment of $\tilde{L}$ approaches 1.

Intuitively, we would like to use the truncation event to prevent "bad behavior" at the intersection of two sets C_1 and C_2. However, we can only state the truncation event in terms of one of these sets. This creates a challenge. For a given set $C \subseteq V$, the potentially problematic intersections are sets $D \subseteq C$ for which $\mathbb{E}_0[e(D)]$ is large. We will denote by $\mathcal{E}_C$ (see (4.51) below) this class of "potentially problematic" sets. The idea is then to construct the truncation event so that it removes the set C from consideration if it contains a subset $D \in \mathcal{E}_C$ for which the number of edges $e(D)$ is significantly larger than its expectation $\mathbb{E}_0[e(D)]$.

To formalize this, it is helpful to express the likelihood ratio in a more convenient form. Namely,

$$\begin{aligned} L_C(g) &= \exp\Bigg(\sum_{i<j \in C} A_{ij} \log\left(\frac{\rho_C p_{ij}}{p_{ij}}\right) + (1-A_{ij}) \log\left(\frac{1-\rho_C p_{ij}}{1-p_{ij}}\right)\Bigg) \qquad (4.49) \\ &= \exp\Bigg(\sum_{i<j \in C} A_{ij} \theta_{ij}(\rho_C p_{ij}) - \Lambda_{ij}(\theta_{ij}(\rho_C p_{ij}))\Bigg), \end{aligned}$$

with

$$\theta_{ij}(q) \coloneqq \log\left(\frac{q(1-p_{ij})}{p_{ij}(1-q)}\right), \qquad \text{and} \qquad \Lambda_{ij}(\theta) \coloneqq \log\left(1 - p_{ij} + p_{ij} e^{\theta}\right).$$

Note that $\Lambda_{ij}(\theta)$ is the cumulant generating function of $\mathrm{Bern}(p_{ij})$, with Fenchel-Legendre transform given by

$$H_{p_{ij}}(q) = \sup_{x\geq 0}\{qx - \Lambda_{ij}(x)\} = q\,\theta_{ij}(q) - \Lambda_{ij}(\theta_{ij}(q)), \qquad \text{for } q \in (p_{ij}, 1), \tag{4.50}$$

where $H_p(q) := q\log(\frac{q}{p}) + (1-q)\log(\frac{1-q}{1-p})$ is the Kullback-Leibler divergence between $\mathrm{Bern}(p)$ and $\mathrm{Bern}(q)$.

Now, to construct the truncation event Γ_C, we begin by defining for each set $C \subseteq V$ a class of "potentially problematic" intersection sets as

$$\mathcal{E}_C := \left\{D \subseteq C \,:\, (\rho_C - 1)^2\,\mathbb{E}_0[e(D)] > (1-\varepsilon/2)|D|\left(\log\left(\frac{n|D|}{r^2}\right) - b_n\right)\right\}, \tag{4.51}$$

where $b_n \to \infty$ very slowly. For concreteness we will take $b_n = \log\log(n/r)$. Using this, we define the numbers ζ_D in the lemma below, the proof of this lemma is mainly technical and is therefore deferred to Section 4.5.6:

Lemma 4.3. *Let Assumption 2, and either Assumption 1.1 or 1.2 hold. Then for any $C \subseteq V$ of size $|C| = r$ and $D \in \mathcal{E}_C$ there exists a unique number $\zeta_D \geq 1$, such that for n large enough,*

$$(1+\varepsilon)\mathbb{E}_0[e(D)]h(\zeta_D - 1) = |D|\log\left(\frac{n}{|D|}\right).$$

Moreover, ζ_D satisfies $\theta_{ij}(\zeta_D\, p_{ij}) \leq 2\theta_{ij}(\rho_C\, p_{ij})$ for every $i, j \in D$.

Using the numbers $\zeta_D \geq 1$ and $\mathcal{E}_C$ from (4.51), we finally define the truncation events as

$$\Gamma_C := \left\{\sum_{i<j\in D} A_{ij}\,\theta_{ij}(\rho_C p_{ij}) \leq \sum_{i<j\in D} p_{ij}\zeta_D\,\theta_{ij}(\rho_C p_{ij}), \quad \text{for all } D \in \mathcal{E}_C\right\}. \tag{4.52}$$

Loosely speaking $\theta_{ij}(\rho_C p_{ij}) \approx \log(\rho_C)$, so the above truncation event will remove all sets $C \subseteq V$ for which there exists a subset $D \in \mathcal{E}_C$ with $e(D) > \zeta_D\mathbb{E}_0[e(D)]$. Utilizing this truncation event, we need to show that both $\mathbb{E}_0[\tilde{L}] \to 1$ and $\mathbb{E}_0[\tilde{L}^2] \to 1$.

First truncated moment. Here we show that $\mathbb{E}_0[\tilde{L}] \to 1$. Since we are simply considering a truncation of the likelihood, it follows from Fubini's theorem that

$$\mathbb{E}_0[\tilde{L}] = \bar{\mathbb{E}}[\mathbb{E}_0[\mathbb{1}_{\Gamma_C} L_C]] = \bar{\mathbb{E}}[\mathbb{P}_C(\Gamma_C)]. \tag{4.53}$$

Hence, it suffices to show that $\mathbb{P}_C(\Gamma_C) \to 1$ for most $C \subseteq V$. Below we will show the slightly stronger result that $\min_{C\subseteq V, |C|=r} \mathbb{P}_C(\Gamma_C) \to 1$, which together with (4.53) shows that $\mathbb{E}_0[\tilde{L}] \to 1$.

Begin by noting that

$$\max_{C\subseteq V,|C|=r}\max_{i,j\in C}\left|\frac{\theta_{ij}(\rho_C p_{ij})}{\log(\rho_C)}-1\right|\to 0,\qquad \text{as } n\to\infty. \tag{4.54}$$

To see this, consider

$$\frac{\theta_{ij}(\rho_C p_{ij})}{\log(\rho_C)}-1=\frac{\log\left(\frac{1-p_{ij}}{1-\rho_C p_{ij}}\right)}{\log(\rho_C)}. \tag{4.55}$$

Using Assumption 2 we see that the above converges to 0 uniformly over all $i,j\in C$, if ρ_C is bounded away from 1. Otherwise, when $\rho_C\to 1$, we can simply use Taylor's theorem to obtain $\log(\frac{1-p_{ij}}{1-\rho_C p_{ij}})/\log(\rho_C)\le p_{ij}/(1-p_{ij})+3p_{ij}$ provided p_{ij} is small enough (e.g., $p_{ij}\le 1/3$ suffices). Hence, also in this case it follows from Assumption 2 that (4.55) converges uniformly to 0. Loosely speaking, this means that $\theta_{ij}(\rho_C p_{ij})\asymp\log(\rho_C)$ for all sets $C\subseteq V$ and $i,j\in C$. This, together with a union bound and Bennett's inequality, allows us to control $\mathbb{P}_C(\Gamma_C)$. Indeed,

$$\begin{aligned}
1-\mathbb{P}_C(\Gamma_C)&\le\sum_{D\in\mathcal{E}_C}\mathbb{P}_C\Big(\sum_{i<j\in D}A_{ij}\theta_{ij}(\rho_C p_{ij})>\sum_{i<j\in D}p_{ij}\zeta_D\theta_{ij}(\rho_C p_{ij})\Big)\\
&\le\sum_{D\in\mathcal{E}_C}\mathbb{P}_C\Big(\sum_{i<j\in D}A_{ij}>(1+o(1))\zeta_D\sum_{i<j\in D}p_{ij}\Big)\\
&\le\sum_{D\in\mathcal{E}_C}\exp\Big(-\mathbb{E}_C[e(D)]\,h\Big((1+o(1))\Big(\frac{\zeta_D}{\rho_C}-1\Big)\Big)\Big)\\
&=\sum_{D\in\mathcal{E}_C}\exp\Big(-(1+o(1))\mathbb{E}_C[e(D)]\,h\Big(\frac{\zeta_D}{\rho_C}-1\Big)\Big),
\end{aligned}$$

where the last step uses a property of the h function, which ensures that for $t\ge 1$, $x\ge 0$ we have $\sqrt{t}h(x)\le h(tx)\le t^2h(x)$.

To show that this vanishes we need the following lemma, the proof of which is mainly technical and therefore deferred to Section 4.5.6. We remark that the definition of a_n in this lemma comes from the exponent in (4.56) below.

Lemma 4.4. *Define the sequence a_n as*

$$a_n:=\min_{C\subseteq V,|C|=r}\min_{D\in\mathcal{E}_C}\left((1-\varepsilon)\frac{\mathbb{E}_C[e(D)]}{|D|}\,h\Big(\frac{\zeta_D}{\rho_C}-1\Big)-\log\Big(\frac{r}{|D|}\Big)\right).$$

When (4.5)*, Assumption 2, and either Assumption 1.1 or 1.2 hold, then $a_n\to\infty$.*

Using Lemma 4.4 and grouping the sets $D \in \mathcal{E}_C$ by their size $|D|$, together with the bound on the binomial coefficient $\binom{r}{k} \le \left(\frac{re}{k}\right)^k$ we conclude that, for n large enough,

$$\begin{aligned}
1 - &\min_{C\subseteq V,|C|=r} \mathbb{P}_C(\Gamma_C) \\
&\le \max_{C\subseteq V,|C|=r} \sum_{D\in\mathcal{E}_C} \exp\left(-(1+o(1))\mathbb{E}_C[e(D)]\, h\left(\frac{\zeta_D}{\rho_C}-1\right)\right) \\
&= \max_{C\subseteq V,|C|=r} \sum_{k=1}^{r} \sum_{D\in\mathcal{E}_C,|D|=k} \exp\left(-(1+o(1))\mathbb{E}_C[e(D)]\, h\left(\frac{\zeta_D}{\rho_C}-1\right)\right) \\
&= \max_{C\subseteq V,|C|=r} \sum_{k=1}^{r} \sum_{D\in\mathcal{E}_C,|D|=k} \frac{1}{(re/k)^k} \exp\Bigg(-k\Bigg((1+o(1))\frac{\mathbb{E}_C[e(D)]}{k}\, h\left(\frac{\zeta_D}{\rho_C}-1\right) \\
&\qquad\qquad - \log(re/k)\Bigg)\Bigg) \\
&\le \max_{C\subseteq V,|C|=r} \sum_{k=1}^{r} \binom{r}{k}^{-1} \sum_{D\in\mathcal{E}_C,|D|=k} \exp\Bigg(-k\Bigg((1+o(1))\frac{\mathbb{E}_C[e(D)]}{k}\, h\left(\frac{\zeta_D}{\rho_C}-1\right) \\
&\qquad\qquad - \log(re/k)\Bigg)\Bigg) \\
&\le \sum_{k=1}^{r} \binom{r}{k}^{-1} \sum_{D\subseteq C,|D|=k} \exp\left(-k\,(a_n-1)\right) \qquad (4.56) \\
&= \sum_{k=1}^{r} \exp\left(-k\,(a_n-1)\right) \le \frac{\exp(-(a_n-1))}{1-\exp(-(a_n-1))} \to 0,
\end{aligned}$$

where the final step follows because $a_n \to \infty$ by Lemma 4.4. Hence, from (4.53) we see that $\mathbb{E}_0[\tilde{L}] \to 1$.

Second truncated moment. Here we show that $\mathbb{E}_0[\tilde{L}^2] \to 1$. In other words,

$$\mathbb{E}_0[\tilde{L}^2] = \bar{\mathbb{E}}^{\otimes 2}\left[\mathbb{E}_0\left[\mathbb{1}_{\Gamma_{C_1}}\mathbb{1}_{\Gamma_{C_2}}L_{C_1}L_{C_2}\right]\right] \le 1 + o(1),$$

where we recall that $\bar{\mathbb{E}}^{\otimes 2}[\cdot]$ denotes expectation with respect to two independently and uniformly chosen sets $C_1, C_2 \subseteq V$ of size $|C_1| = |C_2| = r$. Let $D = C_1 \cap C_2$, then using (4.49) this becomes

$$\begin{aligned}
\mathbb{E}_0[\tilde{L}^2] &= \bar{\mathbb{E}}^{\otimes 2}\left[\mathbb{E}_0\left[\mathbb{1}_{\Gamma_{C_1}}\mathbb{1}_{\Gamma_{C_2}}L_{C_1}L_{C_2}\right]\right] \\
&= \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_1}\cap\Gamma_{C_2}} \exp\Bigg(\sum_{i<j\in D} A_{ij}\left(\theta_{ij}(\rho_{C_1}p_{ij}) + \theta_{ij}(\rho_{C_2}p_{ij})\right) \\
&\qquad\qquad - \Lambda_{ij}(\theta_{ij}(\rho_{C_1}p_{ij})) - \Lambda_{ij}(\theta_{ij}(\rho_{C_2}p_{ij}))\Bigg)\Bigg]\Bigg],
\end{aligned}$$

where we note that the sum runs only over $i < j \in D = C_1 \cap C_2$. The remaining terms in the sum above (i.e., the terms $i, j \in C_1 \cup C_2$ with $i \notin D$ or $j \notin D$) can all be factorized because the A_{ij} are independent, and all these terms have a zero contribution because their expectation equals one.

Using the Cauchy-Schwarz inequality inside the expectation $\bar{\mathbb{E}}^{\otimes 2}[\cdot]$, and that the sets C_1 and C_2 are chosen independently, we obtain

$$\begin{aligned}
\mathbb{E}_0[\tilde{L}^2] &= \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_1}} \exp\Bigg(\sum_{i<j\in D} A_{ij}\theta_{ij}(\rho_{C_1}p_{ij}) - \Lambda_{ij}(\theta_{ij}(\rho_{C_1}p_{ij}))\Bigg) \\
&\qquad \times \mathbb{1}_{\Gamma_{C_2}} \exp\Bigg(\sum_{i<j\in D} A_{ij}\theta_{ij}(\rho_{C_2}p_{ij}) - \Lambda_{ij}(\theta_{ij}(\rho_{C_2}p_{ij}))\Bigg)\Bigg]\Bigg] \\
&\le \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_1}} \exp\Bigg(\sum_{i<j\in D} 2A_{ij}\theta_{ij}(\rho_{C_1}p_{ij}) - 2\Lambda_{ij}(\theta_{ij}(\rho_{C_1}p_{ij}))\Bigg)\Bigg]^{1/2} \\
&\qquad \times \mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_2}} \exp\Bigg(\sum_{i<j\in D} 2A_{ij}\theta_{ij}(\rho_{C_2}p_{ij}) - 2\Lambda_{ij}(\theta_{ij}(\rho_{C_2}p_{ij}))\Bigg)\Bigg]^{1/2}\Bigg] \\
&= \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_1}} \exp\Bigg(\sum_{i<j\in D} 2A_{ij}\theta_{ij}(\rho_{C_1}p_{ij}) - 2\Lambda_{ij}(\theta_{ij}(\rho_{C_1}p_{ij}))\Bigg)\Bigg]\Bigg].
\end{aligned}$$

Next, we split this expectation into two parts based on whether $D \notin \mathcal{E}_{C_1}$ or $D \in \mathcal{E}_{C_1}$. Thus we have the partition

$$\mathbb{E}_0[\tilde{L}^2] \le \mathsf{P}_1 + \mathsf{P}_2,$$

where

$$\begin{aligned}
\mathsf{P}_1 &:= \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\notin\mathcal{E}_{C_1}\}}\mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_1}} \exp\Bigg(\sum_{i<j\in D} 2A_{ij}\theta_{ij}(\rho_{C_1}p_{ij}) - 2\Lambda_{ij}(\theta_{ij}(\rho_{C_1}p_{ij}))\Bigg)\Bigg]\Bigg], \\
\mathsf{P}_2 &:= \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\in\mathcal{E}_{C_1}\}}\mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_1}} \exp\Bigg(\sum_{i<j\in D} 2A_{ij}\theta_{ij}(\rho_{C_1}p_{ij}) - 2\Lambda_{ij}(\theta_{ij}(\rho_{C_1}p_{ij}))\Bigg)\Bigg]\Bigg].
\end{aligned}$$

Using this split, we first show that $\mathsf{P}_1 \le 1 + o(1)$ and then show that $\mathsf{P}_2 \le o(1)$.

Part 1: Here we show that $\mathsf{P}_1 \le 1 + o(1)$. In this part we can simply ignore the truncation events Γ_{C_1} and obtain the bound

$$\begin{aligned}
\mathsf{P}_1 &\le \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\notin\mathcal{E}_{C_1}\}}\mathbb{E}_0\Bigg[\exp\Bigg(\sum_{i<j\in D} 2A_{ij}\theta_{ij}(\rho_{C_1}p_{ij}) - 2\Lambda_{ij}(\theta_{ij}(\rho_{C_1}p_{ij}))\Bigg)\Bigg]\Bigg] \\
&\le \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\notin\mathcal{E}_{C_1}\}}\exp\Bigg(\sum_{i<j\in D}\Delta_{ij}^{(1)}\Bigg)\Bigg],
\end{aligned}$$

where

$$\Delta_{ij}^{(1)} := \log\left(1 + \frac{(\rho_{C_1} p_{ij} - p_{ij})^2}{p_{ij}(1-p_{ij})}\right).$$

Then using $\log(1+x) \le x$ and by Assumption 2, uniformly over all $i, j \in D$,

$$\Delta_{ij}^{(1)} \le \log\big(1 + (1+o(1))(\rho_{C_1}-1)^2 p_{ij}\big) \le (1+o(1))(\rho_{C_1}-1)^2 p_{ij}.$$

Now, by definition of $\mathcal{E}_{C_1}$ it follows that $(1+o(1))(\rho_{C_1}-1)^2\,\mathbb{E}_0[e(D)] \le |D|\big(\log(\frac{n|D|}{r^2}) - b_n\big)$ for every $D \notin \mathcal{E}_{C_1}$. Therefore

$$\begin{aligned}
\mathsf{P}_1 &\le \bar{\mathbb{E}}^{\otimes 2}\left[\mathbb{1}_{\{D\notin\mathcal{E}_{C_1}\}} \exp\Big((1+o(1))(\rho_{C_1}-1)^2 \mathbb{E}_0[e(D)]\Big)\right] \\
&\le \bar{\mathbb{E}}^{\otimes 2}\left[\mathbb{1}_{\{|D|\le 1\}} + \mathbb{1}_{\{|D|>1\}} \exp\left(|D|\left(\log\left(\frac{n|D|}{r^2}\right) - b_n\right)\right)\right] \\
&\le \bar{\mathbb{P}}^{\otimes 2}(|D|\le 1) + \sum_{k=2}^{r} \exp\left(k\left(\log\left(\frac{nk}{r^2}\right) - b_n\right)\right) \bar{\mathbb{P}}^{\otimes 2}(|D| = k) \\
&\le 1 + \sum_{k=2}^{r} \exp\left(k\left(\log\left(\frac{nk}{r^2}\right) - b_n\right)\right) \bar{\mathbb{P}}^{\otimes 2}(|D| = k). \qquad (4.57)
\end{aligned}$$

Note that $|D| = |C_1 \cap C_2|$ has a hypergeometric distribution under $\bar{\mathbb{P}}^{\otimes 2}$, hence

$$\begin{aligned}
\bar{\mathbb{P}}(|D| = k) &= \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \\
&= \left((1+o(1))\,\frac{re}{k}\,\frac{r-k}{n-r}\right)^k \\
&\le \exp\left(-k\left(\log\left(\frac{nk}{r^2}\right) + O(1)\right)\right). \qquad (4.58)
\end{aligned}$$

Plugging this into (4.57), we obtain

$$\begin{aligned}
\mathsf{P}_1 &\le 1 + \sum_{k=2}^{r} \exp\left(k\left(\log\left(\frac{nk}{r^2}\right) - b_n - \log\left(\frac{nk}{r^2}\right) + O(1)\right)\right) \\
&\le 1 + \sum_{k=2}^{r} \exp\Big(k\big(O(1) - b_n\big)\Big) \le 1 + o(1),
\end{aligned}$$

where the final step follows because $b_n = \log\log(n/r) \to \infty$.

Part 2: Here we show that $\mathsf{P}_2 \le o(1)$. First, define

$$\xi := \frac{1}{2}\,\frac{\log(\zeta_D)}{\log(\rho_{C_1})}, \qquad (4.59)$$

where ζ_D was defined in Lemma 4.3. Then, by the same reasoning as in (4.54),

$$\max_{C_1\subseteq V,|C_1|=r}\ \max_{D\in\mathcal{E}_{C_1}}\ \max_{i,j\in D}\left|\frac{\log(\zeta_D)/\log(\rho_{C_1})}{\theta_{ij}(\zeta_D p_{ij})/\theta_{ij}(\rho_{C_1}p_{ij})}-1\right|\to 0,\qquad \text{as } n\to\infty. \tag{4.60}$$

Loosely speaking, this means that, $\xi \asymp \frac{\theta_{ij}(\zeta_D p_{ij})}{2\theta_{ij}(\rho_{C_1}p_{ij})} \le 1$ uniformly over $i, j \in D$.

By definition of the truncation event Γ_{C_1} in (4.52), for any $D \in \mathcal{E}_{C_1}$,

$$\sum_{i<j\in D} A_{ij}\,\theta_{ij}(\rho_C p_{ij}) \le \sum_{i<j\in D} p_{ij}\zeta_D\,\theta_{ij}(\rho_C p_{ij}).$$

Then for $x \in [0,1]$, we obtain the bound

$$\begin{aligned}
\mathsf{P}_2 &= \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\in\mathcal{E}_{C_1}\}}\,\mathbb{E}_0\Bigg[\mathbb{1}_{\Gamma_{C_1}}\exp\Bigg(\sum_{i<j\in D}2A_{ij}\theta_{ij}(\rho_{C_1}p_{ij})-2\Lambda_{ij}\big(\theta_{ij}(\rho_{C_1}p_{ij})\big)\Bigg)\Bigg]\Bigg]\\
&\le \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\in\mathcal{E}_{C_1}\}}\,\mathbb{E}_0\Bigg[\exp\Bigg(\sum_{i<j\in D}2\theta_{ij}(\rho_{C_1}p_{ij})\Big[xA_{ij}+(1-x)\zeta_D p_{ij}\Big]\\
&\qquad\qquad -2\Lambda_{ij}\big(\theta_{ij}(\rho_{C_1}p_{ij})\big)\Bigg)\Bigg]\\
&= \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\in\mathcal{E}_{C_1}\}}\exp\Bigg(\sum_{i<j\in D}\Lambda_{ij}\big(2\theta_{ij}(\rho_{C_1}p_{ij})x\big)\\
&\qquad\qquad +\big(2\theta_{ij}(\rho_{C_1}p_{ij})-2\theta_{ij}(\rho_{C_1}p_{ij})x\big)\zeta_D p_{ij}-2\Lambda_{ij}\big(\theta_{ij}(\rho_{C_1}p_{ij})\big)\Bigg)\Bigg].
\end{aligned}$$

To obtain the best possible bound we optimize the above with respect to x. Here it can be seen from (4.50) that each individual term in the sum is minimal when $x = \frac{\theta_{ij}(\zeta_D p_{ij})}{2\theta_{ij}(\rho_{C_1}p_{ij})}$. Therefore, by (4.60) it follows that the overall optimum is attained at $x = (1+o(1))\xi$, where ξ was defined in (4.59). Plugging this in, and using (4.60), gives

$$\mathsf{P}_2 \le \bar{\mathbb{E}}^{\otimes 2}\Bigg[\mathbb{1}_{\{D\in\mathcal{E}_{C_1}\}}\exp\Bigg(\sum_{i<j\in D}\Delta_{ij}^{(2)}\Bigg)\Bigg],$$

where

$$\begin{aligned}
\Delta_{ij}^{(2)} &:= \Big(\Lambda_{ij}(\theta_{ij}(\zeta_D p_{ij}))-\zeta_D p_{ij}\theta_{ij}(\zeta_D p_{ij})\Big)-2\Big(\Lambda_{ij}\big(\theta_{ij}(\rho_{C_1}p_{ij})\big)-\zeta_D p_{ij}\theta_{ij}(\rho_{C_1}p_{ij})\Big)\\
&= -H_{p_{ij}}(\zeta_D p_{ij})-2\Big(H_{\rho_{C_1}p_{ij}}(\zeta_D p_{ij})-H_{p_{ij}}(\zeta_D p_{ij})\Big)\\
&= H_{p_{ij}}(\zeta_D p_{ij})-2H_{\rho_{C_1}p_{ij}}(\zeta_D p_{ij}),
\end{aligned} \tag{4.61}$$

where we have used (4.50) in the second equality. To relate the Kullback-Leibler divergence $H_p(q)$, appearing in (4.61), to the function $h(x)$ from (4.2) we need the following lemma, the proof of which is deferred to Section 4.5.6:

Lemma 4.5. *For any $0 < p < q < 1/2$ (possibly depending on n) it follows that,*

$$\left| \frac{H_p(q)}{ph\left(\frac{q}{p} - 1\right)} - 1 \right| \leq O(p+q),$$

where $H_p(q)$ is the Kullback-Leibler divergence between Bern(p) and Bern(q), and $h(x)$ is given in (4.2)*.*

Recall that $\zeta_D \leq \rho_C^2$ by Lemma 4.3, and therefore $\max_{i,j\in D} p_{ij}\zeta_D = o(1)$ by Assumption 2. Similarly, it follows that $\max_{i,j\in D} p_{ij}\rho_C = o(1)$ and $\max_{i,j\in D} p_{ij} = o(1)$. Then, using Lemma 4.5 we obtain the bounds, uniformly over $i, j, \in D$,

$$\left| \frac{H_{p_{ij}}(p_{ij}\zeta_D)}{p_{ij}h\Big(\zeta_D - 1\Big)} - 1 \right| = O\big(p_{ij}(\zeta_D + 1)\big) \leq \max_{i,j\in D} O\big(p_{ij}(\zeta_D + 1)\big) = o(1),$$

$$\left| \frac{H_{p_{ij}\rho_C}(p_{ij}\zeta_D)}{p_{ij}\rho_C h\Big(\frac{\zeta_D}{\rho_C} - 1\Big)} - 1 \right| = O\big(p_{ij}(\zeta_D + \rho_C)\big) \leq \max_{i,j\in D} O\big(p_{ij}(\zeta_D + \rho_C)\big) = o(1).$$

Using the uniform bounds above, we can express $\Delta_{ij}^{(2)}$ from (4.61) in terms on the function $h(x)$. This gives, uniformly over $i, j \in D$,

$$\begin{aligned}\Delta_{ij}^{(2)} &= H_{p_{ij}}(p_{ij}\zeta_D) - 2H_{\rho_{C_1}p_{ij}}(p_{ij}\zeta_D) \\ &= (1 + o(1))\left(p_{ij}h\Big(\zeta_D - 1\Big) - 2\rho_{C_1}p_{ij}h\Big(\frac{\zeta_D}{\rho_{C_1}} - 1\Big)\right).\end{aligned}$$

Therefore, for $D \in \mathcal{E}_{C_1}$, we have

$$\begin{aligned}&\frac{1}{|D|}\sum_{i<j\in D}\Delta_{ij}^{(2)} - \log\Big(\frac{n|D|}{r^2}\Big) \\ &\quad= (1 + o(1))\frac{1}{|D|}\sum_{i<j\in D}\left[p_{ij}h\Big(\zeta_D - 1\Big) - 2\rho_{C_1}p_{ij}h\Big(\frac{\zeta_D}{\rho_{C_1}} - 1\Big)\right] - \log\Big(\frac{n|D|}{r^2}\Big) \\ &\quad= (1 + o(1))\left[\frac{\mathbb{E}_0[e(D)]}{|D|}h\Big(\zeta_D - 1\Big) - 2\frac{\mathbb{E}_{C_1}[e(D)]}{|D|}h\Big(\frac{\zeta_D}{\rho_{C_1}} - 1\Big)\right] \\ &\qquad\qquad - \left(\log\Big(\frac{n}{|D|}\Big) - 2\log\Big(\frac{r}{|D|}\Big)\right).\end{aligned}$$

Then, by definition of ζ_D in Lemma 4.3 and a_n in Lemma 4.4, this becomes

$$\begin{aligned}
&\max_{C_1\subseteq V,|C_1|=r}\max_{D\in\mathcal{E}_{C_1}}\frac{1}{|D|}\sum_{i<j\in D}\Delta_{ij}^{(2)}-\log\Big(\frac{n|D|}{r^2}\Big)\\
&\qquad\le\max_{C_1\subseteq V,|C_1|=r}\max_{D\in\mathcal{E}_{C_1}}2\left(\log\Big(\frac{r}{|D|}\Big)-(1+o(1))\frac{\mathbb{E}_{C_1}[e(D)]h\Big(\frac{\zeta_D}{\rho_{C_1}}-1\Big)}{|D|}\right)\\
&\qquad\le -2a_n\to-\infty.
\end{aligned}$$

Combining the above and grouping the sets $D\in\mathcal{E}_{C_1}$ by their size $|D|$, together with (4.58), we obtain

$$\begin{aligned}
\mathsf{P}_2&\le\bar{\mathbb{E}}^{\otimes 2}\Big[\mathbb{1}_{\{D\in\mathcal{E}_{C_1}\}}\exp\Big(\sum_{i<j\in D}\Delta_{ij}^{(2)}\Big)\Big]\le\sum_{k=1}^{r}\exp\left(k\left(-2a_n+\log\Big(\frac{nk}{r^2}\Big)\right)\right)\bar{\mathbb{P}}(|D|=k)\\
&\le\sum_{k=1}^{r}\exp\left(k\left(-2a_n+\log\Big(\frac{nk}{r^2}\Big)-\log\Big(\frac{nk}{r^2}\Big)+O(1)\right)\right)\\
&\le\sum_{k=1}^{r}\exp\big(k(-2a_n+O(1))\big)\ \to\ 0,
\end{aligned}$$

where the final step follows because $a_n\to\infty$ by Lemma 4.4. This shows that $\mathsf{P}_2=o(1)$.

Following our steps, we conclude that $\mathbb{E}_0[\tilde{L}]\to 1$ and $\mathbb{E}_0[\tilde{L}^2]=\mathsf{P}_1+\mathsf{P}_2\le 1+o(1)$, and therefore $\bar{R}_n(T^{\mathrm{LR}})\to 1$. Finally, the risk of any test T_n is bounded by the average risk of the likelihood ratio test, that is $R_n(T_n)\ge\bar{R}_n(T^{\mathrm{LR}})\to 1$, completing the proof of Theorem4.1. □

4.5.6 Auxiliary results

In this section we provide the proofs for Lemmas 4.3, 4.4, and 4.5. To simplify this, we first compile Assumptions 1.1 and 1.2 into a single result. This is the only place in the proof of Theorem 4.1 where Assumptions 1.1 and 1.2 are used directly. Thus, Theorem 4.1 can simply be extended to other assumptions, provided one can prove Lemma 4.6 below under the new set of assumptions made.

Lemma 4.6. *Let* (4.5)*, Assumption 2, and either Assumption 1.1 or 1.2 hold. Then, for all $C\subseteq V$ of size $|C|=r$ and for all $D\in\mathcal{E}_C$,*

$$\frac{\log(r/|D|)}{\log(n/r)}\big(\log(\rho_C)\vee 1\big)=o(1). \tag{4.62}$$

Furthermore, $\log(n/r)/\log(\rho_C)\to\infty$ *for all $C\subseteq V$ of size $|C|=r$.*

4.5.6.1 Proof of Lemma 4.6

Below we consider two cases depending on whether Assumption 1.1 or Assumption 1.2 holds. We note that some of these inequalities below only hold when n is large enough.

Case 1 (Assumptions 2 and 1.1 hold): For all $C \subseteq V$ of size $|C| = r$, define $\eta_C \geq \rho_C$, such that

$$\frac{|C|\,\overline{p}_C h(\eta_C - 1)}{2\log(n/r)} = 1 - \frac{2}{3}\varepsilon,$$

where ε comes from (4.5). Further, by Assumption 1.1 (iii), we obtain

$$h(\eta_C - 1) \leq \frac{2\log(n/r)}{r\,\overline{p}_C} = o(1). \tag{4.63}$$

Hence, $\eta_C \to 1$ and thus $(\eta_C - 1)^2/h(\eta_C - 1) \to 2$ for every $C \subseteq V$ of size $|C| = r$. Using this together with Assumption 1.1 (ii), we obtain, for all $C \subseteq V$ of size $|C| = r$ and for all $D \subseteq C$ of size $|D| < r/(n/r)^{\gamma_n}$, that

$$\begin{aligned}
(\rho_C - 1)^2 \frac{\mathbb{E}_0[e(D)]}{|D|} &\leq (\eta_C - 1)^2 \frac{|D|\,\overline{p}_D}{2} \leq \delta(\eta_C - 1)^2 \frac{|C|\,\overline{p}_C}{2} \\
&= \big(1 - \frac{2}{3}\varepsilon\big)\log\Big(\frac{n}{r}\Big)\,\delta\,\frac{(\eta_C - 1)^2}{h(\eta_C - 1)} \\
&\leq \big(1 - \frac{2}{3}\varepsilon\big)\log\Big(\frac{n}{r}\Big)\,2\delta\,(1 + o(1)) \\
&\leq (1 - \varepsilon/2)\left(\log\left(\frac{n|D|}{r^2}\right) - b_n\right),
\end{aligned} \tag{4.64}$$

where we recall that $b_n = \log\log(n/r)$. Furthermore, the final inequality above (in (4.64)) follows since

$$\begin{aligned}
2\delta\log(n/r) \leq 2\delta\log(n) &\leq \log(n/r^2) + O(1) \leq \log(n|D|/r^2) + O(1) \\
&\leq (1 + o(1))\,(\log(n|D|/r^2) - b_n),
\end{aligned}$$

because $r = O(n^{1/2-\delta})$ by Assumption 1.1 (i).

Therefore, by definition of $\mathcal{E}_C$ (see (4.51)) it follows that, for all $C \subseteq V$ of size $|C| = r$ and $D \in \mathcal{E}_C$, we have $|D| \geq r/(n/r)^{\gamma_n}$, or equivalently $\log(r/|D|)/\log(n/r) \leq \gamma_n = o(1)$. Furthermore, by (4.63) we have $\rho_C \to 1$ for all $C \subseteq V$ of size $|C| = r$. Combining this, we obtain

$$\frac{\log(r/|D|)}{\log(n/r)}\big(\log(\rho_C) \vee 1\big) \leq \frac{\log(r/|D|)}{\log(n/r)} \leq \gamma_n = o(1).$$

This shows that (4.62) holds.

To complete the proof, we need to show that $\log(n/r)/\log(\rho_C) \to \infty$ for all $C \subseteq V$ of size $|C| = r$. This is trivial because $\rho_C \to 1$, and therefore we have proved Lemma 4.6 when Assumptions 1.1 and 2 hold.

Case 2 (Assumptions 2 and 1.2 hold): For all $C \subseteq V$ of size $|C| = r$ we have $\rho_C \overline{p}_C \leq 1$, and therefore

$$\log(\rho_C) \leq \log(1/\overline{p}_C).$$

Hence, by Assumption 1.2 (i) and (ii), we obtain, for all $C \subseteq V$ of size $|C| = r$,

$$\frac{\log(r/|D|)}{\log(n/r)}\big(\log(\rho_C) \vee 1\big) \leq \frac{\log(r)}{\log(n/r)}\big(\log(1/\overline{p}_C) \vee 1\big) = o(1).$$

This shows that (4.62) holds. Similarly, for all $C \subseteq V$ of size $|C| = r$, we obtain

$$\frac{\log(\rho_C)}{\log(n/r)} \leq \frac{\log(r)}{\log(n/r)} \log(1/\overline{p}_C) = o(1),$$

which shows that $\log(n/r)/\log(\rho_C) \to \infty$.

This proves Lemma 4.6 when Assumptions 1.2 and 2 hold. □

4.5.6.2 Proof of Lemma 4.3

Begin by defining $\tilde{q}_{ij}$ by

$$\frac{\tilde{q}_{ij}\, p_{ij}}{1 - \tilde{q}_{ij}\, p_{ij}} = \frac{\big(\rho_C p_{ij}\big)^2}{p_{ij}} \frac{(1 - p_{ij})}{\big(1 - \rho_C p_{ij}\big)^2},$$

which implies that $\theta_{ij}(\tilde{q}_{ij}\, p_{ij}) = 2\theta_{ij}(\rho_C\, p_{ij})$. By Assumption 2 we have $p_{ij} \to 0$ and $\rho_C^2 p_{ij} \to 0$ for every $i, j \in V$ and therefore it follows that $\tilde{q}_{ij} \asymp \rho_C^2$.

We show below that $h(\tilde{q}_{ij} - 1) \geq (2 + o(1))(\rho_C - 1)^2$ for all $i, j \in D$ when n is large enough. Using this and the fact that $D \in \mathcal{E}_C$ gives

$$\begin{aligned}
(1+\varepsilon)\frac{1}{|D|}\,\mathbb{E}_0[e(D)]h(\tilde{q}_{ij} - 1) &\geq (2 + o(1))(1+\varepsilon)(\rho_C - 1)^2 \frac{\mathbb{E}_0[e(D)]}{|D|} \\
&\geq (2 + o(1))(1+\varepsilon)(1 - \varepsilon/2)\left(\log\left(\frac{n\,|D|}{r^2}\right) - b_n\right) \\
&\geq 2(1+\varepsilon/4)\left(\log\left(\frac{n\,|D|}{r^2}\right) - b_n\right) \\
&\geq 2(1+\varepsilon/4)\log\left(\frac{n}{|D|}\frac{|D|^2}{r^2}\frac{1}{\log(n/r)}\right).
\end{aligned}$$

Then by Lemma 4.6, for every $D \in \mathcal{E}_C$, we have $|D|/r \geq (n/r)^{-o(1)}$, and therefore

$$\begin{aligned}(1+\varepsilon)\frac{1}{|D|}\,\mathbb{E}_0[e(D)]h(\tilde{q}_{ij}-1) &\geq 2(1+\varepsilon/4)\log\!\left(\frac{n}{|D|}\,\frac{|D|^2}{r^2}\,\frac{1}{\log(n/r)}\right)\\ &\geq 2\log\!\left(\frac{n}{|D|}\right)+2\log\!\left(\left(\frac{n}{|D|}\right)^{\varepsilon/4}\left(\frac{|D|^2}{r^2}\,\frac{1}{\log(n/r)}\right)^{1+\varepsilon/4}\right)\\ &\geq 2\log\!\left(\frac{n}{|D|}\right)+\underbrace{2\log\!\left(\left(\frac{n}{r}\right)^{\varepsilon/4-o(1)(1+\varepsilon/4)}\right)}_{\to\infty}\\ &\geq 2\log\!\left(\frac{n}{|D|}\right).\end{aligned}$$

Note that $h(x-1)$ is continuous and increasing on $x \geq 1$. This means that, for large enough n, there is a unique solution $\zeta_D \in (1, \min_{i,j\in D}\tilde{q}_{ij})$ such that

$$(1+\varepsilon)\frac{1}{|D|}\,\mathbb{E}_0[e(D)]h(\zeta_D-1) = \log\!\left(\frac{n}{|D|}\right).$$

Moreover, it follows that $\theta_{ij}(\zeta_D\, p_{ij}) \leq \theta_{ij}(\tilde{q}_{ij}\, p_{ij}) = 2\theta_{ij}(\rho_C\, p_{ij})$ for every $i,j \in D$ because $\zeta_D \in (1, \min_{i,j\in D}\tilde{q}_{ij})$.

We are left to show $h(\tilde{q}_{ij}-1) \geq (2+o(1))(\rho_C-1)^2$, which we do by considering different cases depending on the asymptotic behavior of ρ_C (which is sufficient by Remark 4.1).

Case 1 ($\rho_C \to 1$): By definition of $\tilde{q}_{ij}$,

$$\tilde{q}_{ij}-1 = (\rho_C-1)\left(1+\frac{(1-p_{ij})\rho_C^2}{1-p_{ij}(2\rho_C-\rho_C^2)}\right) \asymp 2(\rho_C-1).$$

Then, using the above together with $h(x-1) \asymp (x-1)^2/2$ as $x \to 1$, we obtain

$$h\left(\tilde{q}_{ij}-1\right) \asymp \left(\tilde{q}_{ij}-1\right)^2/2 \asymp 2(\rho_C-1)^2.$$

Case 2 ($\rho_C \to \alpha \in (1,\infty)$): Using $\tilde{q}_{ij} \asymp \rho_C^2$, we obtain

$$\begin{aligned}\frac{h(\tilde{q}_{ij}-1)}{(\rho_C-1)^2} &\asymp \frac{h(\rho_C^2-1)}{(\rho_C-1)^2} \asymp \frac{\rho_C^2\log(\rho_C^2)-\rho_C^2+1}{(\rho_C-1)^2}\\ &\asymp 1+\frac{2\rho_C\big(\rho_C\log(\rho_C)-\rho_C+1\big)}{(\rho_C-1)^2} \geq 2+o(1).\end{aligned}$$

Case 3 ($\rho_C \to \infty$): Using $\tilde{q}_{ij} \asymp \rho_C^2$ and $h(x-1) \asymp x\log(x)$ as $x \to \infty$, we obtain

$$\frac{h(\tilde{q}_{ij}-1)}{(\rho_C-1)^2} = (1+o(1))\frac{\tilde{q}_{ij}\log(\tilde{q}_{ij})}{\rho_C^2} \geq (2+o(1))\log(\rho_C) \to \infty.$$

In particular, $h(\tilde{q}_{ij} - 1) \geq 2(\rho_C - 1)^2$ when n is large enough. □

4.5.6.3 Proof of Lemma 4.4

First note that (4.5) implies

$$h(\rho_C - 1) \leq (1 - \varepsilon)\frac{|D| \log(n/|D|)}{\mathbb{E}_0[e(D)]},$$

and Lemma 4.3 implies

$$h(\zeta_D - 1) = \frac{1}{1 + \varepsilon} \frac{|D| \log(n/|D|)}{\mathbb{E}_0[e(D)]}.$$

Therefore,

$$\frac{h(\zeta_D - 1)}{h(\rho_C - 1)} \geq \frac{1}{1 - \varepsilon^2}. \tag{4.65}$$

To prove the lemma we consider three different cases depending on the asymptotic behavior of ρ_C (any other case is handled as in Remark 4.1).

Case 1 ($\rho_C \to 1$): From the proof of Lemma 4.3 we have $\zeta_D \in \big(1, \min_{i,j \in D} \tilde{q}_{ij}\big)$, where $\tilde{q}_{ij} \asymp \rho_C^2 \to 1$, and therefore $\zeta_D \to 1$. Then using $h(x - 1) \asymp (x - 1)^2/2$ as $x \to 1$ together with (4.65), we obtain

$$\frac{(\zeta_D - 1)^2}{(\rho_C - 1)^2} \asymp \frac{h(\zeta_D - 1)}{h(\rho_C - 1)} \geq \frac{1}{1 - \varepsilon^2}.$$

Using this, we obtain

$$\begin{aligned}\rho_C h\Big(\frac{\zeta_D}{\rho_C} - 1\Big) \asymp \frac{(\zeta_D - \rho_C)^2}{2\rho_C} &= \frac{1}{2}(\zeta_D - 1)^2 \Big(1 - \frac{\rho_C - 1}{\zeta_D - 1}\Big)^2 \\ &\geq (1 + o(1))\, h(\zeta_D - 1)(1 - \sqrt{1 - \varepsilon^2}) = \Omega(1)\, h(\zeta_D - 1).\end{aligned}$$

This result, together with Lemma 4.3, yields

$$\frac{1}{|D|}\, \mathbb{E}_C[e(D)]\, h\Big(\frac{\zeta_D}{\rho_C} - 1\Big) \geq \Omega(1)\, \frac{1}{|D|}\, \mathbb{E}_0[e(D)]\, h\,(\zeta_D - 1) \,\geq \Omega(1)\, \log\,(n/|D|)\,.$$

Finally, by Lemma 4.6 it follows that $r/|D| \leq (n/r)^{o(1)}$, and therefore

$$\begin{aligned}(1 - \varepsilon)\frac{1}{|D|}\, \mathbb{E}_C[e(D)]\, h\Big(\frac{\zeta_D}{\rho_C} - 1\Big) - \log\Big(\frac{r}{|D|}\Big) \\ \geq \Omega(1) \log\Big(\frac{n}{|D|}\Big) - \log\Big(\frac{r}{|D|}\Big) \geq \big(\Omega(1) - o(1)\big) \log\Big(\frac{n}{r}\Big) \to \infty.\end{aligned}$$

Case 2 ($\rho_C \to \alpha \in (1,\infty)$): By (4.65) it clearly follows that $\rho_C \le \zeta_C$. Also, $h(x-1)$ is convex and has derivative $\log(x)$. It follows that $h(x-1)-h(\rho_C-1) \le (x-\rho_C)\log(x)$ for $x \ge \rho_C$. Using this,

$$\log(\zeta_D)(\zeta_D-\rho_C) \ge h(\rho_C-1)\left(\frac{h(\zeta_D-1)}{h(\rho_C-1)}-1\right) \ge h(\rho_C-1)\left(\frac{1}{1-\varepsilon^2}-1\right).$$

In particular, this result implies that ζ_D is lower bounded away from ρ_C (i.e. $\zeta_D \ge \rho_C + \Omega(1)$). Now, using that $h(x) \ge \frac{x}{2}\log(x+1)$ we obtain

$$\begin{aligned}
\rho_C h\left(\frac{\zeta_D}{\rho_C}-1\right) &\ge \frac{\rho_C}{2}\left(\frac{\zeta_D}{\rho_C}-1\right)\log\left(\frac{\zeta_D}{\rho_C}\right)\\
&\ge \frac{\zeta_D-\rho_C}{2}\left(\log(\zeta_D)-\log(\rho_C)\right)\\
&\ge \Omega(1)\frac{\zeta_D-\rho_C}{2}\log(\zeta_D)\\
&\ge \Omega(1)\,h(\rho_C-1),
\end{aligned}$$

where the last step follows from the fact that ζ_D is lower bounded away from ρ_C. To proceed similarly as in case 1, we need to relate $h(\zeta_D-1)$ to $h(\rho_C-1)$. From the proof of case 2 in Lemma 4.3 it follows that $\zeta_D \le \tilde{q}_{ij} \asymp \rho_C^2$, and since ρ_C is bounded away from 1 it follows that $h(\rho_C-1)/h(\zeta_D-1) \ge \Omega(1)$. Therefore we conclude that

$$\rho_C h\left(\frac{\zeta_D}{\rho_C}-1\right) \ge \Omega(1)\,h(\zeta_D-1).$$

From this point onward the proof continues as in case 1.

Case 3 ($\rho_C \to \infty$): We have $\zeta_D \ge \rho_C \to \infty$ and $h(x-1) \asymp x\log(x)$ as $x \to \infty$. Therefore it follows by (4.65) that

$$\frac{1}{1-\varepsilon^2} \le \frac{h(\zeta_D-1)}{h(\rho_C-1)} \asymp \frac{\zeta_D\log(\zeta_D)}{\rho_C\log(\rho_C)} \asymp \frac{\zeta_D}{\rho_C}\left(1+\frac{\log(\zeta_D/\rho_C)}{\log(\rho_C)}\right).$$

Hence, $\zeta_D/\rho_C \ge 1+\Omega(1)$. Further, using $\zeta_D \le \tilde{q}_{ij} \asymp \rho_C^2$, we obtain

$$\begin{aligned}
\frac{\rho_C h(\zeta_D/\rho_C-1)}{h(\zeta_D-1)} &\asymp \frac{\zeta_D\log(\zeta_D/\rho_C)-\zeta_D+\rho_C}{\zeta_D\log(\zeta_D)}\\
&= \frac{\log(\zeta_D/\rho_C)}{\log(\zeta_D)}+o(1)\\
&\ge \Omega(1)\frac{1}{\log(\zeta_D)}\\
&\ge \Omega(1)\frac{1}{\log(\rho_C)}.
\end{aligned}$$

Here it was crucial to use the fact that ζ_D/ρ_C is lower bounded away from 1. Finally, by Lemma 4.6 we obtain $\log(r/|D|) \le o(\log(n/r)/\log(\rho_C))$, and therefore we get

$$\begin{aligned}(1-\varepsilon)\frac{1}{|D|}\,\mathbb{E}_C[e(D)]\,h\Big(\frac{\zeta_D}{\rho_C}-1\Big)-\log\Big(\frac{r}{|D|}\Big)&\\ \ge \Omega(1)\frac{1}{|D|}\,\mathbb{E}_0[e(D)]\,\frac{h(\zeta_D-1)}{\log(\rho_C)}-\log\Big(\frac{r}{|D|}\Big)&\\ \ge \Omega(1)\frac{\log(n/|D|)}{\log(\rho_C)}-\log\Big(\frac{r}{|D|}\Big)&\\ \ge \big(\Omega(1)-o(1)\big)\frac{\log(n/r)}{\log(\rho_C)}\to\infty,&\end{aligned}$$

where $\log(n/r)/\log(\rho_C)\to\infty$ follows from Lemma 4.6. □

4.5.6.4 Proof of Lemma 4.5

Define the function

$$f_p(q) := H_p(q) - p\,h\Big(\frac{q}{p}-1\Big) = (q-p)+(1-q)\log\Big(\frac{1-q}{1-p}\Big).$$

Then the derivatives of $f_p(q)$ are given by

$$\frac{\partial f_p(q)}{\partial q}=\log\Big(\frac{1-p}{1-q}\Big),\qquad \frac{\partial^2 f_p(q)}{\partial q^2}=\frac{1}{1-q},\qquad \frac{\partial^3 f_p(q)}{\partial q^3}=\frac{1}{(1-q)^2}.$$

Therefore, for $0<p<q$, a Taylor expansion of q around p shows that there exists $\xi\in[p,q]$ such that

$$f_p(q)=\frac{1}{2(1-p)}(q-p)^2+\frac{1}{6(1-\xi)^2}(q-p)^3.$$

Now, we continue by considering two cases depending of the value of q/p.

Case 1 ($q/p\le 5$): Here we use that $h(x-1)\ge(x-1)^2/4$ for all $1<x\le 5$. Therefore,

$$\begin{aligned}\frac{f_p(q)}{ph(\frac{q}{p}-1)}\le\frac{4p\,f_p(q)}{(q-p)^2}&=\frac{4p}{(q-p)^2}\left[\frac{(q-p)^2}{2(1-p)}+\frac{(q-p)^3}{6(1-\xi)^2}\right]\\ &=\frac{2p}{1-p}+\frac{2}{3}\,\frac{q-p}{(1-\xi)^2}\\ &\le O(p)+O(q),\end{aligned}$$

for some $\xi\in[p,q]$.

Case 2 ($q/p > 5$): Here we use that $h(x-1) \geq (x-1)$ for all $x > 5$. Therefore,

$$\frac{f_p(q)}{ph\left(\frac{q}{p}-1\right)} \leq \frac{f_p(q)}{q-p} = \frac{q-p}{2(1-p)} + \frac{(q-p)^2}{6(1-\xi)^2} \leq O(p) + O(q),$$

for some $\xi \in [p,q]$.

To complete the proof, note that $f_p(q) \geq 0$ and $ph\left(\frac{q}{p}-1\right) \geq 0$ for all $0 < p < q < 1$. Therefore, it follows that

$$0 \leq \frac{f_p(q)}{ph\left(\frac{q}{p}-1\right)} \leq O(p+q).$$

□

4.5.6.5 Derivation of equation (4.15)

The choice of estimator in Section 4.2.3 is based on the equality from (4.15). In this section we give a more detailed derivation of this equality. First, observe that $\sum_{i\notin D} w_i \geq \sum_{i\in D} w_i$, which is ensured by Assumption 3. To see this, note that $r\frac{w_{\max}}{w_{\min}} \leq r^{4/3} \wedge \sqrt{nr} \leq n^{4/5}$. Hence, for n large enough,

$$\sum_{i\notin D} w_i - \sum_{i\in D} w_i = w_{\min}\left((n-r) - r\frac{w_{\max}}{w_{\min}}\right) \geq w_{\min}\big((1+o(1))n - n^{4/5}\big) > 0.$$

Then, using that $\sum_{i\notin D} w_i \geq \sum_{i\in D} w_i$, we obtain

$$\begin{aligned} 2\textstyle\sum_{i\in D} w_i &= \sqrt{\left(\sum_{i\in V} w_i\right)^2} - \sqrt{\left(\sum_{i\notin D} w_i - \sum_{i\in D} w_i\right)^2} \\ &= \sqrt{\left(\sum_{i\in V} w_i\right)^2} - \sqrt{\left(\sum_{i\notin D} w_i + \sum_{i\in D} w_i\right)^2 - 4\sum_{i\in D}\sum_{j\notin D} w_i w_j} \\ &= \sqrt{2\mathbb{E}_0[e(V)] + \textstyle\sum_{i\in V} w_i^2} - \sqrt{2\mathbb{E}_0[e(V)] + \textstyle\sum_{i\in V} w_i^2 - 4\mathbb{E}_0[e(D,-D)]}, \end{aligned} \tag{4.66}$$

Finally, plugging (4.66) into the definition of $\mathbb{E}_0[e(D)]$, we obtain

$$\begin{aligned} \mathbb{E}_0[e(D)] &= \frac{1}{2}\Big(\sum_{i\in D} w_i\Big)^2 - \frac{1}{2}\sum_{i\in D} w_i^2 = \frac{1}{8}\Big(2\sum_{i\in D} w_i\Big)^2 - \frac{1}{2}\sum_{i\in D} w_i^2 \\ &= \frac{\left(\sqrt{2\mathbb{E}_0[e(V)] + \sum\limits_{i\in V} w_i^2} - \sqrt{2\mathbb{E}_0[e(V)] + \sum\limits_{i\in V} w_i^2 - 4\mathbb{E}_0[e(D,-D)]}\right)^2}{8} - \frac{1}{2}\sum_{i\in D} w_i^2 \\ &= \frac{\left(\sqrt{\mathbb{E}_0[e(V)] + \frac{1}{2}\sum\limits_{i\in V} w_i^2} - \sqrt{\mathbb{E}_0[e(V)] + \frac{1}{2}\sum\limits_{i\in V} w_i^2 - 2\mathbb{E}_0[e(D,-D)]}\right)^2}{4} - \frac{1}{2}\sum_{i\in D} w_i^2. \end{aligned}$$

Chapter 5

Detecting a botnet in a random geometric graph

Based on:
Detecting a botnet in a network,
G. Bet, K. Bogerd, R. M. Castro, and R. van der Hofstad,
Submitted.

We formalize the problem of detecting the presence of a botnet in a network as a hypothesis testing problem where we observe a single instance of a graph. The null hypothesis, corresponding to the absence of a botnet, is modeled as a random geometric graph where every vertex is assigned a location on a d-dimensional torus and two vertices are connected when their distance is smaller than a certain threshold. The alternative hypothesis is similar, except that there is a small number of vertices, called the botnet, that ignore this geometric structure and simply connect randomly to every other vertex with a prescribed probability.

We present two tests that are able to detect the presence of such a botnet. The first test is based on the idea that botnet vertices tend to form large isolated stars that are not present under the null hypothesis. The second test uses the average graph distance, which becomes significantly shorter under the alternative hypothesis. We show that both these tests are asymptotically optimal. However, numerical simulations show that the isolated star test performs significantly better than the average distance test on networks of moderate size. Finally, we construct a robust scheme based on the isolated star test that is also able to identify the vertices in the botnet.

5.1 Introduction

Complex networks are often described in terms of a large number of vertices that are connected using the same underlying probabilistic mechanism. In practice, however, these networks might contain a small number of vertices that follow different connection criteria. Examples are fake user profiles in a social network (like Facebook or LinkedIn) or servers infected by a computer virus on the internet. We refer to such a set of anomalous vertices as a *botnet*. Typically a botnet represents a potentially malicious anomaly in the network, and thus it is of great practical interest to detect its presence and, when detected, to identify the corresponding vertices. Accordingly, numerous empirical studies have analyzed botnet detection problems and techniques, see [75, 88, 89, 129, 163] and the references therein. In this work we look at the problem from a statistical point of view, and characterize the difficulty of detecting a botnet based only on structural information from the observed network.

More precisely, we formalize this problem as a hypothesis testing problem where we observe a single instance of a random graph. Under the null hypothesis, this graph is a sample from a random geometric graph [91, 151] on n vertices where every vertex is assigned a location on a d-dimensional torus and two vertices are connected when their Euclidean distance on the torus is less than a given radius. Under the alternative hypothesis there is a small number k of vertices, called the botnet, that ignore the geometric structure and instead connect to every other vertex with a prescribed probability. In other words, $n - k$ vertices still connect based on the underlying geometry, while each of the k botnet vertices forms connections uniformly at random with every other vertex (botnet or not). In practice, botnets are built to imitate regular nodes in the network, and so we assume that the expected degree of every vertex is the same under the null and alternative hypothesis. This assumption rules out trivial scenarios where the botnet can be detected simply by looking at the edge density or degree structure.

Our contribution. We propose two different tests to detect whether an observed graph contains a botnet. The first test is a local test, based on the number of isolated stars that can be observed in the given graph. For convenience we refer to this test as the *isolated star test*. For a given vertex, its isolated star is the largest subset of its neighbors such that none of them are connected to each other by an edge. Hence, an isolated star is the largest independent set on the subgraph induced by the neighbors of a vertex. Under the null hypothesis, none of the vertices can become a large isolated star because the underlying geometry ensures that most neighbors are directly connected. However, because the botnet vertices are connected uniformly at random throughout the graph they are likely to become large isolated stars.

Our second test is based on graph distances in the observed graph and thus it has a more global nature. We refer to this test as the *average distance test*. Under the null hypothesis, vertices that are separated by a large Euclidean distance will also

be separated by a large graph distance. However, under the alternative hypothesis, the botnet vertices typically create shortcuts, making many paths much shorter. Under appropriate assumptions, the effect of the shortcuts is large enough to significantly decrease the average graph distance. This phenomenon was first investigated by Watts and Strogatz [162].

Both of our methods can be used to test for the presence of a botnet. Our results show that a botnet can be detected, with high probability, when the expected number of edges connected to all botnet vertices is diverging (i.e., when the expected vertex degree diverges or when the botnet size is unbounded). Remarkably, this means that a single botnet vertex can be detected provided that the graph is not of bounded average degree. We also show that this result is optimal, meaning that it is impossible for any test to detect the presence of a botnet when the expected number of botnet edges is bounded. We complement our theoretical results for the $n \to \infty$ asymptotic regime with numerical simulations that illustrate the performance of our tests on graphs of finite size. These results empirically show that the isolated star test performs much better than the average distance test, with the difference being more pronounced when the dimension of the underlying geometry is large.

Related work. Recently there has been an increasing interest in the development of statistical techniques and algorithms that exploit the structure of large complex-network data to analyze networks more efficiently. In particular, several recent papers have studied hypothesis testing for random graph models. In [11, 12], the authors consider the problem of detecting a denser subset of vertices in an Erdős-Rényi random graph, or in an inhomogeneous random graph [28].

The setting of [47] is perhaps the closest to our setting. The authors consider the problem of deciding whether a given graph is generated by some underlying spatial mechanism. More specifically, in their model, the null hypothesis is an Erdős-Rényi random graph, and this is compared to a high-dimensional random geometric graph under the alternative. As the dimension tends to infinity, the two random graphs become indistinguishable, and they identify how large the dimension can be so that these models can still be distinguished.

The authors of [84] propose a test based on observed frequencies of small subgraphs to distinguish between an Erdős-Rényi random graph, seen as the null hypothesis, and a general class of alternative models that include stochastic block models and the configuration model. Similarly, [40] proposes a test to distinguish between mean-field models and structured Gibbs models. Finally, [102, 134, 147] investigate detection problems in a dynamical setting, where the goal is to detect changes in the graph structure over time.

In this paper we specifically consider the problem of detecting a botnet in undirected graphs. For instance, servers infected by a computer virus on the internet or fake user profiles in social networks like Facebook or LinkedIn. A related and very interesting problem is that of detecting botnets in directed networks, such as Twit-

ter. These are heavily involved in the spread of fake news [22, 57, 129, 154]. In both settings we are trying to identify nodes in the network that are anomalous or disruptive. However, the way these anomalous nodes manifest themselves is rather different than in our model.

5.2 Model formulation and results

In this section we formalize the problem of detecting a botnet in a network as a hypothesis testing problem for graphs. We are given a single observation of a random graph $G = (V, E)$, where $V = \{1, \dots, n\}$ is the vertex set of size $|V| = n$ and $E \subseteq \{(i, j) \in V \times V : i < j\}$ is the random set of edges. We use $i \leftrightarrow j$ to indicate that $i, j \in V$ are connected. That is, we write $i \leftrightarrow j$ when $(i, j) \in E$ and $i \nleftrightarrow j$ otherwise. In particular, G is a simple graph, so it does not contain any self-loops or multiple edges.

Under the null hypothesis, denoted by H_0, the observed graph G is a realization of a d-dimensional random geometric graph $\mathbb{G}(n, d, p)$ on n vertices and with average edge probability p. Formally, let $\mathrm{T}^d := [0, 1]^d$ be the d-dimensional unit torus, with distance function

$$D_{\mathrm{T}}(x, y) = \sqrt{\sum_{j=1}^{d} \min\bigl(|x_j - y_j|, 1 - |x_j - y_j|\bigr)^2}, \qquad \text{for } x, y \in \mathrm{T}^d. \tag{5.1}$$

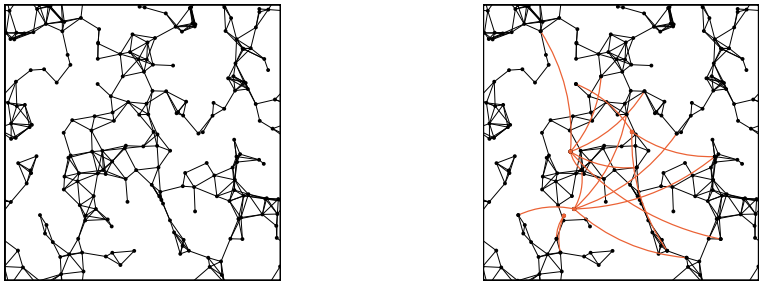
This is simply the Euclidean distance on the unit (hyper-)cube with the ability to "wrap around" the boundaries. We refer to T^d as the embedding space. For each vertex $i \in V$, let X_i be a d-dimensional vector-valued random variable uniformly distributed on T^d. We denote the components of this random vector by $X_i = (X_{i,1}, \dots, X_{i,d})$ and note that these components are independent uniform random variables on the unit interval $[0, 1]$.

For an edge probability p, two vertices $i, j \in V$ are connected when $D_{\mathrm{T}}(X_i, X_j) \le r$, where r is chosen such that the average edge probability is p, that is $\mathbb{P}(D_{\mathrm{T}}(X_i, X_j) \le r) = p$. In other words, r is such that the probability of a random point X_i landing in a ball of radius r is equal to p, which gives the explicit relation $p = (\sqrt{\pi}\, r)^d / \Gamma(d/2 + 1)$, where $\Gamma(\cdot)$ denotes the gamma function. Throughout the rest of this paper we assume that $p \to 0$ as $n \to \infty$, so the average degree is sub-linear in the graph size n. For further details on this model and many of its properties we refer the reader to [151].

The alternative hypothesis, denoted by H_1, is similar except for a small subset of vertices called the botnet. These vertices ignore the geometric structure and simply connect to every other vertex independently with probability p. Formally, the observed graph under the alternative hypothesis is a realization from $\mathbb{G}(n, d, p; k)$, which is a random geometric graph on $n - k$ vertices together with a subset of vertices $B \subseteq V$ of size $|B| = k$, called the botnet. That is, each pair of vertices $i, j \in V \setminus B$ is connected precisely when $D_{\mathrm{T}}(X_i, X_j) \le r$. The remaining vertices in the botnet B

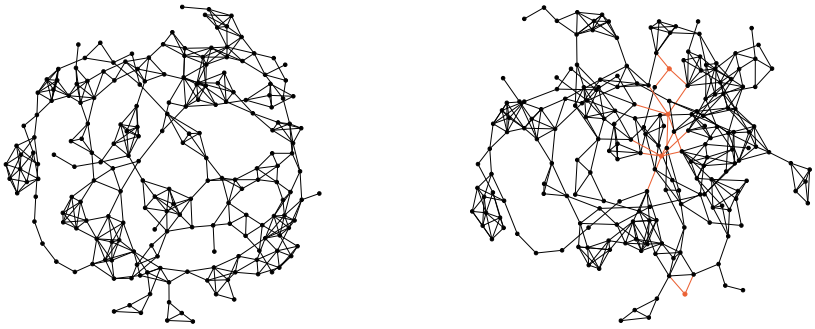
are connected independently and with probability p to every other vertex in V. Note that, by construction, the expected number of edges under the alternative hypothesis is exactly the same as under the null hypothesis.

Another way to sample a graph $\mathbb{G}(n, d, p; k)$ from the alternative hypothesis is to first sample a graph $\mathbb{G}(n, d, p)$ from the null hypothesis. Then randomly select k vertices and delete all edges incident to them, and finally reconnect these vertices to every other vertex independently and with probability p. An example of this is shown in Figures 5.1 and 5.2, where we compare the model under the null and alternative hypothesis in 2 dimensions. However, remember that the vertex locations as shown in Figure 5.1 are not available for the inference problem and we can only observe which vertices are connected. In Figure 5.2 a representation of the graph that does not rely on the Euclidean embedding is given, illustrating how the botnet edges faintly "shorten" the connections between different parts of the network.



(a) Null model $\mathbb{G}(n, d, p)$. (b) Alternative model $\mathbb{G}(n, d, p; k)$.

Figure 5.1: Example of the model under the null and alternative model in 2 dimensions, were we identified opposite sides of the square so that edges can "wrap-around" the sides. Note that this representation uses the embedding of the vertices in the torus that is not available for the inference problem. The graph contains $n = 200$ vertices with $k = 4$ botnet vertices and average degree $np = 5$. The botnet is highlighted in red.



(a) Null model $\mathbb{G}(n, d, p)$. (b) Alternative model $\mathbb{G}(n, d, p; k)$.

Figure 5.2: The same example graphs as in Figure 5.1, but drawn using a force field layout.

General assumptions and notation. Throughout the rest of this paper all unspecified limits are assumed to be taken as the graph size n tends to ∞. We also use standard asymptotic notation: $a_n = O(b_n)$ when a_n/b_n is bounded, $a_n = \Omega(b_n)$ when $b_n = O(a_n)$, $a_n = \Theta(b_n)$ when $a_n = O(b_n)$ and $a_n = \Omega(b_n)$, and $a_n = o(b_n)$ when $a_n/b_n \to 0$. Furthermore, we write $a_n \asymp b_n$ to indicate that $a_n = (1 + o(1))b_n$, and $a_n \ll b_n$ when $a_n = o(b_n)$, or $a_n \gg b_n$ when $b_n = o(a_n)$. Finally, we say that a sequence of events holds with high probability if it holds with probability tending to 1 as $n \to \infty$.

Given two vertices $i, j \in V$, we write $i \leftrightarrow j$ when these vertices are directly connected by an edge, and $i \leftsquigarrow j$ when there exists a path between them. Further, we assume that the dimension $d \geq 2$ remains fixed, but the edge probability p and the botnet size k are allowed to depend on n, although this dependence is left implicit in the notation. We also require that $p \to 0$ in such a way that $np = \Omega(1)$ because otherwise the resulting graphs will be such that most vertices are isolated. Finally, we assume that the botnet size k satisfies $1 \leq k \leq o(n)$.

5.2.1 Detecting a botnet

In this section we obtain a necessary condition for detecting the presence of a planted botnet in the asymptotic regime $n \to \infty$. Given an observed graph, we want to decide whether it was sampled from H_0 or from H_1. To this end, define a test T as a function mapping G to $\{0, 1\}$, where $T(G) = 1$ indicates the null hypothesis is rejected (i.e., the test indicates that the graph contains a botnet), and $T(G) = 0$ otherwise. The *worst-case risk* of such a test is defined as

$$R(T) \coloneqq \mathbb{P}_0(T(G) \neq 0) + \max_{B \subseteq V,\, |B|=k} \mathbb{P}_B(T(G) \neq 1)\,, \tag{5.2}$$

where $\mathbb{P}_0(\cdot)$ denotes the distribution of the random geometric graph under the null hypothesis, and $\mathbb{P}_B(\cdot)$ denotes the distribution of a graph with the botnet $B \subseteq V$ under the alternative hypothesis.

Our goal is to determine when can we distinguish H_0 and H_1 as the graph size n diverges. To this end we consider a sequence of tests $(T_n)_{n=1}^{\infty}$ and we call such a sequence asymptotically powerful when it has vanishing risk, that is $R(T_n) \to 0$ as $n \to \infty$. Hence, a sequence of tests is asymptotically powerful when it identifies the underlying model correctly in the limit $n \to \infty$.

Before we introduce our tests, we define the threshold (in terms of the model parameters) below which it becomes impossible for any test to be asymptotically powerful. We later show that above this threshold the isolated star test is asymptotically powerful. The average distance test is also asymptotically powerful in this regime, assuming some additional technical assumptions are satisfied. This threshold is given in terms of the parameters of the alternative model. Intuitively, it corresponds to the setting where the expected number of edges connected to all botnet ver-

tices is bounded, which happens precisely when both the average degree np and the botnet size k are bounded. In this case, there is a positive probability that all botnet vertices are isolated. When this happens it becomes impossible to reliably distinguish the null and alternative hypothesis. This is formalized in the following theorem, the proof of which is postponed to Section 5.5.4:

Theorem 5.1. *When $npk = O(1)$ no test can be asymptotically powerful (i.e., all tests have risk that is strictly larger than zero).*

In the rest of this section we present the two different tests that can detect the presence of a planted botnet in the regime $npk \to \infty$.

5.2.1.1 Isolated star test

In this section we define a test that can detect whether an observed graph contains a planted botnet based on the presence of isolated stars. For a given vertex $i \in V$, let $N(i) = \{j \in V : (i,j) \in E\}$ denote the subset of its neighbors. The isolated star $S(i) \subseteq N(i)$, at vertex $i \in V$, is the largest independent set on the subgraph of G induced by $N(i)$. In other words, every $j \in S(i)$ is directly connected by an edge to i, and no pair of vertices in $S(i)$ are directly connected (i.e., for every $j,k \in S(i)$ we have $(j,k) \notin E$).

Intuitively, under H_0, the observed graph does not contain large isolated stars because of the underlying geometric structure. In fact, any isolated star under H_0 cannot be larger than the kissing number κ_d, which is the maximum number of non-overlapping spheres of the same radius that can be placed tangent to some central sphere in dimension d. To see this, note that our model is equivalent to the model where every vertex is the center of a sphere of radius $r/2$, and two vertices are connected when their spheres touch or overlap. This means that, under H_0, it is impossible to observe an isolated star that is larger than the kissing number κ_d. For example, the kissing number for dimension $d = 2$ is $\kappa_2 = 6$, so it is impossible to have more than six vertices in a given neighborhood without some of them being connected, see Figure 5.3 for an example.

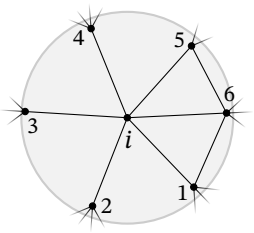


Figure 5.3: Example of an isolated star around the vertex $i \in V$. Although the neighborhood consists of vertices $N(i) = \{1,\dots,6\}$, the largest isolated star is $S(i) = \{1,\dots,5\}$.

However, under the alternative hypothesis the observed graph can, and likely will, contain large isolated stars. In particular, a botnet vertex is quite likely to have an isolated star that is almost as large as its degree. Therefore, it will be likely to observe a few isolated stars that are larger than the kissing number κ_d. Hence, we can scan the graph and compute the size of the isolated star at every vertex. Then we reject H_0 when we see an isolated star that is larger than the kissing number.

Definition 5.1. Let κ_d be the kissing number in dimension d. The *isolated star test* rejects the null hypothesis for a given graph G when $\max_{i \in V} |S(i)| > \kappa_d$.

Checking whether there exists a vertex that has an isolated star that is larger than the kissing number can be done in $O\big(\sum_{i \in V} d_i^{\kappa_d^2}\big)$ time, with d_i the degree of vertex $i \in V$. This scales polynomially in the number of vertices. However, in practice this is not feasible on large graphs, unless all vertices have quite small degree. Instead we can use a greedy algorithm to obtain lower bounds on the size of an isolated star, for example as described in [34]. Moreover, note that the kissing number κ_d depends on the underlying dimension d, and the exact kissing number κ_d is unknown for many dimensions. However, there exist good upper bounds which can be used instead. For dimensions $d \leq 24$, the best known upper bounds can be found in [131], and for larger dimensions one could use the upper bound $\kappa_d \ll 1.3233^d$ [113].

Next we present the main result of this section, where we give conditions for the isolated star test to be asymptotically powerful. The proof of this result is postponed until Section 5.5.1.

Theorem 5.2. *If $npk \to \infty$ then the isolated star test from Definition 5.1 is asymptotically powerful, meaning that it has a risk converging to zero.*

5.2.1.2 Average distance test

In this section we define a test that can detect whether an observed graph contains a planted botnet based on the difference in graph distances under the null and alternative hypothesis. Here we require that p is large enough to ensure that the graph is connected with high probability.

Given two connected vertices $i, j \in V$, let $D_{\mathrm{G}}(i, j)$ be the graph distance between i and j. That is, $D_{\mathrm{G}}(i, j)$ is the length of the shortest path in the graph G that connects i to j. Also, we define the average graph distance as

$$D_{\mathrm{G}}^{\mathrm{avg}}(G) := \frac{\sum_{1 \leq i < j \leq n} \mathbb{1}_{\{i \leftrightsquigarrow j\}} D_{\mathrm{G}}(i, j)}{\sum_{1 \leq i < j \leq n} \mathbb{1}_{\{i \leftrightsquigarrow j\}}}. \tag{5.3}$$

Under the null hypothesis, the observed graph is a random geometric graph and therefore the average graph distance will be large. To see this, consider first the average Euclidean distance between two uniformly chosen points on the torus. This can be lower bounded by

$$\mathbb{E}_0[D_{\mathrm{T}}(X_1,X_2)] = \int_{[0,1]^d} \sqrt{\textstyle\sum_{j=1}^d \min\big(|x_j-1/2|, 1-|x_j-1/2|\big)^2}\ \mathrm{d}x_1\cdots\mathrm{d}x_d \tag{5.4}$$

$$= \int_{[0,1]^d} \sqrt{\textstyle\sum_{j=1}^d |x_j-1/2|^2}\ \mathrm{d}x_1\cdots\mathrm{d}x_d \tag{5.5}$$

$$\geq \int_{[0,1]^d} \max_{1\leq j\leq d} |x_j-1/2|\ \mathrm{d}x_1\cdots\mathrm{d}x_d = \frac{d}{2(d+1)}\,, \tag{5.6}$$

where the final step follows by symmetry and is simply the expectation of the maximum of d independent uniform random variables on $[0,1/2]$. Hence, two uniformly chosen vertices have an expected Euclidean distance of at least $d/(2d+2)$ on the torus. Then, consider the following lower bound on the average graph distance, which holds with high probability

$$D_{\mathrm{G}}^{\mathrm{avg}}(G) \geq \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \frac{D_{\mathrm{T}}(X_i,X_j)}{r}\,, \tag{5.7}$$

because we assumed that the graph is connected with high probability and because every edge can only connect two vertices when they are within distance r, so $D_{\mathrm{G}}(i,j) \geq D_{\mathrm{T}}(X_i,X_j)/r$. Note that, the right-hand side of (5.7) can be seen as a U-statistic. Therefore, using [103, Theorem 5.2], we obtain

$$\mathrm{Var}_0\left(\tbinom{n}{2}^{-1} \textstyle\sum_{1\leq i<j\leq n} D_{\mathrm{T}}(X_i,X_j)\right) \leq \frac{2}{n}\,\mathrm{Var}_0\left(D_{\mathrm{T}}(X_1,X_2)\right) \to 0\,. \tag{5.8}$$

Hence, Chebyshev's inequality ensures that $\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} D_{\mathrm{T}}(X_i,X_j)$ is concentrated around $\mathbb{E}_0[D_{\mathrm{T}}(X_1,X_2)]$ with probability tending to one. Therefore, using (5.6) and (5.7), we obtain for any $\varepsilon > 0$ the following with high probability lower bound

$$D_{\mathrm{G}}^{\mathrm{avg}}(G) \geq (1-\varepsilon)\,\frac{\mathbb{E}_0[D_{\mathrm{T}}(X_1,X_2)]}{r} \geq (1-\varepsilon)\,\frac{d}{2(d+1)}\cdot\frac{1}{r}\,, \tag{5.9}$$

As we show below, the average graph distance is significantly smaller under the alternative hypothesis. Therefore, we consider the following test based on the average graph distance in the observed graph:

Definition 5.2. Fix $\varepsilon > 0$. The *average distance test* rejects the null hypothesis for a given graph G when

$$D_{\mathrm{G}}^{\mathrm{avg}}(G) < (1-\varepsilon)\frac{d}{2(d+1)}\cdot\frac{1}{r}\,. \tag{5.10}$$

This brings us to the main result of this section, which identifies when the average distance test is asymptotically powerful. We postpone the proof of this theorem to Section 5.5.2.

Theorem 5.3. *If $npk \to \infty$ and p is large enough to ensure that the subgraph induced by all non-botnet vertices is connected with high probability, then the average distance test from Definition 5.2 is asymptotically powerful.*

Note that the assumption of connectedness implies that $np \geq \Omega(\log(n))$ [151], and together with the fact that $k \geq 1$ this implies $npk \to \infty$. We include the latter condition to be able to compare the theorem above to Theorem 5.2. The requirement of connectivity is only a technical assumption that we make to considerably simplify the proof. This leads us to conjecture that Theorem 5.3 also holds under the milder condition that p is large enough to ensure the existence of a giant component. This is also supported by our numerical simulations.

5.2.1.3 Unknown dimension and connection radius

Computing the isolated star test requires knowledge of the dimension d of the embedding space, and the average distance test requires the knowledge of the dimension d as well as the connection radius r. In this section we show how to estimate these parameters from the observed graph.

To estimate the dimension d we use the clustering coefficient [59]. This is defined as the probability that two random neighbors of a given vertex are themselves connected. Under the null hypothesis, the clustering coefficient can be computed analytically and the resulting quantity only depends on the dimension d. Using [100, see (15)], for distinct $i, j, k \in V$, we obtain

$$C_d = \mathbb{P}_0(j \leftrightarrow k \,|\, i \leftrightarrow j, i \leftrightarrow k) \tag{5.11}$$

$$= \mathbb{P}\left(\mathrm{Beta}\left(\frac{d+1}{2}, \frac{1}{2}\right) \leq \frac{3}{4}\right) + \mathbb{P}\left(\mathrm{Beta}\left(\frac{d+1}{2}, \frac{d+1}{2}\right) \leq \frac{1}{4}\right), \tag{5.12}$$

where $\mathrm{Beta}(\cdot,\cdot)$ denotes a random variable with a beta distribution. Moreover, for a given graph, the clustering coefficient can be estimated by

$$\widehat{C}_d = \frac{\sum_{1\leq i,j,k\leq n} \mathbb{1}_{\{i\leftrightarrow j, i\leftrightarrow k, j\leftrightarrow k\}}}{\sum_{1\leq i,j,k\leq n} \mathbb{1}_{\{i\leftrightarrow j, i\leftrightarrow k\}}}. \tag{5.13}$$

To estimate the dimension d we can estimate the clustering coefficient $\widehat{C}_d$ using (5.13) and then invert the relation in (5.11) to obtain an estimate for the dimension $\hat{d}$. This method of estimating the dimension gives a consistent estimator, under the null as well as the alternative hypothesis. This is shown in the following lemma, which we prove in Section 5.5.5.

Lemma 5.1. *Using the clustering coefficient to estimate the dimension $\hat{d}$ is consistent under both the null and alternative hypothesis, in the sense that $\hat{d} \xrightarrow{\mathbb{P}_0} d$ and $\hat{d} \xrightarrow{\mathbb{P}_B} d$.*

The average distance test also requires knowledge of the connection radius r. To estimate this, we use that the edge probability p is given by

$$p = \mathbb{P}_0(i \leftrightarrow j) = \frac{(\sqrt{\pi} r)^d}{\Gamma(d/2+1)}, \tag{5.14}$$

where $\Gamma(\cdot)$ denotes the Gamma function. For a given graph, the edge probability can be estimated by

$$\widehat{p} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbb{1}_{\{i \leftrightarrow j\}}. \tag{5.15}$$

To obtain an estimate of the connection radius r, we can estimate the edge probability using (5.15) and then invert the relation in (5.14), using our estimate of d, to obtain an estimate for the connection radius $\hat{r}$. This method gives a consistent estimator of $p = p_n$, as the next lemma shows.

Lemma 5.2. *Using $\widehat{p}$ to estimate $p = p_n$ is consistent both under the null and alternative hypothesis, in the sense that both $\widehat{p}/p \xrightarrow{\mathbb{P}_0} 1$ and $\widehat{p}/p \xrightarrow{\mathbb{P}_B} 1$.*

We postpone the proof of Lemma 5.2 to Section 5.5.6. Since the radius r is given in terms of a continuous function of p in (5.14), this also shows that $\hat{r}/r \xrightarrow{\mathbb{P}_0} 1$ and $\hat{r}/r \xrightarrow{\mathbb{P}_B} 1$ by the continuous mapping theorem. Therefore, our estimate for the connection radius $\hat{r}$ is also consistent under the null and alternative hypotheses.

5.2.2 Identifying the botnet

When a test rejects the null hypothesis, we would also like to identify the vertices that are part of the botnet. To this end, let $\widehat{B} \subseteq V$ be an estimator of the vertices in the botnet. We assume that the size of the botnet $|B| = k$ is known and that $|\widehat{B}| = k$. To measure the performance of our estimator we use the risk function

$$R_{\text{est}}(\widehat{B}) := \mathbb{E}_B\left[\frac{|\widehat{B} \triangle B|}{2|B|}\right], \tag{5.16}$$

where $\widehat{B} \triangle B = ((V \setminus \widehat{B}) \cap B) \cup (\widehat{B} \cap (V \setminus B))$ is the symmetric difference between an estimator $\widehat{B}$ of the botnet and the true botnet B. The reason for the normalization in (5.16) is that $|\widehat{B} \triangle B|$ could be unbounded, while $0 \le |\widehat{B} \triangle B|/|B| \le 2$.

We say that a method achieves *exact recovery* when $R_{\text{est}}(\widehat{B}) \to 0$, and *partial recovery* when $R_{\text{est}}(\widehat{B}) \to \alpha$ for $\alpha \in (0,1)$. In other words, partial recovery corresponds to identifying a positive proportion of the botnet vertices while exact recovery corresponds to identifying the majority of the botnet vertices. Note that, partial recovery is most interesting when the botnet size diverges. To see this, consider partial recovery of a single botnet vertex $k = 1$, in this case $R_{\text{est}}(\widehat{B}) = \mathbb{P}_B(\widehat{B} \neq B)$. That is, the bot-

net vertex is identified correctly only a fraction of the time, and remains unidentified otherwise.

Intuitively, our procedure identifies a botnet vertex when that vertex has a large enough isolated star $|S(i)|$. However, in this case, non-botnet vertices could have an isolated star that is larger than the kissing number κ_d, because it is connected to one or more botnet vertices. Therefore, in order to control the number of false positives, we introduce a parameter $\xi_n > 0$ to artificially increase the threshold κ_d that was used when detecting the presence of a botnet. This leads to the following definition of the isolated star estimator:

Definition 5.3. Let κ_d be the kissing number in dimension d. The *isolated star estimator* is

$$\widehat{B} \coloneqq \{i \in V \,:\, |S(i)| > \kappa_d + \xi_n\}\,, \tag{5.17}$$

with ξ_n given by

$$\xi_n \coloneqq (1+\varepsilon)\frac{\log(n/k)}{\mathcal{W}_0\left(\log(n/k)/(k\mathrm{pe})\right)}\,, \tag{5.18}$$

where $\varepsilon > 0$ is arbitrary, and $\mathcal{W}_0(\cdot)$ denotes the Lambert-W function[1].

Comparing this estimator with the isolated star test from Section 5.2.1.1, we see that the detection threshold is increased by ξ_n. In fact, we have chosen ξ_n to be slightly larger than the maximum number of botnet vertices that are likely to connect to any non-botnet vertex. In other words, the addition of ξ_n ensures that the number of false positives remains vanishingly small.

The performance of our test depends crucially on the asymptotic behavior of the expected number of edges npk that are connected to any botnet vertex. We will concisely refer to these as *botnet edges*. Intuitively, when npk grows slowly, the botnet edges do not influence the largest isolated star of a typical vertex and thus ξ_n is a constant. On the other hand, when npk is large, the largest isolated star of a typical vertex grows with n and consequently ξ_n also increases with n.

More precisely, we show that when $npk \leq n^\beta$ with $\beta \in (0,1)$ our method always achieves at least partial recovery. This corresponds to the most common situation where there is a small botnet in a sparse graph. In this case, ξ_n can be shown to converge to a constant, and thus every vertex with an isolated star that is only slightly larger than the kissing number κ_d is considered a botnet vertex. On the other hand, if npk grows linearly in n or faster, then the typical size of the largest isolated star is significantly larger than the kissing number κ_d and additional technical assumptions are required for our method to achieve at least partial recovery. We make the above considerations precise in the main result of this section, which is presented below.

[1]The function $\mathcal{W}_0(\cdot)$ denotes one of the branches of the Lambert-W function. This is the solution in $y \in [-1,\infty)$ of the equation $x = y\mathrm{e}^y$, with $x \geq -1/\mathrm{e}$. For a detailed overview of this function and its properties see [56].

Theorem 5.4. *Suppose that one of the following conditions holds:*
(i) $npk \leq n^{\beta}$ *for some* $\beta \in (0, 1)$,
(ii) $n^{1-o(1)} \leq npk \leq o(n \log(n/k))$, *and* $\log(n/k)^2/n \leq p \leq \log(n/k)^{-2}$,
(iii) $npk \geq \Omega(n \log(n/k))$, *and* $p = o(k^{-2/3})$.
Then the isolated star estimator from Definition 5.3 has exact recovery if $np \to \infty$, *and partial recovery otherwise.*

Note that, when taken together, conditions (i)–(iii) describe all possible asymptotic behaviors of npk, but additional technical assumptions are required when $npk \geq n^{1-o(1)}$. The proof of Theorem 5.4 is given in Section 5.5.3.

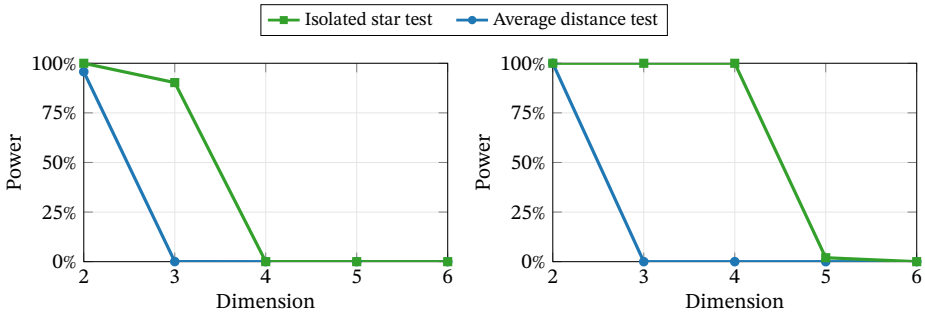
5.3 Simulations

We have shown that the tests introduced in the previous sections are asymptotically powerful when $npk \to \infty$. In this section, we study the finite sample performance of these tests using simulations in order to compare their efficiency in practice on relatively small graphs. As specified both our tests have type-1 error that is nearly zero, so they will almost always correctly identify a graph without a botnet. Therefore, the focus of these simulations is on the type-2 error, which indicates how often a planted botnet is detected when it is actually present.

For our first simulation study we estimate the graph parameters with the consistent estimators described in Section 5.2.1.3 and use these to compute the thresholds for rejecting the null hypothesis as explained in Sections 5.2.1.1 and 5.2.1.2. The results of this can be seen in Figure 5.4. Here we can see that both the isolated star test and average distance test perform quite well, even on relatively small graphs, provided that the underlying dimension is small. Nevertheless, the isolated star test performs better than the average distance test, especially when np is large.

Note that using the estimated model parameters as described in Section 5.2.1.3 instead of the true values could introduce some errors, which in turn could lead to our tests being incorrectly calibrated and result in a type-1 error that is too large. To investigate this issue we repeated the simulation with no botnet (i.e., $k = 0$). Both our tests were always correct and did not reject the null hypothesis in any of the trials. Furthermore, the dimension was correctly estimated in all cases. So it would have made no difference if we used the true dimension d instead of the estimated value $\hat{d}$. We did see some estimation errors for the dimension, but these were only present when the underlying dimension d was larger than 10. Moreover, for the average distance test we also need to estimate the connection radius. The errors introduced by using the estimator $\hat{r}$ compared to the true value r were minimal, and using r instead of $\hat{r}$ yields essentially the same performance as in Figure 5.4.

The results in Figure 5.4 show that both the isolated star test and average distance test can perform well even on relatively small graphs. However, we see that their performance quickly deteriorates as the dimension increases. This happens be-



(a) Average degree $np = 10$.

(b) Average degree $np = 30$.

Figure 5.4: The power of the isolated star test and the average distance test as a function of the dimension d. The threshold for rejecting the null hypothesis is as described in Sections 5.2.1.1 or 5.2.1.2, using estimated model parameters as described in Section 5.2.1.3. The parameters are: graph size $n = 10000$, botnet size $k = 10$, and each simulation contains 5000 samples.

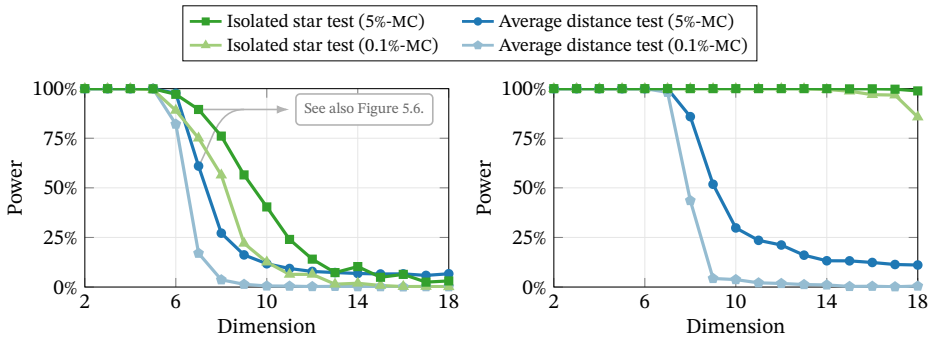
cause the rejection thresholds as described in Sections 5.2.1.1 and 5.2.1.2 are much too conservative.

To better understand the properties of our two test statistics we conduct another simulation study, this time with clairvoyant knowledge of the dimension d and connection radius r, which allows us to correctly calibrate these tests using a simple Monte Carlo method. That is, we sample 5000 graphs from the null model (i.e., $k = 0$) and use these to compute the empirical distributions of either $\max_{i \in V} |S(i)|$ (for the isolated star test) and $D_{\mathrm{G}}^{\mathrm{avg}}(G)$ (for the average distance test). We then take an appropriate quantile of these empirical distributions to obtain the rejection thresholds for a given significance level. The results of this can be seen in Figure 5.5. This shows that the isolated star test outperforms the average distance test in most cases, especially when the dimension d or the average degree np is large.

We note that the Monte Carlo method described above can also be applied when the dimension d or the connection radius r are unknown, but then using the estimated parameter as described in Section 5.2.1.3. However, the problem with this approach is that errors in the parameter estimation could lead to an incorrectly calibrated test, with a type-1 error that is possibly larger than the prescribed α.

In Figure 5.5 we can see that the isolated star test has good performance when the dimension d is small and the average degree np large. The reason for this is that the isolated star test rejects the null hypothesis when the graph contains an isolated star that is larger than a certain rejection threshold (i.e., the kissing number κ_d in Figure 5.4, or the threshold found by Monte Carlo calibration in Figure 5.5). This rejection threshold is lower when the dimension d is small, and the graph is more likely to contain a large isolated star when the average degree np is large. Hence, we see the best performance when the dimension d is small and the average degree np large.

The performance of the average distance test is also related to the dimension d and average degree np of the graph. To understand this, note that the botnet vertices can create shortcuts between vertices that are far away in the embedding space. When the average degree np is large, there is a higher probability that more shortcuts are created, which in turn decreases the average graph distance. On the other hand, as the dimension d increases the average graph distance among the non-botnet vertices decreases, so the shortcuts created by any potential botnet vertices have a less pronounced effect. Thus, here we also see the best performance when the dimension d is small and the average degree np large.



(a) Average degree $np = 10$.

(b) Average degree $np = 30$.

Figure 5.5: The power of the isolated star test and the average distance test. The threshold for rejecting the null hypothesis is obtained by Monte Carlo calibration that ensures respectively $\alpha = 5\%$ and $\alpha = 0.1\%$ type-1 error, assuming that the dimension d and connection radius r are known. The parameters are: graph size $n = 10000$, botnet size $k = 10$, and each simulation contains 5000 samples.

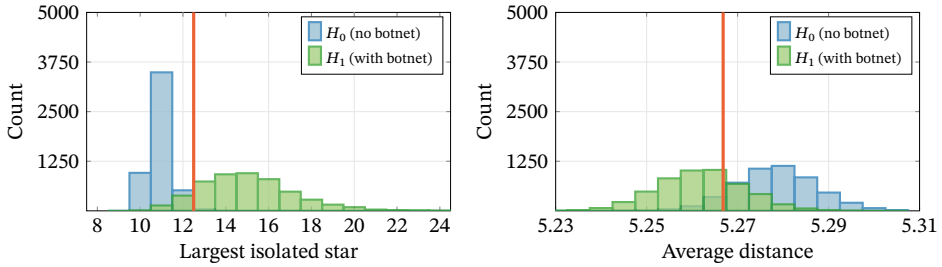


Figure 5.6: Histograms comparing the empirical distributions of the largest isolated star $\max_{i \in V} |S(i)|$ and the average distance $D_G^{\text{avg}}(G)$ statistics under the null and alternative hypothesis. The threshold for rejecting the null hypothesis at the $\alpha = 0.05$ significance level is shown in red. The parameters are: graph size $n = 10000$, botnet size $k = 10$, average degree $np = 10$, dimension $d = 7$, and each histogram contains 5000 samples.

Finally, another reason why both tests have worse performance when the dimension d increases is because the effect of the underlying geometry disappears when $d \to \infty$, as was shown in [47]. Hence the difference between the null and alternative hypothesis is more pronounced when the dimension d is small.

5.4 Discussion

In this section we remark on our results and discuss some possible directions for future work.

Different null hypothesis. Our results show that it is possible to detect an arbitrarily small planted botnet, provided that $npk \to \infty$. However, these results hinge on the underlying geometric structure of the model. Many other network models have been developed that are based on a different geometry than the one assumed by our model [18, 29, 42, 61, 117]. Therefore, it would be interesting to see what the effect of the underlying geometry is, and to what extent our results can be extended to models that have a different underlying geometric structure.

Our tests and analytical approach is fairly robust against minor changes in the underlying geometry. For instance, our results remain true when the embedding space is a slightly deformed torus or sphere, or the points are distributed in the embedding space in a slightly non-uniform way. However, when the changes in geometry are more drastic we expect the nature of the results to change. In particular, when the geometry causes the resulting graph to become a small world we expect the average distance test to fail, and when the geometry causes considerable inhomogeneity in vertex degrees we expect the isolated star test to fail.

Smaller isolated stars for higher power. The isolated star test rejects the null hypothesis when the largest observed isolated star is bigger than the kissing number κ_d, which automatically ensures that the type-1 error is zero. However, for dimensions $d > 2$, the typical largest isolated star in a random geometric graph is much smaller than the kissing number κ_d. For example, numerical simulations suggest that in dimension $d = 4$, the size of the typical isolated star is smaller than 10, whereas the kissing number is $\kappa_4 = 24$ [131, 139]. This suggests that, depending on the significance level, one might use a much smaller threshold value, which would greatly increase the power of the test.

One possible way to achieve this is to calibrate the test using a Monte Carlo approach, as we did in Section 5.3. However, this is a computationally expensive approach which could be avoided with better knowledge of the behavior of isolated star sizes in higher dimensions.

Diverging dimension. From a theoretical perspective it would be interesting to know whether our results can be extended to the setting where the dimension d is diverging together with the graph size n, similar to the problem considered in [47].

For the isolated star test, we can use the following bound on the kissing number $\kappa_d \ll 1.3233^d$ [113]. In this case, the same arguments as in the proof of Theorem 5.2 suggest that the isolated star test is asymptotically powerful when $1 \ll np \ll n^{1/3}$ and

$$d \leq \frac{\log(np)}{\log(1.3233)}\,. \tag{5.19}$$

However, a better understanding of the distribution of isolated stars in graphs with large underlying dimension could significantly improve this result and possibly show that the isolated star test can still be applied even when the dimension grows much faster than (5.19).

Estimating the botnet size. In Section 5.2.2 we show that, under some technical conditions, it is possible to asymptotically identify all botnet vertices provided $np \to \infty$, and that a part of the botnet can be recovered when $np = O(1)$. It could be an interesting possibility for future research to see whether it is possible to estimate the botnet size $|B|$. In the setting where we have exact recovery (i.e., $np \to \infty$) this is of course trivial, but it would be very interesting to see how well that botnet size $|B|$ can be estimated when $np = O(1)$.

5.5 Proofs

This section is devoted to the proofs of the results stated in Sections 5.2.1 and 5.2.2.

5.5.1 Proof of Theorem 5.2: Isolated star test is powerful

As explained in Section 5.2.1.1, the isolated star test has zero type-1 error (i.e., it always correctly identifies a random geometric graph without a botnet). Therefore, to show that the isolated star test is asymptotically powerful, we must show that under the alternative hypothesis, the probability of having an isolated star larger than the kissing number κ_d tends to one. This is done in two steps. First, let $\deg_{V\setminus B}(i)$ be the non-botnet degree of a vertex $i \in V$. That is, $\deg_{V\setminus B}(i)$ denotes the number of non-botnet neighbors of i. Then, we show that any botnet vertex $i \in B$, with $\deg_{V\setminus B}(i) \geq \kappa_d + 1$, will form an isolated star of size $|S(i)| \geq \kappa_d + 1$ with high probability. Second, we show that, with high probability, there exists a botnet vertex that has arbitrarily large non-botnet degree.

Given a botnet vertex $i \in B$, define the event $\mathsf{D}(i) \coloneqq \{\deg_{V\setminus B}(i) \geq \kappa_d + 1\}$. Then, conditionally on the event $\mathsf{D}(i)$, let $\{v_1, \dots, v_{\kappa_d+1}\}$ be a subset of $\kappa_d + 1$ non-botnet neighbors of i. We reveal these vertices one at a time. For every vertex v_j revealed this way, let q_j be the probability that v_j is not connected to any of the previously revealed vertices given that all these previously revealed vertices are themselves not

connected. For $j \in [\kappa_d + 1] = \{1, \dots, \kappa_d + 1\}$ we obtain

$$\begin{aligned} q_j &\coloneqq \mathbb{P}_B\big(v_j \nleftrightarrow v_k \; \forall\, k \in [j-1] \,\big|\, \mathsf{D}(i),\, v_k \nleftrightarrow v_l \; \forall\, k < l \in [j-1]\big) \\ &= \mathbb{P}_B\big(D_{\mathrm{T}}(X_{v_j}, X_{v_k}) > r \; \forall\, k \in [j-1] \,\big|\, \mathsf{D}(i),\, v_k \nleftrightarrow v_l \; \forall\, k < l \in [j-1]\big) \\ &\geq 1 - (j-1)p\,, \end{aligned} \tag{5.20}$$

where we note that, because $i \in B$ is a botnet vertex, conditioning on the event $\mathsf{D}(i)$ does not affect the distribution of the vertex locations (i.e., these remain uniform random variables on the torus). Furthermore, observe that (5.20) becomes an equality precisely when the torus distance between every pair of previously revealed vertices is larger than $2r$. Then, a lower bound on the probability that $i \in B$ forms an isolated star of size at least $\kappa_d + 1$ is given by

$$\mathbb{P}_B\big(|S(i)| \geq \kappa_d + 1 \,\big|\, \mathsf{D}(i)\big) \geq \mathbb{P}_B\big(v_j \nleftrightarrow v_k \; \forall\, j < k \in [\kappa_d + 1] \,\big|\, \mathsf{D}(i)\big) \tag{5.21}$$

$$= \prod_{j=1}^{\kappa_d+1} q_j \geq (1 - \kappa_d p)^{\kappa_d} \to 1\,, \tag{5.22}$$

where the convergence to 1 follows because $p \to 0$ and κ_d is constant. Hence, any botnet vertex $i \in B$ with $\deg_{V\setminus B}(i) \geq \kappa_d + 1$ will form an isolated star of size $|S(i)| \geq \kappa_d + 1$ with probability tending to one.

For the second part of the proof, we will show that there indeed exists a botnet vertex $i \in B$ with $\deg_{V\setminus B}(i) \geq \kappa_d + 1$. First observe that for all $i \in B$ the non-botnet degrees $\deg_{V\setminus B}(i)$ are independent random variables distributed as $\mathrm{Bin}(n-k, p)$. Moreover, by the Stein-Chen method [52, 107], it follows that

$$\Big\| \deg_{V\setminus B}(i) - \mathrm{Poi}((n-k)p) \Big\|_{\mathrm{TV}} \leq 2p \to 0\,, \tag{5.23}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. Now, because $npk \to \infty$ and $k = o(n)$ it follows that either $(n-k)p \to \infty$, or $(n-k)p = \Theta(1)$ and $k \to \infty$. When $(n-k)p \to \infty$ every botnet vertex will eventually have non-botnet degree larger than $\kappa_d + 1$ with high probability. On the other hand, if $(n-k)p = \Theta(1)$ then by (5.23) there is a positive probability that $\deg_{V\setminus B}(i) \geq \kappa_d + 1$, independently for each botnet vertex $i \in B$, and since $k \to \infty$ there exists a botnet vertex with non-botnet degree larger than $\kappa_d + 1$ with high probability. Finally, combining this with (5.22) shows that the graph will contain an isolated star larger than $\kappa_d + 1$ with high probability. □

5.5.2 Proof of Theorem 5.3: Average distance test is powerful

As given in (5.10), under the null hypothesis we have the high probability lower bound

$$D_{\mathrm{G}}^{\mathrm{avg}}(G) \geq (1 - \varepsilon)\, \frac{d}{2(d+1)} \cdot \frac{1}{r}\,, \tag{5.24}$$

Therefore, the average distance test has vanishing type-1 error (i.e., it will correctly identify a geometric random graph with no botnet with high probability). To show that this test is asymptotically powerful, we are left to show that the type-2 error also vanishes. This is done by showing that, under the alternative, there is a botnet vertex that creates a shortcut between most pairs of non-botnet vertices, as shown in Figure 5.7. Using this, we show that, with high probability, the average graph distance is at most $o(1)/r$, which is much smaller than the threshold in (5.10).

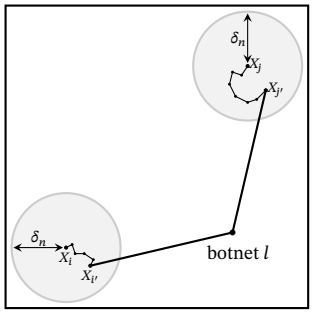


Figure 5.7: Example of botnet vertex $l \in B$ creating a shortcut between vertices $i, j \in V \setminus B$.

For a non-botnet vertex $i \in V \setminus B$, let $B(X_i; \delta_n)$ denote the ball of radius $\delta_n := (V_d \log(np))^{-1/d}$ around the location X_i, where $V_d := \pi^{d/2}/\Gamma(d/2+1)$ denotes the volume of a d-dimensional unit ball. Also, let $A_i \subseteq V \setminus B$ denote the non-botnet vertices with location in $B(X_i; \delta_n)$, that is

$$A_i := \{i' \in V \setminus B : X_{i'} \in B(X_i; \delta_n)\} . \tag{5.25}$$

Note that, because $k = o(n)$, we have

$$\mathbb{E}_B[|A_i|] = \sum_{i' \in V \setminus B} \mathbb{P}_B(i' \in A_i) = (n-k) V_d \delta_n^d = \frac{n-k}{\log(np)} = (1+o(1)) \frac{n}{\log(np)} . \tag{5.26}$$

Therefore, using the relative Chernoff bound [97, see (7)] or [132, Theorem 4.5], for any $\xi > 0$, we obtain

$$\begin{aligned} \mathbb{P}_B\left(|A_i| \geq (1-\xi)\frac{n}{\log(np)}\right) &= 1 - \mathbb{P}_B\left(|A_i| < (1-\xi)\frac{n}{\log(np)}\right) && (5.27)\\ &\geq 1 - \mathbb{P}_B\big(|A_i| < (1-\xi/2)\mathbb{E}[|A_i|]\big) && (5.28)\\ &\geq 1 - \exp\left(-\frac{\xi^2 n}{8\log(np)}\right) \to 1 . && (5.29) \end{aligned}$$

Now, let $l \in B$ be an arbitrary botnet vertex, and consider the probability that there

exists a vertex $i' \in A_i$ that connects to the botnet vertex l. This gives

$$\begin{aligned}
&\mathbb{P}_B\big(\exists i' \in A_i : i' \leftrightarrow l\big) \\
&\quad \geq \mathbb{P}_B\left(\exists i' \in A_i : i' \leftrightarrow l \,\middle|\, |A_i| \geq (1-\xi)\frac{n}{\log(np)}\right) \mathbb{P}_B\left(|A_i| \geq (1-\xi)\frac{n}{\log(np)}\right) \\
&\quad \geq (1+o(1))\Big(1-(1-p)^{(1-\xi)n/\log(np)}\Big) \\
&\quad \geq (1+o(1))\Big(1-\mathrm{e}^{-\xi np/\log(np)}\Big) \to 1\,,
\end{aligned}$$

where the convergence to 1 follows because $np/\log(np) \to \infty$. To continue, we use an existing result relating the torus distance and the graph distance [38, 65, 69, 83, 140]. Translated to our notation, this result is as follows:

Theorem (see [83, Theorem 3] or [69, Theorem 8])**.** *There exists a constant K independent of n such that for any pair of vertices in the same connected component $i, j \in V$ with $D_{\mathrm{T}}(X_i, X_j) \gg \frac{\log(n)}{n\, r^{d-1}}$ we obtain $D_{\mathrm{G}}(i,j) \leq K D_{\mathrm{T}}(X_i, X_j)/r$ with high probability.*

Define the event $\mathsf{C} \coloneqq \{G_{V\setminus B} \text{ is connected}\}$, where $G_{V\setminus B}$ denotes the subgraph induced by all non-botnet vertices. Note that $\mathbb{P}_B(\mathsf{C}) \to 1$ by assumption. Then, given the event C, the result above guarantees that there exists a path of length at most $O(\delta_n)/r$ between i and every $i' \in A_i$. Hence, for a given $i \in V \setminus B$,

$$\begin{aligned}
&\mathbb{P}_B\big(D_{\mathrm{G}}(i,l) \leq 1 + O(\delta_n)/r\big) \\
&\qquad = \mathbb{P}_B\big(\mathsf{C} \cap \big\{D_{\mathrm{G}}(i,l) \leq 1 + O(\delta_n)/r\big\}\big) - o(1) \\
&\qquad \geq \mathbb{P}_B\big(\mathsf{C} \cap \big\{\exists i' \in A_i : i' \leftrightarrow l, D_{\mathrm{G}}(i,i') \leq O(\delta_n)/r\big\}\big) - o(1) \to 1\,. \qquad (5.30)
\end{aligned}$$

Then, by definition of δ_n and applying (5.30) twice, we obtain for an arbitrary pair of non-botnet vertices $i, j \in V \setminus B$ and botnet vertex $l \in B$,

$$\begin{aligned}
&\mathbb{P}_B\big(D_{\mathrm{G}}(i,j) \leq o(1)/r\big) \\
&\qquad \geq \mathbb{P}_B\big(D_{\mathrm{G}}(i,j) \leq 2 + 2\,O(\delta_n)/r\big) \\
&\qquad \geq \mathbb{P}_B\big(D_{\mathrm{G}}(i,l) \leq 1 + O(\delta_n)/r,\, D_{\mathrm{G}}(j,l) \leq 1 + O(\delta_n)/r\big) \to 1\,. \qquad (5.31)
\end{aligned}$$

By observing that every botnet vertex connects to several non-botnet vertices with high probability (as explained at the end of the proof of Theorem 5.2), the above can be strengthened to also include the botnet vertices, and show that the distance between any given pair of vertices is at most $o(1)/r$ with high probability. This brings us to the central result of this proof, namely that for an arbitrary pair $i, j \in V$ it follows that

$$\mathbb{P}_B\big(D_{\mathrm{G}}(i,j) \leq o(1)/r\big) \to 1\,, \qquad (5.32)$$

We continue by showing that the diameter of the graph G is at most $O(1)/r$ with

high probability. To this end, we first consider the diameter of $G_{V\setminus B}$, this gives

$$\begin{aligned}\mathbb{P}_B\Big(\max_{i,j\in V\setminus B} D_{\mathrm{G}}(i,j) \le O(1)/r\Big) &\\ &= \mathbb{P}_B\big(\mathrm{diam}(G_{V\setminus B}) \le O(1)/r\big)\\ &= \mathbb{P}_B\big(\mathsf{C} \cap \big\{\mathrm{diam}(G_{V\setminus B}) \le O(1)/r\big\}\big) - o(1) \to 1\,, \quad (5.33)\end{aligned}$$

where the convergence to 1 follows from the theorem stated above (see also [83, Corollary 6]). Similarly to what we did above, this can be extended to the diameter of G by showing that every botnet vertex connects to at least one non-botnet vertex. Let $l \in B$ denote an arbitrary botnet vertex, then

$$\begin{aligned}\mathbb{P}_B\Big(\min_{i\in B} \deg_{V\setminus B}(i) \ge 1\Big) &= 1 - \big(\mathbb{P}_B(\deg_{V\setminus B}(l) = 0)\big)^k \quad (5.34)\\ &= 1 - \big((1-p)^{n-k}\big)^k \ge 1 - \mathrm{e}^{-(1+o(1))npk} \to 1\,. \quad (5.35)\end{aligned}$$

Hence, using (5.33) and (5.35), we obtain

$$\mathbb{P}_B\Big(\max_{i,j\in V} D_{\mathrm{G}}(i,j) \le O(1)/r\Big) = \mathbb{P}_B\big(\mathrm{diam}(G) \le O(1)/r\big) \to 1\,. \quad (5.36)$$

Finally, it follows from the dominated convergence theorem and (5.32) that $\mathbb{E}_B[\mathbb{1}_{\{\mathrm{diam}(G)\le O(1)/r\}} D_{\mathrm{G}}^{\mathrm{avg}}(G)] = o(1)/r$. Combining this with (5.36) and Markov's inequality we obtain, for any $a > 0$,

$$\begin{aligned}\mathbb{P}_B\left(D_{\mathrm{G}}^{\mathrm{avg}}(G) \ge \frac{a}{r}\right) &= \mathbb{P}_B\left(\mathbb{1}_{\{\mathrm{diam}(G)\le O(1)/r\}} D_{\mathrm{G}}^{\mathrm{avg}}(G) \ge \frac{a}{r}\right) - o(1) \quad (5.37)\\ &\le \frac{r}{a}\,\mathbb{E}_B\big[\mathbb{1}_{\{\mathrm{diam}(G)\le O(1)/r\}} D_{\mathrm{G}}^{\mathrm{avg}}(G)\big] - o(1) \to 0\,. \quad (5.38)\end{aligned}$$

In particular, choosing $a = (1-\varepsilon)\frac{d}{2(d+1)}$ gives $\mathbb{P}_B\big(D_{\mathrm{G}}^{\mathrm{avg}}(G) < (1-\varepsilon)\frac{d}{2(d+1)}\frac{1}{r}\big) \to 1$. This shows that the average distance test is asymptotically powerful. □

5.5.3 Proof of Theorem 5.4: Performance of the isolated star estimator

We need to show that $R_{\mathrm{est}}(\widehat{B}) \to 0$, for the estimator $\widehat{B}$ from Definition 5.3. First, we decompose the risk $R_{\mathrm{est}}(\widehat{B})$ as

$$\begin{aligned}R_{\mathrm{est}}(\widehat{B}) = \mathbb{E}_B\left[\frac{|\widehat{B} \vartriangle B|}{2|B|}\right] &= \frac{\mathbb{E}_B\big[|(V\setminus\widehat{B})\cap B|\big] + \mathbb{E}_B\big[|\widehat{B}\cap(V\setminus B)|\big]}{2|B|} \quad (5.39)\\ &= \frac{1}{2|B|}\sum_{j\in B}\mathbb{P}_B\big(j\notin\widehat{B}\big) + \frac{1}{2|B|}\sum_{j\in V\setminus B}\mathbb{P}_B\big(j\in\widehat{B}\big) \quad (5.40)\\ &= \frac{1}{2|B|}\sum_{j\in B}\mathbb{P}_B\big(|S(j)|\le\kappa_d+\xi_n\big) + \frac{1}{2|B|}\sum_{j\in V\setminus B}\mathbb{P}_B\big(|S(j)|>\kappa_d+\xi_n\big)\,. \quad (5.41)\end{aligned}$$

We start by showing that the second term in (5.39) vanishes. Note that, for any non-botnet vertex $i \in V \setminus B$, the size of its isolated star $|S(i)|$ is bounded by the kissing number κ_d plus the amount of botnet vertices connected to it. Therefore,

$$\begin{aligned}\frac{1}{2|B|}\sum_{j\in V\setminus B}\mathbb{P}_B\big(|S(j)|>\kappa_d+\xi_n\big) &= \frac{n-k}{2k}\,\mathbb{P}_B\big(|S(i)|>\kappa_d+\xi_n\big)\\ &\le \frac{n-k}{2k}\,\mathbb{P}_B\big(i \text{ is connected to at least } \xi_n \text{ botnet vertices}\big)\\ &= \frac{n-k}{2k}\,\mathbb{P}\big(\mathrm{Bin}(k,p)>\xi_n\big) \le \frac{n-k}{2k}\left(\frac{kpe}{\xi_n}\right)^{\xi_n} \to 0\,, \end{aligned} \tag{5.42}$$

where the convergence to 0 follows from the definition of ξ_n in (5.18). In fact, the definition of ξ_n was chosen precisely to ensure this convergence.

To complete the proof, we analyze the first term on the right-hand side of (5.39). Let $i \in B$ be an arbitrary botnet vertex, then

$$\begin{aligned}\frac{1}{|B|}\sum_{j\in B}\mathbb{P}_B\big(|S(j)|\le\kappa_d+\xi_n\big) &= \mathbb{P}_B\big(|S(i)|\le\kappa_d+\xi_n\big)\\ &= 1-\mathbb{P}_B\big(|S(i)|>\kappa_d+\xi_n\big)\\ &= 1-\mathbb{P}_B\big(|S(i)|>\kappa_d+\xi_n \,\big|\, \deg(i)>\kappa_d+\xi_n\big)\,\mathbb{P}_B\big(\deg(i)>\kappa_d+\xi_n\big)\,.\end{aligned}$$

Now, using the same argument as in (5.22), we obtain

$$\begin{aligned}&\mathbb{P}_B\big(|S(i)|>\kappa_d+\xi_n \,\big|\, \deg(i)>\kappa_d+\xi_n\big)\\ &\qquad\ge \prod_{j=1}^{\kappa_d+\xi_n+1}\min\big\{(1-(\kappa_d+\xi_n)p),(1-p)^{\kappa_d+\xi_n}\big\}, \end{aligned} \tag{5.43}$$

which converges to 1 provided that $\xi_n^2 p \to 0$. Combining the above, we obtain

$$R_{\mathrm{est}}(\widehat{B}) = \frac{1}{2|B|}\sum_{j\in B}\mathbb{P}_B\big(|S(j)|\le\kappa_d+\xi_n\big) + \frac{1}{2|B|}\sum_{j\in V\setminus B}\mathbb{P}_B\big(|S(j)|>\kappa_d+\xi_n\big) \tag{5.44}$$

$$= \frac{1}{2}\big(1-\big(1-(\kappa_d+\xi_n)p\big)^{\kappa_d+\xi_n+1}\big)\mathbb{P}_B\big(\deg(i)>\kappa_d+\xi_n\big)\big) + o(1)\,, \tag{5.45}$$

where $i \in B$ is an arbitrary botnet vertex. Therefore, the isolated star estimator has exact recovery when $\xi_n^2 p \to 0$ and $\mathbb{P}_B\big(\deg(i)>\kappa_d+\xi_n\big)\to 1$, and partial recovery when $\xi_n^2 p \to 0$ and $\mathbb{P}_B\big(\deg(i)>\kappa_d+\xi_n\big) = \Omega(1)$. To show this, we consider the three different cases from the theorem statement.

Case (i): From our assumption it follows that $kp \le n^{-\alpha}$ for some $\alpha \in (0,1)$. Recall that $\mathcal{W}_0(x)$ denotes the Lambert-W function, which can be approximated by $\mathcal{W}_0(x) \asymp \log(x)$ when $x \to \infty$ [56]. We obtain $\xi_n \asymp 2\log(n/k)/\log(n^\alpha) = O(1)$. Hence, it

follows that $\xi_n^2 p \to 0$. Moreover,

$$\mathbb{P}_B\big(\deg(i) > \kappa_d + \xi_n\big) = \mathbb{P}\big(\mathrm{Bin}(n-1,p) > O(1)\big) \tag{5.46}$$

$$= \begin{cases} 1 - o(1) & \text{if } np \to \infty\,, \\ \Omega(1) & \text{otherwise}\,. \end{cases} \tag{5.47}$$

Therefore, the isolated star estimator achieves exact recovery when $np \to \infty$, and partial recovery otherwise.

Case (ii): From our assumption it follows that $n^{-o(1)} \leq kp \leq o(\log(n/k))$. Using $\mathcal{W}_0(x) \to \infty$ when $x \to \infty$, we obtain

$$\xi_n \leq \frac{2\log(n/k)}{\mathcal{W}_0\big(\log(n/k)/o(\log(n/k))\big)} = o(\log(n/k))\,. \tag{5.48}$$

Hence, it follows that $\xi_n^2 p \leq o(\log(n/k)^2)\log(n/k)^{-2} \to 0$. Moreover, from the assumptions for this case it follows that $np \gg \log(n/k) \to \infty$, and therefore

$$\mathbb{P}_B\big(\deg(i) > \kappa_d + \xi_n\big) = \mathbb{P}\big(\mathrm{Bin}(n-1,p) > \kappa_d + o(\log(n/k))\big) \tag{5.49}$$

$$\geq \mathbb{P}\big(\mathrm{Bin}(n-1,p) > \log(n/k)\big) \to 1\,. \tag{5.50}$$

Hence, the isolated star estimator has exact recovery.

Case (iii): From our assumption it follows that $kp \geq \Omega(\log(n/k))$. When $kp \gg \log(n/k)$ we use that $\mathcal{W}_0(x) \asymp x$ when $x \to 0$ [56], and obtain $\xi_n = \Theta(kp)$. Otherwise, when $kp = \Theta(\log(n/k))$, it also holds that $\xi_n = \Theta(\log(n/k)) = \Theta(kp)$. In both cases, it follows that $\xi_n^2 p = \Theta(k^2 p^3) \to 0$. Furthermore, note that $np \gg kp \to \infty$ and therefore $\mathbb{P}_B\big(\deg(i) > \kappa_d + \xi_n\big) = \mathbb{P}\big(\mathrm{Bin}(n-1,p) > O(kp)\big) \to 1$, so the isolated star estimator achieves exact recovery. □

5.5.4 Proof of Theorem 5.1: When no test is powerful

We start by considering a simpler version of the problem where the set of potential botnet vertices $B \subseteq V$ is known. Now, we no longer have a composite alternative hypothesis, and this problem corresponds to a hypothesis test between two simple hypotheses. That is, given a set $B \subseteq V$, we consider the risk

$$R^*(T) = \mathbb{P}_0(T(G) \neq 0) + \mathbb{P}_B(T(G) \neq 1)\,. \tag{5.51}$$

Note that, for every test T, the risk $R^*(T)$ is a lower bound for the worst-case risk $R(T)$ in (5.2). Using a result by Tsybakov [157, Proposition 2.1], for every test T it follows that

$$R(T) \geq R^*(T) \geq \sup_{\tau > 0} \left\{ \frac{\tau}{\tau + 1}\, \mathbb{P}_0(L(G) \geq \tau) \right\}, \tag{5.52}$$

where $L(g) = \mathbb{P}_B(G = g)/\mathbb{P}_0(G = g)$ is the likelihood ratio. Therefore, to show that no test is asymptotically powerful it suffices to show that $\mathbb{P}_0(L(G) \geq \tau)$ remains bounded away from zero, for some τ independent of the graph size n. To this end, define the event

$$\mathsf{A} \coloneqq \big\{\text{all vertices in } B \text{ are isolated in the graph } G\big\}. \tag{5.53}$$

For every graph g such that $\mathbb{P}_0(G = g \,|\, \mathsf{A}) > 0$ (i.e., a graph that could be a sample from the null hypothesis with all vertices in B being isolated), it follows that

$$\mathbb{P}_0(G = g) \leq \mathbb{P}_0(G_{V\setminus B} = g_{V\setminus B}) = \mathbb{P}_B(G_{V\setminus B} = g_{V\setminus B}) = \frac{\mathbb{P}_B(G = g)}{(1-p)^{(n-k)k+k(k-1)/2}}, \tag{5.54}$$

where we have used $\{G_{V\setminus B} = g_{V\setminus B}\}$ to indicate the event where the subgraphs induced by the non-botnet vertices $V \setminus B$ are equal. Hence, for all g in which the vertices of B are isolated, we obtain

$$L(g) = \frac{\mathbb{P}_B(G = g)}{\mathbb{P}_0(G = g)} \geq (1-p)^{(n-k)k+k(k-1)/2} = \mathrm{e}^{-(1+o(1))npk}, \tag{5.55}$$

which remains strictly positive as $n \to \infty$ by the assumption that $npk = O(1)$. Therefore, we can choose $\tau > 0$ small enough such that $\mathbb{P}_0(L(G) \geq \tau \,|\, \mathsf{A}) = 1$ for all n large enough. Finally, using the same reasoning as in (5.22), observe that

$$\begin{aligned}
\mathbb{P}_0(L(G) \geq \tau) &\geq \mathbb{P}_0(L(G) \geq \tau \,|\, \mathsf{A})\, \mathbb{P}_0(\mathsf{A}) && (5.56)\\
&= \mathbb{P}_0(\mathsf{A}) && (5.57)\\
&\geq \left(\prod_{i=0}^{k-1}(1-ip)\right)(1-kp)^{n-k} && (5.58)\\
&= \mathrm{e}^{-(1+o(1))npk}. && (5.59)
\end{aligned}$$

which remains strictly positive as $n \to \infty$ by the assumption that $npk = O(1)$. Plugging this into (5.52) shows that, for every test T, the risk $R(T) \geq R^*(T)$ remains bounded away from zero, and therefore that no test can be asymptotically powerful. □

5.5.5 Proof of Lemma 5.1: Consistency of the dimension estimator

We start by showing that $\widehat{C}_d \xrightarrow{\mathbb{P}_0} C_d$, and from this it follows that $\hat{d} \xrightarrow{\mathbb{P}_0} d$ by the continuous mapping theorem and because (5.11) is continuous. Using (5.13) we obtain

$$\widehat{C}_d(G) = \frac{n^{-3}\sum_{1\leq i,j,k\leq n} \mathbb{1}_{\{i\leftrightarrow j, i\leftrightarrow k, j\leftrightarrow k\}}/p^2}{n^{-3}\sum_{1\leq i,j,k\leq n} \mathbb{1}_{\{i\leftrightarrow j, i\leftrightarrow k\}}/p^2}. \tag{5.60}$$

Here we will show that the numerator converges in probability to C_d, and the denominator converges in probability to 1. Since the computations regarding the denominator are largely similar to those of the numerator these will be omitted for brevity, and we will focus on the numerator.

Let $X_{ijk} = \mathbb{1}_{\{i \leftrightarrow j, i \leftrightarrow k, j \leftrightarrow k\}}/p^2$ and $\bar{X} = n^{-3}\sum_{1\leq i,j,k\leq n} X_{ijk}$, then $\bar{X}$ is precisely the numerator in (5.60). Consider the first moment of $\bar{X}$, this is given by

$$\begin{aligned}\mathbb{E}_0[\bar{X}] &= n^{-3} \sum_{1\leq i,j,k\leq n} \mathbb{E}_0[X_{ijk}] && (5.61)\\ &= n^{-3} \sum_{1\leq i,j,k\leq n} \frac{\mathbb{P}_0(i \leftrightarrow j)\mathbb{P}_0(i \leftrightarrow k)\mathbb{P}_0(j \leftrightarrow k \,|\, i \leftrightarrow j, i \leftrightarrow k)}{p^2} = (1 + o(1))\, C_d\,.\end{aligned}$$

Moreover, the second moment of $\bar{X}$ can be computed by splitting between the number of common vertices in the two triangles involved. This gives

$$\begin{aligned}\mathbb{E}_0[\bar{X}^2] &= n^{-6} \sum_{1\leq i,j,k,i',j',k'\leq n} \mathbb{E}_0[X_{ijk} X_{i'j'k'}] && (5.62)\\ &= n^{-6} \sum_{\substack{1\leq i,j,k,i',j',k'\leq n\\ \text{distinct}}} \mathbb{E}_0[X_{ijk} X_{i'j'k'}] \;+\; 3\,n^{-6} \sum_{\substack{1\leq i,j,k,j',k'\leq n\\ \text{distinct}}} \mathbb{E}_0[X_{ijk} X_{ij'k'}] && (5.63)\\ &\qquad +\; 3\,n^{-6} \sum_{\substack{1\leq i,j,k,k'\leq n\\ \text{distinct}}} \mathbb{E}_0[X_{ijk} X_{ijk'}] \;+\; n^{-6} \sum_{\substack{1\leq i,j,k\leq n\\ \text{distinct}}} \mathbb{E}_0[X_{ijk}^2] && (5.64)\\ &= (1 + o(1))\left[C_d^2 + 3\,\frac{C_d^2}{n} + 3\,\frac{C_d^2}{n^2 p} + \frac{C_d}{n^3 p^2}\right] = (1 + o(1))\, C_d^2\,, && (5.65)\end{aligned}$$

where the final step follows from the assumption that $p \geq \Omega(1/n)$. Hence, $\mathrm{Var}_0(\bar{X}) = \mathbb{E}_0[\bar{X}^2] - \mathbb{E}_0[\bar{X}]^2 = o(1)$, and therefore it follows by Chebyshev's inequality that $\bar{X} \xrightarrow{\mathbb{P}_0} C_d$. This shows that the numerator of (5.60) converges in probability to C_d and the denominator of (5.60) converges in probability to 1, so we have $\widehat{C}_d \xrightarrow{\mathbb{P}_0} C_d$. Finally, it follows from the continuous mapping theorem that $\hat{d} \xrightarrow{\mathbb{P}_0} d$, and we conclude that our estimator for the dimension is consistent under the null hypothesis.

Under the alternative hypothesis, the proof is largely similar. Because the botnet size $k = o(n)$ is small, it can be seen that the first and second moment of $\bar{X}$ converge to the same values, and therefore $\bar{X} \xrightarrow{\mathbb{P}_B} C_d$. Finally, we can again apply the continuous mapping theorem to show that our estimator for the dimension is consistent under the alternative hypothesis. □

5.5.6 Proof of Lemma 5.2: Consistency of the connection probability estimator

We start by showing that $\widehat{p}/p \xrightarrow{\mathbb{P}_0} 1$. Using the estimator $\widehat{p}$ from (5.15) it follows directly that $\mathbb{E}_0\big[\widehat{p}/p\big] = 1$. Therefore, we are left to compute

$$\mathbb{E}_0[(\widehat{p}/p)^2] = \binom{n}{2}^{-2} \sum_{\substack{1\le i<j\le n \\ 1\le i'<j'\le n}} \mathbb{E}_0\left[\frac{\mathbb{1}_{\{i\leftrightarrow j\}}}{p}\frac{\mathbb{1}_{\{i'\leftrightarrow j'\}}}{p}\right] \tag{5.66}$$

$$= \binom{n}{2}^{-2}\left(\left[\binom{n}{2}^2 - \binom{n}{2}\right] + \binom{n}{2}\frac{1}{p}\right) \tag{5.67}$$

$$= \left(1 - \binom{n}{2}^{-1} + \binom{n}{2}^{-1}\frac{1}{p}\right) = 1 + o(1)\,, \tag{5.68}$$

where we obtained the second equality by splitting between the case where $i \neq i'$ and $j \neq j'$, and the case where $i = i'$ and $j = j'$. Moreover, the final step followed from the assumption $p \geq \Omega(1/n)$. Therefore, it follows that $\mathrm{Var}_0(\widehat{p}/p) = \mathbb{E}_0\big[(\widehat{p}/p)^2\big] - \mathbb{E}_0\big[\widehat{p}/p\big]^2 = o(1)$, and hence $\widehat{p}/p \xrightarrow{\mathbb{P}_0} 1$ by Chebyshev's inequality. Moreover, for any distinct triplet $i, j, k \in V$,

$$p = \mathbb{P}_0(i \leftrightarrow j) = \mathbb{P}_B(i \leftrightarrow j)\,, \qquad p^2 = \mathbb{P}_0(i \leftrightarrow j, i \leftrightarrow k) = \mathbb{P}_B(i \leftrightarrow j, i \leftrightarrow k)\,. \tag{5.69}$$

Hence, performing the above computations under the measure $\mathbb{P}_B$ shows that $\widehat{p}/p \xrightarrow{\mathbb{P}_B} 1$ as well. □

Chapter 6

Changepoint detection in the preferential attachment model

Based on:
Detecting a late changepoint in the preferential attachment model,
G. Bet, K. Bogerd, R. M. Castro, and R. van der Hofstad,
In preparation.

Motivated by the problem of detecting an anomalous evolution of a network, we consider the preferential attachment random graph model with a *time-dependent* attachment function. We cast this problem as a hypothesis testing problem where the null hypothesis is a preferential attachment model with a constant affine attachment parameter δ_0, and the alternative hypothesis is a preferential attachment model where the affine attachment parameter changes from δ_0 to δ_1 at an unknown changepoint time τ_n. We focus on the regime where the changepoint may only occur rather late in time (i.e., $\tau_n = n - cn^\gamma$ with $c \geq 0$ and $\gamma \in (0,1)$). We present an asymptotically powerful test that is able to distinguish between the null and alternative hypothesis when $\gamma > 1/2$. Our test is based on precise estimates of the expected number of *minimal* degree vertices in the alternative model together with their fluctuations.

6.1 Introduction

One of the most celebrated successes of complex network theory has been the recognition that simple *dynamical* random graph models with local connection rules are able to successfully explain important macroscopic features observed in real-world networks. The preferential attachment model and its generalizations are perhaps the most successful of such models. Barabási and Albert [16] proposed this model to explain the occurrence of power-law degree sequences, which are often observed in

real-world networks such as the world wide web [4, 44] or internet [71], biological networks [72, 110, 130], or even in collaboration networks of movie actors [7, 86]. Furthermore, the typical distance between vertices in the preferential attachment model is small. This is called the *small-world* phenomenon, see [161, 162].

The preferential attachment model considers the entire evolution of a network by adding vertices one by one using a simple *preferential attachment* rule. Informally, as new vertices are added to the graph, they are more likely to attach to vertices that already have a large degree, hence further increasing the degree of these vertices. Accordingly, the degree of the oldest vertices grows as new vertices attach to the graph. On the other hand, the degree of the last few vertices to join is typically quite small. Since its introduction in [16], the preferential attachment model has received a tremendous amount of attention thanks to its early explanatory successes. The structural properties of the model are investigated formally in [32, 33], see also [104, 105] for a detailed overview on this model and many of its properties.

In general, however, not all vertices in a network follow the same connection criterion. For example, some major event could cause a change in the subsequent evolution of the network. To model this, we consider here a time-inhomogeneous affine preferential attachment model, where a new vertex v_t that enters the graph at time $1 \leq t \leq n$ connects to a pre-existing vertex with degree k with probability proportional to $f(k) = k + \delta(t)$. We consider the hypothesis testing problem where $\delta(t) = \delta_0$ remains constant under the null hypothesis, whereas under the alternative the affine attachment parameter $\delta(t)$ changes from δ_0 to δ_1 at an unknown changepoint τ_n. We focus on the regime where the changepoint may only occur late (i.e., $\tau_n = n - cn^\gamma$ with $c \geq 0$ and $\gamma \in (0, 1)$). This is explained in more detail in Section 6.2.

Related work. Our work nicely complements earlier results [14, 21] that focused on the detection of a changepoint in the setting of preferential attachment trees, where every vertex that enters the graph connects to $m = 1$ other edge. There are also some differences. First, our results consider the more general case of preferential attachment graphs, where vertices may enter the graph with $m \geq 1$ edges. The other difference between is that we focus on a late changepoint $\tau_n = n - cn^\gamma$, whereas [14, 21] focus on a changepoint that happens at a linear time $O(n)$. Thus, in our setting a much smaller number of vertices enter the graph after changepoint, making it harder to detect. We believe that our results are robust enough to be easily extended to the setting of a linear changepoint as well.

Although different from this work, there has been much interest in understanding and detecting the effect of an initial seed graph on the evolution of the preferential attachment model [46, 48, 49, 58, 123]. Here one starts with a given initial graph at time $t = 1$ and then grows the remaining graph according to the preferential attachment (or uniform attachment). The goal is then to estimate the initial graph based on an observation of the graph at a much later time.

Finally, changepoint detection has also received much attention in the setting of

dynamic stochastic block models [23, 152, 159, 160, 164]. There the aim is primarily to understand the evolution of the network's community structure.

6.2 Model and results

We formalize the problem of detecting a changepoint in a dynamical network as a hypothesis testing problem on random graphs. We first explain the model that we use in general, and then define concrete versions of this model for the null and alternative hypothesis.

We are given a single observation of a random graph $G_n = (V_n, E_n)$ with vertex set $V_n := \{v_0, \dots, v_n\}$ and $E_n \subseteq \{(i,j) : i,j \in V_n\}$ is the random set of edges. The observed graph G_n is then a sample from the affine preferential attachment model with parameters $m \geq 1$ and $\delta(t) > -m$. This model actually generates a sequence of graphs $(G_t)_{t=1}^n$, from which we observe only the final snapshot G_n.

There exist various versions of the preferential attachment model, each following slightly different conventions for adding new vertices. Here we consider the following model. The first graph G_1, also called the seed graph, consists of two vertices v_0 and v_1 connected by m edges. For $2 \leq t \leq n$, the graph G_t is constructed from G_{t-1} by adding a vertex v_t with m new edges one by one and with intermediate updating of degrees. To this end, define $G_{t,0}$ as the graph G_{t-1} together with the vertex v_t without any edges, and let $G_{t,1}, G_{t,2}, \dots, G_{t,m}$ be the intermediate graphs for each of the m edges emanating from v_t. For $1 \leq i \leq m$, the graph $G_{t,i}$ is constructed from $G_{t,i-1}$ by connecting v_t to a randomly selected vertex $v_s \in \{v_0, \dots, v_{t-1}\}$. This is where the parameter $\delta(t)$ comes in, because the conditional probability given $G_{t,i-1}$ that the ith edge of v_t connects to v_s is given by

$$\mathbb{P}\big(v_{t,i} \leftrightarrow v_s \,|\, G_{t,i-1}\big) = \frac{\deg_{v_s}(G_{t,i-1}) + \delta(t)}{\sum_{j=0}^{t-1}(\deg_{v_j}(G_{t,i-1}) + \delta(t))}, \tag{6.1}$$

where $\deg_{v_s}(G_{t,i-1})$ denotes the degree of v_s in $G_{t,i-1}$. After all m edges have been added to the vertex v_t we obtain the graph $G_t = G_{t,m}$.

The model above is rather general, as there are almost no restrictions on the function $\delta(t)$. For our hypothesis testing problem we will consider this is either a constant or a step function.

Under the null hypothesis, denoted by H_0, the observed graph $G_n = G_n^{(\delta_0)}$ is generated according the the preferential attachment model with fixed parameter $\delta(t) = \delta_0$. Thus, the attachment rule from (6.1) becomes

$$\mathbb{P}\big(v_{t,i} \leftrightarrow v_s \,|\, G_{t,i-1}^{(\delta_0)}\big) = \frac{\deg_{v_s}(G_{t,i-1}^{(\delta_0)}) + \delta_0}{2(t-1)m + t\delta_0 + (i-1)}, \tag{6.2}$$

where $v_s \in \{v_0, \dots, v_{t-1}\}$. Note that this is a special case of the preferential attachment

model from [60, 87], where every vertex enters the graph with exactly m edges.

Under the alternative hypothesis, denoted by H_1, the observed graph $G_n = G_n^{(\delta_0,\delta_1)}$ is similar except the parameter $\delta(t)$ changes from δ_0 to δ_1 for a small number of vertices at the very end of the process. Specifically, for a graph of size n, the changepoint happens at time $\tau_n = n - cn^\gamma$, where $\gamma \in (0,1)$ and $c \in (0,\infty)$ are fixed constants. That is, $\delta(t)$ is given by $\delta(t) = \mathbb{1}_{\{t<\tau_n\}}\delta_0 + \mathbb{1}_{\{t\geq\tau_n\}}\delta_1$. In this case, the attachment rule from (6.1) becomes

$$\mathbb{P}\big(v_{t,i} \leftrightarrow v_s \,|\, G_{t,i-1}^{(\delta_0,\delta_1)}\big) = \begin{cases} \dfrac{\deg_{v_s}(G_{t,i-1}^{(\delta_0,\delta_1)}) + \delta_0}{2(t-1)m + t\delta_0 + (i-1)}, & \text{if } t < \tau_n, \\ \dfrac{\deg_{v_s}(G_{t,i-1}^{(\delta_0,\delta_1)}) + \delta_1}{2(t-1)m + t\delta_1 + (i-1)}, & \text{if } t \geq \tau_n. \end{cases} \tag{6.3}$$

To summarize, the graph under the null hypothesis is denoted by $G_n = G_n^{(\delta_0)}$, and the graph with a changepoint under the alternative hypothesis is denoted by $G_n = G_n^{(\delta_0,\delta_1)}$. Both these models also depend on the parameter m, although this dependence is left implicit to avoid notational clutter. Furthermore, the total number of edges in a graph of size n is mn, and thus m can be considered known.

Assumptions and notation. Our primary goal is to characterize the asymptotic distinguishability between the null and alternative hypothesis in the asymptotic regime where n tends to ∞. Throughout this paper, when limits are unspecified they are taken as the graph size $n \to \infty$. The other parameters m, δ_0, δ_1, c, and γ are assumed to remain constant. We also use standard asymptotic notation as defined in Section 1.5.

6.2.1 Minimal degree test

In this section we present a test that can distinguish between the null and alternative hypotheses based on just the final snapshot of the graph G_n. Here we do not know the vertex labels and therefore do not know which vertices entered the graph early or which entered the graph late. The goal of our test is to determine whether the last $n - \tau_n = cn^\gamma$ vertices that entered the graph had parameter δ_0 or δ_1. To this end, define a test T as a function mapping the observed graph G_n to $\{0,1\}$, where $T(G_n) = 1$ indicates that the null hypothesis is rejected (i.e., the test indicates that the graph contains a changepoint), and $T(G_n) = 0$ otherwise.

To define our test we first need some additional notation. Let $N_k(G_n)$ be the number of vertices of degree k in the graph G_n, that is

$$N_k(G_n) := \sum_{t=1}^{n} \mathbb{1}_{\{\deg_{v_t}(G_n)=k\}}, \tag{6.4}$$

where we recall that $\deg_{v_t}(G_n)$ denotes the degree of v_t in G_n. For our model, under

both the null and alternative hypotheses, $N_k(G_n) = 0$ for $k < m$, and $N_m(G_n)$ denotes the number of vertices with minimal degree in G_n. The latter quantity will plays a crucial role in our test.

It is well-known that in the classical preferential attachment model $G_n^{(\delta_0)}$ the number of vertices of degree k is highly concentrated around $n\,p_k(\delta_0)$ for some sequence $(p_k(\delta_0))_{k=m}^{\infty}$. This motivates the interpretation of $p_k(\delta_0)$ as the limiting degree distribution of the random graph $G_n^{(\delta_0)}$. The expression for the probability mass of the minimal degree m is especially simple [60, 104], and is given by

$$p_m(\delta_0) := \frac{2 + \delta_0/m}{m + \delta_0 + 2 + \delta_0/m}. \tag{6.5}$$

We are now able to introduce our test, which is based on a comparison between the observed number of minimal degree vertices $N_m(G_n)$ to its asymptotic expected value under the null hypothesis $n\,p_m(\delta_0)$. When $N_m(G_n)$ deviates too much from $n\,p_m(\delta_0)$ we can reject the null hypothesis. This results in the following test:

Definition 6.1 (Minimal degree test)**.** Given a graph G_n of size $n \geq 1$ and significance level $\alpha \in (0, 1)$, the minimal degree test rejects the null hypothesis H_0 if

$$\left|N_m(G_n) - n\,p_m(\delta_0)\right| \geq m\sqrt{8n\log(2/\alpha)}. \tag{6.6}$$

The minimal degree test as defined above can written as the function $T(G_n) = \mathbb{1}\{|N_m(G_n) - np_m(\delta_0)| \geq m\sqrt{8n\log(2/\alpha)}\}$.

This brings us to the main result, where we show that the minimal degree test is asymptotically powerful if $\gamma > 1/2$. That is, when $\gamma > 1/2$ then it is possible to make the type-II error arbitrarily small for any significance level $\alpha \in (0, 1)$. Furthermore, when $\gamma = 1/2$ then it is possible to show that the type-II error is bounded by a constant that depends on the specific model parameters. The proof of this result is postponed to Section 6.4.

Theorem 6.1. *The type-I and type-II error of the minimal degree test from Definition 6.1 are bounded by*

$$\mathbb{P}(T(G_n^{(\delta_0)}) \neq 0) \leq (1 + o(1))\,\alpha, \tag{6.7}$$

$$\begin{aligned}\mathbb{P}(T(G_n^{(\delta_0,\delta_1)}) \neq 1) & \\ \leq \begin{cases} o(1) & \text{if } \gamma > 1/2, \\ (2 + o(1)) \exp\left(-\left(\left(\frac{c\,|1 - p_m(\delta_0)/p_m(\delta_1)|}{m\sqrt{8}} - \sqrt{\log(2/\alpha)}\right) \vee 0\right)^2\right) & \text{if } \gamma = 1/2. \end{cases}\end{aligned} \tag{6.8}$$

The reason that the type-I error is bounded by $\alpha \in (0, 1)$ is a direct consequence of how we have defined the minimal degree test. In fact, the threshold for rejecting the null hypothesis in (6.6) is chosen such that the Azuma-Hoeffding inequality directly implies that the type-I error is bounded by chosen significance level α.

To bound the type-II error the idea is to analyze $N_m(G_n^{(\delta_0,\delta_1)}) - n\,p_m(\delta_0)$ in two steps. First we consider the expected difference given by $\mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - n\,p_m(\delta_0)$ and show that this is approximately $cn^\gamma(1 - p_m(\delta_0)/p_m(\delta_1))$, and then we control the deviations of $N_m(G_n^{(\delta_0,\delta_1)}) - \mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})]$ again using the Azuma-Hoeffding inequality.

The downside of using the Azuma-Hoeffding inequality is that our test is likely to be overly conservative, and that the resulting error bounds from Theorem 6.1 are quite loose. To remedy this, we would need to more precisely quantify the deviations $N_m(G_n^{(\delta_0)})$ around its mean, both under the null and alternative. This is discussed in more detail in Section 6.3, where we conjecture that $N_m(G_n^{(\delta_0)})$ and $N_m(G_n^{(\delta_0,\delta_1)})$ are asymptotically normal.

6.3 Discussion

Here we remark on the results in this chapter, and explain some extensions of these results that we are working on. In Chapter 7 we also discuss several open problems that are related to the results in this chapter and that might be interesting to pursue.

Central limit theorem for $N_m(G_n)$. In the proof of Theorem 6.1 we use the Azuma-Hoeffding inequality to bound the deviations of $N_m(G_n)$ around its mean. As explained above, when the asymptotic distribution of $N_m(G_n)$ is known then this would make it possible to compute the type-I and type-II error asymptotically exactly. This would result in two main improvements. First, when the asymptotic distribution of $N_m(G_n^{(\delta_0)})$ is known under the null hypothesis then it possible to asymptotically calibrate the minimal degree test, and ensure the type-I error converges to a given significance level $\alpha \in (0,1)$. Second, when the asymptotic distribution of $N_m(G_n^{(\delta_0,\delta_1)})$ is also known under the alternative hypothesis then it is possible to optimize the constants for the case $\gamma = 1/2$ in Theorem 6.1, and make it possible to exactly compute the asymptotic power of the minimal degree test.

Under the null hypothesis and for $m = 1$, it is known that $N_k(G_n^{(\delta_0)})$, with $k \geq m$, admits a central limit theorem [153]. In particular, this shows that $N_m(G_n^{(\delta_0)})$ is asymptotically normally distributed. Furthermore, our preliminary computations suggest that it is possible to extend this result to the case $m > 1$ and derive a central limit theorem for $N_m(G_n^{(\delta_0)})$ that is valid for general $m \geq 1$. If this is indeed the case, then it would be possible to replace the rejection threshold in (6.6) by a quantile of the normal distribution with the appropriate variance. Furthermore, using the same reasoning as in the proof of Theorem 6.1, it can then be shown that the resulting test has a type-I error that converges to the given significance level $\alpha \in (0,1)$. Thus, this would guarantee that the test is asymptotically correctly calibrated.

Under the alternative hypothesis, there are no known results about the asymptotic normality of $N_m(G_n^{(\delta_0,\delta_1)})$. However, our preliminary calculations suggest that $N_m(G_n^{(\delta_0,\delta_1)})$ also admits a central limit theorem with different mean but with the same variance as under the null hypothesis. The reason for this is that in our model the

changepoint happens very late. Because of this, the number of vertices that enter after the changepoint is too small to change the asymptotic variance, but it is large enough to result in a considerably different mean.

Test for unknown δ_0. The minimal degree test from Definition 6.1 requires knowledge of the parameter δ_0. When this is not available, a possible approach could be to estimate it and use an estimator $\widehat{\delta}$ as a plug-in instead of the true value δ_0. For this approach to work, we would need to bound how much N_m and $\widehat{\delta}$ are correlated, and adjust our test accordingly to compensate for this.

A good candidate for this would be the estimator proposed in [87]. Furthermore, this estimator is shown to be asymptotically normally distributed. Viewing this in light of the previous discussion point, if we could show that $N_m(G_n)$ admits a central limit theorem then it could also be possible to derive a joint central limit theorem for $(N_m, \widehat{\delta})$. Such a result would be very useful because that would make it possible to precisely define a test where the asymptotic type-I error is exactly equal to a given significance level $\alpha \in (0,1)$.

6.4 Proof

Here we prove Theorem 6.1 and compute the asymptotic type-I and type-II error of the minimal degree test from Definition 6.1. We consider the type-I and type-II error separately.

Type-I error: We start by using [60, Proposition 2.2] (see also [104, Proposition 8.7]), which states that there exists a constant $C_0 = C_0(\delta_0, m)$ such that, for all $n \geq 1$,

$$|\mathbb{E}[N_m(G_n^{(\delta_0)})] - n\, p_m(\delta_0)| \leq C_0,$$

where $p_m(\delta_0)$ is given by (6.5). Furthermore, we use the Azuma-Hoeffding inequality together with [104, Lemmas 8.5 and 8.6]. This gives, for any $x > 0$,

$$\mathbb{P}\big(|N_m(G_n^{(\delta_0)}) - \mathbb{E}[N_m(G_n^{(\delta_0)})]| \geq x\big) \leq 2\mathrm{e}^{-x^2/8m^2n}.$$

Combining the above, we obtain that the type-I error of the minimal degree test from Definition 6.1 is bounded by

$$\begin{aligned}
\mathbb{P}(T(G_n^{(\delta_0)}) \neq 0) &= \mathbb{P}\Big(|N_m(G_n^{(\delta_0)}) - n\, p_m(\delta_0)| \geq m\sqrt{8n\log(2/\alpha)}\Big) \\
&\leq \mathbb{P}\Big(|N_m(G_n^{(\delta_0)}) - \mathbb{E}[N_m(G_n^{(\delta_0)})]| \geq m\sqrt{8n\log(2/\alpha)} - C_0\Big) \\
&\leq 2\exp\left(-\frac{(m\sqrt{8n\log(2/\alpha)} - C_0)^2}{8m^2n}\right) = (1 + o(1))\alpha.
\end{aligned}$$

This shows that the type-I error is bounded by α, completing the first part of the proof.

Type-II error: For the type-II error we first quantify $\mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - n\,p_m(\delta_0)$, and then we apply the Azuma-Hoeffding inequality to control the deviations similarly as in the previous part.

Given a graph G_n, define $N_m^\star(G_n)$ as the number vertices with degree m in G_n that entered the graph after the changepoint τ_n. That is,

$$N_m^\star(G_n) := \sum_{t=\tau_n}^{n} \mathbb{1}_{\{\deg_{v_t}(G_n)=m\}}.$$

Because we consider an affine preferential attachment model, vertices that enter the graph after the changepoint have exactly the same attachment dynamics when the graph $G_n = G_n^{(\delta_1)}$ is a preferential attachment model with parameter δ_1, as well as when the graph $G_n = G_n^{(\delta_0,\delta_1)}$ is a preferential attachment model with a changepoint (i.e., the model which has δ_0 at time $1 \le t < \tau_n$ and changes to δ_1 for the remaining time $\tau_n \le t \le n$). Therefore, because the quantity $N_m^\star(G_n)$ only count vertices that enter the graph after the changepoint, we can relate $N_m^\star(G_n)$ in both models. This gives the relation

$$\begin{aligned}
&\mathbb{E}[N_m(G_n^{(\delta_1)})] - \mathbb{E}[N_m(G_{\tau_n}^{(\delta_1)})] \prod_{t=\tau_n}^{n} \prod_{i=1}^{m} \Big(1 - \frac{m+\delta_1}{t(2m+\delta_1) - 2m + i - 1}\Big) \\
&\quad = \mathbb{E}[N_m^\star(G_n^{(\delta_1)})] \\
&\quad = \mathbb{E}[N_m^\star(G_n^{(\delta_0,\delta_1)})] \\
&\quad = \mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - \mathbb{E}[N_m(G_{\tau_n}^{(\delta_0)})] \prod_{t=\tau_n}^{n} \prod_{i=1}^{m} \Big(1 - \frac{m+\delta_1}{t(2m+\delta_1) - 2m + i - 1}\Big).
\end{aligned} \tag{6.9}$$

Note that the product term in (6.9) gives the probability that a given vertex with degree m at the changepoint time τ_n receives no further connections (i.e., that vertex remains a degree m vertex in the graph at time n). Hence, the first and last equation in (6.9) compute $\mathbb{E}[N_m^\star(G_n)]$ by first counting the expected number of degree m vertices in G_n, and then subtracting the expected number of degree m vertices that entered the graph before the changepoint τ_n.

To continue, we will derive an approximation for the product in (6.9). Let A be the product term in (6.9), and note that there exists a $\lambda \in [1,2]$ such that

$$\begin{aligned}
A &:= \prod_{t=\tau_n}^{n} \prod_{i=1}^{m} \Big(1 - \frac{m+\delta_1}{t(2m+\delta_1) - 2m + i - 1}\Big) = \prod_{t=\tau_n}^{n} \prod_{i=1}^{m} \Big(1 - \frac{m+\delta_1}{t(2m+\delta_1) - \lambda m}\Big) \\
&= \left(\frac{\Gamma\Big(n + 1 - \frac{(1+\lambda)m+\delta_1}{2m+\delta_1}\Big) \Big/ \Gamma\Big(\tau_n - \frac{(1+\lambda)m+\delta_1}{2m+\delta_1}\Big)}{\Gamma\Big(n + 1 - \frac{\lambda m}{2m+\delta_1}\Big) \Big/ \Gamma\Big(\tau_n - \frac{\lambda m}{2m+\delta_1}\Big)} \right)^m .
\end{aligned} \tag{6.10}$$

To simplify notation, let $\alpha := ((1+\lambda)m+\delta_1)/(2m+\delta_1)$ and $\beta := \lambda m/(2m+\delta_1)$. Then, using Stirling approximation $\log(\Gamma(x)) = (x-1/2)\log(x) - x + \log(2\pi)/2 + O(1/x)$, and taking the logarithm of (6.10) we obtain

$$\begin{aligned}\log(A) &= m\log\left(\frac{\Gamma(n+1-\alpha)\big/\Gamma(\tau_n-\alpha)}{\Gamma(n+1-\beta)\big/\Gamma(\tau_n-\beta)}\right)\\ &= m\big[(n+1/2-\alpha)\log(n+1-\alpha) - (n+1/2-\beta)\log(n+1-\beta)\big]\\ &\quad - m\big[(\tau_n-\alpha-1/2)\log(\tau_n-\alpha) - (\tau_n-\beta-1/2)\log(\tau_n-\beta)\big] + O\Big(\frac{1}{n}\Big).\end{aligned}$$

Using a Taylor expansion around $n=\infty$ we can simplify the above. This yields

$$\begin{aligned}\log(A) &= m(\beta-\alpha)(1+\log(n)) - m(\beta-\alpha)(1+\log(\tau_n)) + O\Big(\frac{1}{n}\Big)\\ &= m(\beta-\alpha)\log\Big(\frac{n}{\tau_n}\Big) + O\Big(\frac{1}{n}\Big)\\ &= \frac{m(m+\delta_1)}{2m+\delta_1}\log\Big(\frac{\tau_n}{n}\Big) + O\Big(\frac{1}{n}\Big)\\ &= \frac{m(m+\delta_1)}{2m+\delta_1}\log\Big(1-\frac{cn^\gamma}{n}\Big) + O\Big(\frac{1}{n}\Big).\end{aligned}$$

Therefore, using another Taylor expansion, we have the following approximation for A from (6.10),

$$\begin{aligned}A &= \exp\left(\frac{m(m+\delta_1)}{2m+\delta_1}\log\Big(1-\frac{cn^\gamma}{n}\Big) + O\Big(\frac{1}{n}\Big)\right)\\ &= 1 - \frac{m(m+\delta_1)}{2m+\delta_1}\frac{cn^\gamma}{n} + O\bigg(\Big(\frac{cn^\gamma}{n}\Big)^2\bigg).\end{aligned} \tag{6.11}$$

Following our steps and using the relation between $G_n^{(\delta_1)}$ and $G_n^{(\delta_0,\delta_1)}$ from (6.9) together with the definition of A, we obtain

$$\mathbb{E}[N_m(G_n^{(\delta_1)})] - A\,\mathbb{E}[N_m(G_{\tau_n}^{(\delta_1)})] = \mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - A\,\mathbb{E}[N_m(G_{\tau_n}^{(\delta_0)})].$$

Rearranging the terms above and plugging in (6.11), we obtain

$$\begin{aligned}&\mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - \mathbb{E}[N_m(G_n^{(\delta_1)})]\\ &\quad= -A\big(\mathbb{E}[N_m(G_{\tau_n}^{(\delta_1)})] - \mathbb{E}[N_m(G_{\tau_n}^{(\delta_0)})]\big)\\ &\quad= \left(\frac{m(m+\delta_1)}{2m+\delta_1}\frac{cn^\gamma}{n} - 1 + O\bigg(\Big(\frac{cn^\gamma}{n}\Big)^2\bigg)\right)\big(\mathbb{E}[N_m(G_{\tau_n}^{(\delta_1)})] - \mathbb{E}[N_m(G_{\tau_n}^{(\delta_0)})]\big).\end{aligned} \tag{6.12}$$

To continue, we again use [60, Proposition 2.2] (see also [104, Proposition 8.7]), which

states that there exists constants $C_0 = C_0(\delta_0, m)$ and $C_1 = C_1(\delta_1, m)$ such that, for all $n \geq 1$,

$$\left|\mathbb{E}[N_m(G_n^{(\delta_0)})] - n\, p_m(\delta_0)\right| \leq C_0, \qquad \text{and} \qquad \left|\mathbb{E}[N_m(G_n^{(\delta_1)})] - n\, p_m(\delta_1)\right| \leq C_1,$$

where $p_m(\delta_0)$ and $p_m(\delta_1)$ are given by (6.5). Combining this with (6.12), we obtain the difference in the expected number of vertices of degree m at time n between the null model and alternative model. This shows that

$$\begin{aligned}
&\mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - n\, p_m(\delta_0) \\
&\quad = \mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - \mathbb{E}[N_m(G_n^{(\delta_1)})] + \mathbb{E}[N_m(G_n^{(\delta_1)})] - n\, p_m(\delta_0) \\
&\quad = \left(\frac{m(m+\delta_1)}{2m+\delta_1}\frac{cn^\gamma}{n} - 1 + O\left(\left(\frac{cn^\gamma}{n}\right)^2\right)\right)(n - cn^\gamma)\big(p_m(\delta_1) - p_m(\delta_0)\big) \\
&\qquad\qquad + n\big(p_m(\delta_1) - p_m(\delta_0)\big) + O(1) \\
&\quad = cn^\gamma\left(1 + \frac{m(m+\delta_1)}{2m+\delta_1} + O\left(\frac{cn^\gamma}{n}\right)\right)\big(p_m(\delta_1) - p_m(\delta_0)\big) + O(1).
\end{aligned}$$

Hence, for $1/2 \leq \gamma < 1$, it follows that

$$\begin{aligned}
\mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - n\, p_m(\delta_0) &= cn^\gamma\left(1 + \frac{m(m+\delta_1)}{2m+\delta_1} + o(1)\right)\big(p_m(\delta_1) - p_m(\delta_0)\big) \\
&= cn^\gamma\left(1 - \frac{p_m(\delta_0)}{p_m(\delta_1)} + o(1)\right),
\end{aligned}$$

where $p_m(\delta_0)$ and $p_m(\delta_1)$ are given by (6.5).

Similarly as for the type-I error, we use the Azuma-Hoeffding inequality (together with [104, Lemmas 8.5 and 8.6]). It can easily be checked that this result also holds in the preferential attachment model with a changepoint. This gives

$$\mathbb{P}\left(\left|N_m(G_n^{(\delta_0,\delta_1)}) - \mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})]\right| \geq x\right) \leq 2\mathrm{e}^{-x^2/8m^2n}.$$

Combining the above

$$\begin{aligned}
\mathbb{P}(T(G_n^{(\delta_0,\delta_1)}) \neq 1) &= \mathbb{P}\left(\left|N_m(G_n^{(\delta_0,\delta_1)}) - n\, p_m(\delta_0)\right| < m\sqrt{8n\log(2/\alpha)}\right) \\
&\leq \mathbb{P}\Big(\left|N_m(G_n^{(\delta_0,\delta_1)}) - \mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})]\right| \\
&\qquad\qquad \geq \left(\left|\mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})] - n\, p_m(\delta_0)\right| - m\sqrt{8n\log(2/\alpha)}\right) \vee 0\Big) \\
&\leq \mathbb{P}\Big(\left|N_m(G_n^{(\delta_0,\delta_1)}) - \mathbb{E}[N_m(G_n^{(\delta_0,\delta_1)})]\right| \\
&\qquad\qquad \geq \left(cn^\gamma\left|1 - p_m(\delta_0)/p_m(\delta_1) + o(1)\right| - m\sqrt{8n\log(2/\alpha)}\right) \vee 0\Big) \\
&\leq 2\exp\left(-\frac{\left((cn^\gamma\left|1 - p_m(\delta_0)/p_m(\delta_1) + o(1)\right| - m\sqrt{8n\log(2/\alpha)}\right) \vee 0\right)^2}{8m^2n}\right).
\end{aligned}$$

Finally, considering the cases $\gamma > 1/2$ and $\gamma = 1/2$ separately, this gives

$$\mathbb{P}(T(G_n^{(\delta_0,\delta_1)}) \neq 1) \leq \begin{cases} o(1) & \text{if } \gamma > 1/2, \\ (2+o(1)) \exp\left(-\left(\left(\frac{c|1-p_m(\delta_0)/p_m(\delta_1)|}{m\sqrt{8}} - \sqrt{\log(2/\alpha)}\right) \vee 0\right)^2\right) & \text{if } \gamma = 1/2. \end{cases}$$

This completes the second part of the proof. □

Chapter 7

Discussion and open problems

In this thesis we have presented and analyzed several methods that can be used to detect planted structures in random graphs. This resulted in new ideas and interesting directions for future research. In this final chapter, we discuss some of these open problems and how they are related to the results obtained in the previous chapters.

7.1 Two-point concentration of the clique and quasi-clique number

Two-point concentration of the clique number in an Erdős-Rényi random graph has been known since the seventies [125, 126]. More recently this result has been extended to other random graph models. For instance, two-point concentration of the clique number is proven for random geometric graphs [137], and in Chapter 2 for rank-1 random graphs. Moreover, slightly weaker concentration results are known for dense inhomogeneous random graphs, see [66] and Chapter 3. When the largest clique is allowed to have a fraction of edges missing then this is called a quasi-clique. Two-point concentration for the size of the largest quasi-clique was only recently shown for dense Erdős-Rényi random graphs [13], and in Chapter 3 we have obtained a slightly weaker concentration result in the inhomogeneous setting.

These results seem to suggest that there might be a more general underlying principle that ensures these concentration results. It would therefore be interesting to investigate which properties a random graph model needs to have in order to guarantee two-point concentration of the clique or quasi-clique number, and in particular to what extend this is affected by inhomogeneity and underlying geometry of the random graph model.

7.2 Detecting a botnet in a geometric inhomogeneous random graph

In Chapter 5 we have considered the problem of detecting a planted botnet in a random geometric graph. However, for many applications the random geometric graph model is not the most appropriate, and it would be interesting to investigate the possibilities of extending the results from Chapter 5 to other random graph models. A good candidate for this would be the geometric inhomogeneous random graph from [41, 42], or the related hyperbolic random graph [29, 117]. This model better reflects the clustering, degree inhomogeneity, and distances observed in many real-world networks [77, 81, 94, 138, 156], and might therefore be more suitable to use as null hypothesis in the setting of botnet detection.

To be more concrete, consider the following model. For each vertex $i \in V$, let W_i be the *weight* sampled from a power-law distribution with exponent $\tau > 2$, and let X_i be the *location* which is a uniform sample on the d-dimensional torus T^d. Given a constant $\alpha > 1$ and conditionally on the weights and locations, two vertices are connected independently with probability

$$\begin{aligned} p_{ij} &:= \mathbb{P}\big((i,j) \in E \,\big|\, (X_k)_{k\in V}, (W_k)_{k\in V}\big) \\ &= \begin{cases} \Theta\Big(\Big(\frac{W_i W_j}{n\|X_i - X_j\|^d}\Big)^{\alpha} \wedge 1\Big), & \text{if } i,j \in V \setminus B, \\ \Theta\Big(\frac{W_i W_j}{n} \wedge 1\Big), & \text{if } i \in B \text{ or } j \in B, \end{cases} \end{aligned} \tag{7.1}$$

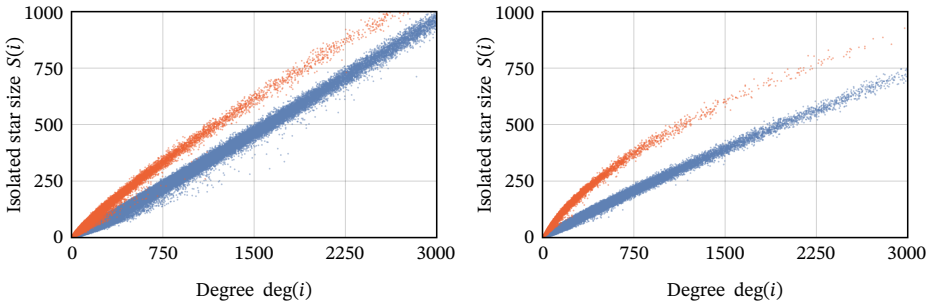
where $B \subseteq V$ denotes the set of botnet vertices, which could be the empty-set to indicate that there is no botnet.

A pair of vertices with no endpoints in the botnet is connected with probability equal to that in the geometric inhomogeneous random graph [41, 42], and when one of the endpoints is in the botnet then these vertices connect according to the definition of the Chung-Lu model [53, 54, 55]. A nice property of this setup is that the marginal edge probabilities are the same for all pairs of vertices, regardless of whether the endpoints belong to the botnet or not. That is, for every pair of vertices $i, j \in V$ it follows that $\mathbb{P}((i,j) \in E \,|\, (W_k)_{k\in V}) = \Theta(W_i W_j/n \wedge 1)$ by integrating out the locations X_i and X_j, see [42, Lemma 3.3]. This means that it is possible to choose the constants hidden behind the $\Theta(\cdot)$ in (7.1) such that the expected degree is same for both botnet and non-botnet vertices.

Under the null hypothesis we assume that there is no botnet, so $B = \emptyset$. On the other hand, under the alternative hypothesis the graph does contain a small number botnet vertices. The goal is then similar to that in Chapter 5, namely to develop a method that can accurately test for the presence of a botnet in a given observed graph.

Intuitively, it seems reasonable to expect that the isolated star test from Chapter 5 can be adapted to the model above. The reason is that the geometry underlying the

non-botnet vertices will produce many local connections, resulting in relatively small isolated stars. On the other hand, the botnet vertices ignore this geometry and will therefore tend to form larger isolated stars. This is confirmed by the simulation study presented in Figure 7.1, where the results of 2500 samples of the model described above are collated. Here we can see a clear separation in the size of the largest isolated star $S(i)$ of normal and botnet vertices. As predicted, a botnet vertex seems to always have a larger isolated star than a normal vertex of similar degree, with the difference being more pronounced in vertices with large degree. Thus a test that rejects the null hypothesis when the observed graph contains a vertex $i \in V$ that has both a large degree $\deg(i)$ and a large degree corrected isolated star $S(i)/\deg(i)$ could be a good candidate to detect the presence of a botnet in a geometric inhomogeneous random graph, where we used $\deg(i)$ and $S(i)$ to denote the degree and the size of the largest isolated star of $i \in V$ respectively. This suggest that such a degree corrected isolated star test is asymptotically powerful when the graph contains at least some botnet vertex with large enough degree.



(a) Power-law weights with exponent $\tau = 2.2$.

(b) Power-law weights with exponent $\tau = 2.8$.

Figure 7.1: The isolated star size as a function of vertex degree for every vertex from 2500 samples of a geometric inhomogeneous random graph with a planted botnet. The botnet vertices are red while the normal vertices are blue. The parameters are: graph size $n = 10000$, botnet size $|B| = 500$, dimension $d = 2$, and $\alpha = 2$. The constants hidden behind the $\Theta(\cdot)$ in (7.1) are chosen such that the average degree is equal to 20.

7.3 Changepoint detection in the preferential attachment model

In Chapter 6 we have considered the preferential attachment model and we have studied the possibility of detecting a change in the attachment function. In particular, we focused on the setting where the changepoint, if present, happens at a very late point in time, meaning that only a sub-linear amount of vertices enter the graph after the changepoint. We showed that it is indeed possible to detect such a changepoint, and that this can simply be achieved by counting the number of vertices with minimal

degree $N_m(n)$ in the observed graph. This project can be extended in several directions that would each be interesting to explore further. We will discuss these separately below.

Linear time changepoint. Instead of considering a late changepoint, we could consider a changepoint that happens at time $\tau_n = cn$, with $c \in (0,1)$. In this case, the problem actually becomes easier, provided δ_0 is known (i.e., when the null model is completely specified). However, when δ_0 is unknown then it needs to be estimated and this is where the main difficulty lies.

For a late changepoint it is relatively easy to estimate δ_0 because the graph is mostly composed of vertices that were incorporated before the changepoint. As explained in Chapter 6, a good candidate would be to estimate δ_0 using the estimator $\widehat{\delta}$ described in [87]. However, when the changepoint happens earlier, say at time $\tau_n = cn$ or even earlier, then there could be a significant bias in the estimator $\widehat{\delta}$. Here it would be interesting to see what the earliest changepoint time could be such that it is still possible to use the estimator $\widehat{\delta}$ from [87] instead of using the true value δ_0. Furthermore, it would be interesting to investigate whether it is possible to adjust the estimator $\widehat{\delta}$ from [87] so that it becomes less biased under the alternative hypothesis, and thereby making it possible to reliably detect an even earlier changepoint.

Estimating the changepoint. We were mostly interested in detecting whether a given graph contains a changepoint or not. Another very interesting problem would be to estimate the changepoint time. This problem of estimating a changepoint was considered under some restrictions ($m = 1$, changepoint linear in time) in [14, 21]. It would be interesting to investigate whether it is also possible to estimate a changepoint that happens very late, or in a preferential attachment model with $m > 1$.

Detecting alternating vertex types. In all the above we consider the detection of a changepoint, where all vertices before the changepoint have parameter δ_0 and all vertices after the changepoint have parameter δ_1. However, the two types of vertices could also enter the graph in alternating or random order. That is, given a parameter $p \in [0,1]$, each vertex that enters the graph has parameter δ_0 with probability p and it has parameter δ_1 with probability $1-p$. We would then like to test whether all vertices had parameter δ_0 (i.e., $p = 1$) versus the alternative where the graph contains both types of vertices.

From a statistical point of view this would make the problem more difficult. It would therefore be interesting to see to what extent it is still possible to distinguish between the null and alternative models, especially in the setting where the parameters δ_0 and δ_1 are not known.

Bibliography

[1] Abbe, E. 'Community detection and stochastic block models: recent developments'. *Journal of Machine Learning Research* 18.177 (2017), pp. 1–86.

[2] Abello, J., Pardalos, P. M., and Resende, M. G. C. 'On maximum clique problems in very large graphs'. *External memory algorithms*. Ed. by J. M. Abello and J. S. Vitter. Vol. 50. American Mathematical Society, 1999, pp. 119–130.

[3] Abello, J., Resende, M. G., and Sudarsky, S. 'Massive quasi-clique detection'. *LATIN 2002: Theoretical Informatics*. Ed. by S. Rajsbaum. Vol. 2286. LATIN 2002. Lecture Notes in Computer Science. Springer, 2002, pp. 598–612.

[4] Adamic, L. A., Huberman, B. A., Barabási, A.-L., Albert, R., Jeong, H., and Bianconi, G. 'Power-law distribution of the world wide web'. *Science* 287.5461 (2000), p. 2115.

[5] Addario-Berry, L., Broutin, N., Devroye, L., and Lugosi, G. 'On combinatorial testing problems'. *The Annals of Statistics* 38.5 (2010), pp. 3063–3092.

[6] Alba, R. D. 'A graph-theoretic definition of a sociometric clique'. *The Journal of Mathematical Sociology* 3.1 (1973), pp. 113–126.

[7] Albert, R. and Barabási, A. L. 'Topology of evolving networks: Local events and universality'. *Physical Review Letters* 85.24 (2000), pp. 5234–5237.

[8] Alon, N., Krivelevich, M., and Sudakov, B. 'Finding a large hidden clique in a random graph'. *Random Structures & Algorithms* 13.3-4 (1998), pp. 457–466.

[9] Arias-Castro, E., Candès, E. J., and Durand, A. 'Detection of an anomalous cluster in a network'. *The Annals of Statistics* 39.1 (2011), pp. 278–304.

[10] Arias-Castro, E., Candès, E. J., Helgason, H., and Zeitouni, O. 'Searching for a trail of evidence in a maze'. *The Annals of Statistics* 36.4 (2008), pp. 1726–1757.

[11] Arias-Castro, E. and Verzelen, N. 'Community detection in dense random networks'. *The Annals of Statistics* 42.3 (2014), pp. 940–969.

[12] Arias-Castro, E. and Verzelen, N. 'Community detection in sparse random networks'. *The Annals of Applied Probability* 25.6 (2015), pp. 3465–3510.

[13] Balister, P., Bollobás, B., Sahasrabudhe, J., and Veremyev, A. 'Dense subgraphs in random graphs'. *Discrete Applied Mathematics* 260 (2019), pp. 66–74.

[14] Banerjee, S., Bhamidi, S., and Carmichael, I. 'Fluctuation bounds for continuous time branching processes and nonparametric change point detection in growing networks' (2018).

[15] Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T. 'Evolution of the social network of scientific collaborations'. *Physica A: Statistical Mechanics and its Applications* 311.3-4 (2002), pp. 590–614.

[16] Barabási, A. L. and Albert, R. 'Emergence of scaling in random networks'. *Science* 286.5439 (1999), pp. 509–512.

[17] Baraud, Y. 'Non-asymptotic minimax rates of testing in signal detection'. *Bernoulli* 8.5 (2002), pp. 577–606.

[18] Barthélemy, M. 'Spatial networks'. *Physics Reports* 499.1-3 (2011), pp. 1–101.

[19] Bet, G., Bogerd, K., Castro, R. M., and Hofstad, R. van der. 'Detecting a botnet in a network'. 2020. arXiv: 2005.10650.

[20] Bet, G., Bogerd, K., Castro, R. M., and Hofstad, R. van der. 'Detecting a change-point in the preferential attachment model' (2020+).

[21] Bhamidi, S., Jin, J., and Nobel, A. 'Change point detection in network models: Preferential attachment and long range dependence'. *Annals of Applied Probability* 28.1 (2018), pp. 35–78.

[22] Bhamidi, S., Steele, J. M., and Zaman, T. 'Twitter event networks and the superstar model'. *Annals of Applied Probability* 25.5 (2015), pp. 2462–2502.

[23] Bhattacharjee, M., Banerjee, M., and Michailidis, G. 'Change point estimation in a dynamic stochastic block model'. Tech. rep. 107. 2020, pp. 1–59.

[24] Bianconi, G. and Marsili, M. 'Emergence of large cliques in random scale-free network'. *Europhysics Letters (EPL)* 74.4 (2005), pp. 740–746.

[25] Bianconi, G. and Marsili, M. 'Number of cliques in random scale-free network ensembles'. *Physica D: Nonlinear Phenomena* 224.1-2 (2006), pp. 1–6.

[26] Bogerd, K. 'Quasi-cliques in inhomogeneous random graphs'. 2020. arXiv: 2009.04945.

[27] Bogerd, K., Castro, R. M., and Hofstad, R. van der. 'Cliques in rank-1 random graphs: the role of inhomogeneity'. *Bernoulli* 26.1 (2020), pp. 253–285.

[28] Bogerd, K., Castro, R. M., Hofstad, R. van der, and Verzelen, N. 'Detecting a planted community in an inhomogeneous random graph'. 2019. arXiv: 1909.03217.

[29] Boguñá, M., Papadopoulos, F., and Krioukov, D. 'Sustaining the internet with hyperbolic mapping'. *Nature Communications* 1.62 (2010).

[30] Bollobás, B. and Erdős, P. 'Cliques in random graphs'. *Mathematical Proceedings of the Cambridge Philosophical Society* 80.4191 (1976), pp. 419–427.

[31] Bollobás, B., Janson, S., and Riordan, O. 'The phase transition in inhomogeneous random graphs'. *Random Structures & Algorithms* 31.1 (2007), pp. 3–122.

[32] Bollobás, B. and Riordan, O. 'The diameter of a scale-free random graph'. *Combinatorica* 24.1 (2004), pp. 5–34.

[33] Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. 'The degree sequence of a scale-free random graph process'. *Random Structures & Algorithms* 18.3 (2001), pp. 279–290.

[34] Boppana, R. and Halldórsson, M. M. 'Approximating maximum independent sets by excluding subgraphs'. *BIT* 32.2 (1992), pp. 180–196.

[35] Bordenave, C., Lelarge, M., and Massoulié, L. 'Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs'. *The Annals of Probability* 46.1 (2018), pp. 1–71.

[36] Boucheron, S., Lugosi, G., and Massart, P. 'Concentration inequalities: a nonasymptotic theory of independence'. Oxford University Press, 2013.

[37] Bourgeois, N., Giannakos, A., Lucarelli, G., Milis, I., Paschos, V. T., and Pottié, O. 'The MAX QUASI-INDEPENDENT SET problem'. *Journal of Combinatorial Optimization* 23.1 (2012), pp. 94–117.

[38] Bradonjić, M., Elsässer, R., Friedrich, T., Sauerwald, T., and Stauffer, A. 'Efficient broadcast on random geometric graphs'. *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete algorithms*. 2010, pp. 1412–1421.

[39] Brennan, M., Bresler, G., and Huleihel, W. 'Reducibility and computational lower bounds for problems with planted sparse structure'. *Proceedings of the 31st Conference on Learning Theory*. Vol. 75. Proceedings of Machine Learning Research. 2018, pp. 48–166.

[40] Bresler, G. and Nagaraj, D. 'Optimal single sample tests for structured versus unstructured network data'. *Proceedings of the 31st Conference on Learning Theory*. Proceedings of Machine Learning Research. 2018.

[41] Bringmann, K., Keusch, R., and Lengler, J. 'Average distance in a general class of scale-free networks with underlying geometry'. 2016. arXiv: 1602.05712.

[42] Bringmann, K., Keusch, R., and Lengler, J. 'Geometric inhomogeneous random graphs'. *Theoretical Computer Science* 760 (2019), pp. 35–54.

[43] Britton, T., Deijfen, M., and Martin-L of, A. 'Generating simple random graphs with prescribed degree distribution'. *Journal of Statistical Physics* 124.6 (2006), pp. 1377–1397.

[44] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. 'Graph structure in the web'. *Computer Networks* 33.1 (2000), pp. 309–320.

[45] Brunato, M., Hoos, H. H., and Battiti, R. 'On effectively finding maximal quasi-cliques in graphs'. *Learning and Intelligent Optimization*. Ed. by V. Maniezzo, R. Battiti, and J.-P. Watson. Vol. 5313. LION 2007. Lecture Notes in Computer Science. Springer, 2008, pp. 41–55.

[46] Bubeck, S., Devroye, L., and Lugosi, G. 'Finding Adam in random growing trees'. *Random Structures & Algorithms* 50.2 (2017), pp. 158–172.

[47] Bubeck, S., Ding, J., Eldan, R., and Rácz, M. Z. 'Testing for high-dimensional geometry in random graphs'. *Random Structures & Algorithms* 49.3 (2016), pp. 503–532.

[48] Bubeck, S., Eldan, R., Mossel, E., and Rácz, M. Z. 'From trees to seeds: On the inference of the seed from large trees in the uniform attachment model'. *Bernoulli* 23.4A (2017), pp. 2887–2916.

[49] Bubeck, S., Mossel, E., and Rácz, M. Z. 'On the influence of the seed graph in the preferential attachment model'. *IEEE Transactions on Network Science and Engineering* 2.1 (2015), pp. 30–39.

[50] Butucea, C. and Ingster, Y. I. 'Detection of a sparse submatrix of a high-dimensional noisy matrix'. *Bernoulli* 19.5B (2011), pp. 2652–2688.

[51] Caltagirone, F., Lelarge, M., and Miolane, L. 'Recovering asymmetric communities in the stochastic block model'. *IEEE Transactions on Network Science and Engineering* 5.3 (2016), pp. 237–246.

[52] Chen, L. H. Y. 'Poisson approximation for dependent trials'. *Annals of Probability* 3.3 (1975), pp. 534–545.

[53] Chung, F. and Lu, L. 'Connected components in random graphs with given expected degree sequences'. *Annals of Combinatorics* 6.2 (2002), pp. 125–145.

[54] Chung, F. and Lu, L. 'The average distance in a random graph with given expected degrees'. *Internet Mathematics* 1.1 (2003), pp. 91–113.

[55] Chung, F. and Lu, L. 'The volume of the giant component of a random graph with given expected degrees'. *SIAM Journal on Discrete Mathematics* 20.2 (2006), pp. 395–411.

[56] Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. 'On the Lambert W function'. *Advances in Computational Mathematics* 5.1 (1996), pp. 329–359.

[57] Crane, H. and Xu, M. 'Inference on the history of a randomly growing tree'. 2020. arXiv: 2005.08794.

[58] Curien, N., Duquesne, T., Kortchemski, I., and Manolescu, I. 'Scaling limits and influence of the seed graph in preferential attachment trees'. *Journal de l'École polytechnique — Mathématiques* 2 (2015), pp. 1–34.

[59] Dall, J. and Christensen, M. 'Random geometric graphs'. *Physical Review E* 66.1 (2002).

[60] Deijfen, M., Esker, H. van den, Hofstad, R. van der, and Hooghiemstra, G. 'A preferential attachment model with random initial degrees'. *Arkiv for Matematik* 47.1 (2007), pp. 41–72.

[61] Deijfen, M., Hofstad, R. van der, and Hooghiemstra, G. 'Scale-free percolation'. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 49.3 (2011), pp. 817–838.

[62] Dekel, Y., Gurel-Gurevich, O., and Peres, Y. 'Finding hidden cliques in linear time with high probability'. *Combinatorics, Probability and Computing* 23.01 (2014), pp. 29–49.

[63] Deshpande, Y. and Montanari, A. 'Finding hidden cliques of size $\sqrt{N/\mathrm{e}}$ in nearly linear time'. *Journal Foundations of Computational Mathematics* 15.4 (2015), pp. 1069–1128.

[64] Devroye, L. and Fraiman, N. 'Connectivity of inhomogeneous random graphs'. *Random Structures & Algorithms* 45.3 (2014), pp. 408–420.

[65] Díaz, J., Mitsche, D., Perarnau, G., and Pérez-Giménez, X. 'On the relation between graph distance and Euclidean distance in random geometric graphs'. *Advances in Applied Probability* 48.3 (2016), pp. 848–864.

[66] Doležal, M., Hladký, J., and Máthé, A. 'Cliques in dense inhomogeneous random graphs'. *Random Structures & Algorithms* 51.2 (2017), pp. 275–314.

[67] Donoho, D. and Jin, J. 'Higher criticism for detecting sparse heterogeneous mixtures'. *The Annals of Statistics* 32.3 (2004), pp. 962–994.

[68] Dorogovtsev, S. N., Mendes, J. F., and Samukhin, A. N. 'Structure of growing networks with preferential linking'. *Physical Review Letters* 85.21 (2000), pp. 4633–4636.

[69] Ellis, R. B., Martin, J. L., and Yan, C. 'Random geometric graph diameter in the unit ball'. *Algorithmica* 47.4 (2007), pp. 421–438.

[70] Erdős, P. and Rényi, A. 'On random graphs'. *Publicationes Mathematicae* 6 (1959), pp. 290–297.

[71] Faloutsos, M., Faloutsos, P., and Faioutsos, C. 'On power-law relationships of the internet topology'. *Computer Communication Review* 29.4 (1999), pp. 251–261.

[72] Farkas, I., Jeong, H., Vicsek, T., Barabási, A. L., and Oltvai, Z. N. 'The topology of the transcription regulatory network in the yeast, Saccharomyces cerevisiae'. *Physica A: Statistical Mechanics and its Applications* 318.3-4 (2003), pp. 601–612.

[73] Feige, U., Goldwasser, S., Lovasz, L., Safra, S., and Szegedy, M. 'Approximating clique is almost NP-complete'. *Proceedings 32nd Annual Symposium of Foundations of Computer Science*. IEEE, 1991, pp. 2–12.

[74] Feige, U. and Krauthgamer, R. 'Finding and certifying a large hidden clique in a semirandom graph'. *Random Structures & Algorithms* 16.2 (2000), pp. 195–208.

[75] Feily, M., Shahrestani, A., and Ramadass, S. 'A survey of botnet and botnet detection'. *Proceedings of the 3rd International Conference on Emerging Security Information, Systems and Technologies*. 2009, pp. 268–273.

[76] Fortunato, S. 'Community detection in graphs'. *Physics Reports* 486.3 (2010), pp. 75–174.

[77] Fountoulakis, N., Hoorn, P. van der, Müller, T., and Schepers, M. 'Clustering in a hyperbolic model of complex networks'. 2020. arXiv: 2003.05525.

[78] Fountoulakis, N., Kang, R. J., and McDiarmid, C. 'The t-stability number of a random graph'. *Electronic Journal of Combinatorics* 17.1 (2010), pp. 1–29.

[79] Fountoulakis, N., Kang, R. J., and McDiarmid, C. 'Largest sparse subgraphs of random graphs'. *European Journal of Combinatorics* 35 (2014), pp. 232–244.

[80] Fraiman, N. and Mitsche, D. 'The diameter of inhomogeneous random graphs'. *Random Structures & Algorithms* 53.2 (2018), pp. 308–326.

[81] Friedrich, T. and Krohmer, A. 'On the diameter of hyperbolic random graphs'. *ICALP 2015: Automata, Languages, and Programming*. Vol. 9135. Springer Verlag, 2015, pp. 614–625.

[82] Friedrich, T. and Krohmer, A. 'Parameterized clique on inhomogeneous random graphs'. *Discrete Applied Mathematics* 184 (2015), pp. 130–138.

[83] Friedrich, T., Sauerwald, T., and Stauffer, A. 'Diameter and broadcast time of random geometric graphs in arbitrary dimensions'. *Algorithmica* 67.1 (2013), pp. 65–88.

[84] Gao, C. and Lafferty, J. 'Testing network structure using relations between small subgraph probabilities'. 2017. arXiv: 1704.06742.

[85] Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. 'Community detection in degree-corrected block models'. *The Annals of Statistics* 46.5 (2018), pp. 2153–2185.

[86] Gao, F. 'Modeling and interference of the internet movie database'. Master Thesis. Eindhoven University of Technology, 2011.

[87] Gao, F. and Vaart, A. van der. 'On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees'. *Stochastic Processes and their Applications* 127.11 (2017), pp. 3754–3775.

[88] García, S., Grill, M., Stiborek, J., and Zunino, A. 'An empirical comparison of botnet detection methods'. *Computers and Security* 45 (2014), pp. 100–123.

[89] García, S., Zunino, A., and Campo, M. 'Survey on network-based botnet detection methods'. *Security and Communication Networks* 7.5 (2014), pp. 878–903.

[90] Gilbert, E. N. 'Random graphs'. *The Annals of Mathematical Statistics* 30.4 (1959), pp. 1141–1144.

[91] Gilbert, E. N. 'Random plane networks'. *Journal of the Society for Industrial and Applied Mathematics* 9.4 (1961), pp. 533–543.

[92] Girvan, M. and Newman, M. E. J. 'Community structure in social and biological networks'. *Proceedings of the National Academy of Sciences of the United States of America* 99.12 (2002), pp. 7821–6.

[93] Grimmett, G. R. and McDiarmid, C. J. H. 'On colouring random graphs'. *Mathematical Proceedings of the Cambridge Philosophical Society* 77.02 (1975), p. 313.

[94] Gugelmann, L., Panagiotou, K., and Peter, U. 'Random hyperbolic graphs: degree sequence and clustering'. *ICALP 2012: Automata, Languages, and Programming*. Vol. 7392. Springer Berlin Heidelberg, 2012, pp. 573–585.

[95] Gulikers, L., Lelarge, M., and Massoulié, L. 'A spectral method for community detection in moderately sparse degree-corrected stochastic block models'. *Advances in Applied Probability* 49.3 (2017), pp. 686–721.

[96] Gulikers, L., Lelarge, M., and Massoulié, L. 'An impossibility result for reconstruction in a degree-corrected planted-partition model'. *The Annals of Applied Probability* 28.5 (2018), pp. 3002–3027.

[97] Hagerup, T. and Rüb, C. 'A guided tour of Chernoff bounds'. *Information Processing Letters* 33.6 (1990), pp. 305–308.

[98] Hajek, B., Wu, Y., and Xu, J. 'Computational lower bounds for community detection on random graphs'. *Proceedings of the 28th Conference on Learning Theory*. Vol. 40. 2015, pp. 899–928.

[99] Hall, P. and Jin, J. 'Innovated higher criticism for detecting sparse signals in correlated noise'. *Annals of Statistics* 38.3 (2010), pp. 1686–1732.

[100] Hammersley, J. M. 'The distribution of distance in a hypersphere'. *The Annals of Mathematical Statistics* 21.3 (1950), pp. 447–452.

[101] Håstad, J. 'Clique is hard to approximate within $n^{1-\varepsilon}$'. *Acta Mathematica* 182.1 (1999), pp. 105–142.

[102] Heard, N. A., Weston, D. J., Platanioti, K., and Hand, D. J. 'Bayesian anomaly detection methods for social networks'. *Annals of Applied Statistics* 4.2 (2010), pp. 645–662.

[103] Hoeffding, W. 'A class of statistics with asymptotically normal distribution'. *The Annals of Mathematical Statistics* 19.3 (1948), pp. 293–325.

[104] Hofstad, R. van der. 'Random graphs and complex networks'. Vol. 1. Cambridge University Press, 2017.

[105] Hofstad, R. van der. 'Random graphs and complex networks'. Vol. 2. 2020+.

[106] Holland, P. W., Laskey, K. B., and Leinhardt, S. 'Stochastic blockmodels: First steps'. *Social Networks* 5.2 (1983), pp. 109–137.

[107] Hollander, F. den. 'Probability theory: the coupling method'. 2012.

[108] Ingster, Y. I. 'Some problems of hypothesis testing leading to infinitely divisible distributions'. *Mathematical Methods of Statistics* 6.1 (1997), pp. 47–69.

[109] Janson, S., Łuczak, T., and Norros, I. 'Large cliques in a power-law random graph'. *Journal of Applied Probability* 47.04 (2010), pp. 1124–1135.

[110] Jeong, H., Tombor, B., Albert, R., Oltval, Z. N., and Barabásl, A. L. 'The large-scale organization of metabolic networks'. *Nature* 407.6804 (2000), pp. 651–654.

[111] Jin, J., Ke, Z. T., and Luo, S. 'Optimal adaptivity of signed-polygon statistics for network testing'. 2019. arXiv: 1904.09532.

[112] Jin, J., Ke, Z., and Luo, S. 'Network global testing by counting graphlets'. *Proceedings of the 35th International Conference on Machine Learning*. 2018, pp. 2333–2341.

[113] Kabatiansky, G. A. and Levenshtein, V. I. 'On bounds for packings on a sphere and in space'. *Problems of Information Transmission* 14.1 (1978), pp. 1–17.

[114] Kang, R. J. and McDiarmid, C. 'The t-improper chromatic number of random graphs'. *Combinatorics Probability and Computing* 19.1 (2010), pp. 87–98.

[115] Karp, R. M. 'Reducibility among combinatorial problems'. *Complexity of Computer Computations*. Ed. by R. E. Miller, J. W. Thatcher, and J. D. Bohlinger. Springer, 1972, pp. 85–103.

[116] Karrer, B. and Newman, M. E. J. 'Stochastic blockmodels and community structure in networks'. *Physical Review E* 83.1 (2011), p. 016107.

[117] Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., and Boguñá, M. 'Hyperbolic geometry of complex networks'. *Physical Review E* 82.3 (2010).

[118] Lehmann, E. L. and Romano, J. P. 'Testing statistical hypotheses'. 3rd edition. Springer-Verlag New York, 2005.

[119] Lovász, L. 'Large networks and graph limits'. American Mathematical Society, 2012.

[120] Lovász, L. and Szegedy, B. 'Limits of dense graph sequences'. *Journal of Combinatorial Theory, Series B* 96.6 (2006), pp. 933–957.

[121] Luce, R. D. 'Connectivity and generalized cliques in sociometric group structure'. *Psychometrika* 15.2 (1950), pp. 169–190.

[122] Luce, R. D. and Perry, A. D. 'A method of matrix analysis of group structure'. *Psychometrika* 14.2 (1949), pp. 95–116.

[123] Marchand, D. C. and Manolescu, I. 'Influence of the seed in affine preferential attachment trees'. *Bernoulli* 26.3 (2020), pp. 1665–1705.

[124] Massoulié, L. 'Community detection thresholds and the weak Ramanujan property'. *Proceedings of the Annual ACM Symposium on Theory of Computing*. 2014, pp. 694–703.

[125] Matula, D. W. 'The employee party problem'. *Notices Of The American Mathematical Society* 19.2 (1972), pp. 89–156.

[126] Matula, D. W. 'The largest clique size in a random graph'. *Tech Report CS 7608, Department of Computer Science and Engineering, Southern Methodist University* (1976).

[127] McDiarmid, C. 'Colouring random graphs'. *Annals of Operations Research* 1.3 (1984), pp. 183–200.

[128] McKinley, G. 'Superlogarithmic cliques in dense inhomogeneous random graphs'. *SIAM Journal on Discrete Mathematics* 33.3 (2019), pp. 1772–1800.

[129] Mesnards, N. G. des, Hunter, D. S., Hjouji, Z. el, and Zaman, T. 'Detecting bots and assessing their impact in social networks'. 2018. arXiv: 1810.12398.

[130] Middendorf, M., Ziv, E., and Wiggins, C. H. 'Inferring network mechanisms: The Drosophila melanogaster protein interaction network'. *Proceedings of the National Academy of Sciences of the United States of America* 102.9 (2005), pp. 3192–3197.

[131] Mittelmann, H. D. and Vallentin, F. 'High accuracy semidefinite programming bounds for kissing numbers'. *Experimental Mathematics* 19.2 (2010), pp. 174–178.

[132] Mitzenmacher, M. and Upfal, E. 'Probability and computing'. 2nd edition. Cambridge University Press, 2017.

[133] Mokken, R. J. 'Cliques, clubs and clans'. *Quality & Quantity* 13.2 (1979), pp. 161–173.

[134] Mongiovì, M., Bogdanov, P., Ranca, R., Papalexakisy, E. E., Faloutsos, C., and Singh, A. K. 'NetSpot: Spotting significant anomalous regions on dynamic networks'. *Proceedings of the 2013 SIAM International Conference on Data Mining*. 2013, pp. 28–36.

[135] Mossel, E., Neeman, J., and Sly, A. 'Reconstruction and estimation in the planted partition model'. *Probability Theory and Related Fields* 162.3-4 (2015), pp. 431–461.

[136] Mossel, E., Neeman, J., and Sly, A. 'A proof of the block model threshold conjecture'. *Combinatorica* 38.3 (2018), pp. 665–708.

[137] Müller, T. 'Two-point concentration in random geometric graphs'. *Combinatorica* 28.5 (2008), pp. 529–545.

[138] Müller, T. and Staps, M. 'The diameter of KPKVB random graphs'. *Advances in Applied Probability* 51.2 (2019), pp. 358–377.

[139] Musin, O. R. 'The problem of the twenty-five spheres'. *Russian Mathematical Surveys* 58.4 (2003), pp. 794–795.

[140] Muthukrishnan, S. and Pandurangan, G. 'The bin-covering technique for thresholding random geometric graph properties'. *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*. 2005, pp. 989–998.

[141] Newman, M. E. J. 'The structure of scientific collaboration networks'. *Proceedings of the National Academy of Sciences* 98.2 (2001), pp. 404–409.

[142] Newman, M. E. J. 'Modularity and community structure in networks'. *Proceedings of the National Academy of Sciences* 103.23 (2006), pp. 8577–8582.

[143] Newman, M. E. J. and Girvan, M. 'Finding and evaluating community structure in networks'. *Physical Review E* 69.2 (2004), p. 026113.

[144] Neyman, J. and Pearson, E. S. 'On the problem of the most efficient tests of statistical hypotheses'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 231.694-706 (1933), pp. 289–337.

[145] Neyman, J. and Pearson, E. S. 'The testing of statistical hypotheses in relation to probabilities a priori'. *Mathematical Proceedings of the Cambridge Philosophical Society* 29.4 (1933), pp. 492–510.

[146] Norros, I. and Reittu, H. 'On a conditionally Poissonian graph process'. *Advances in Applied Probability* 38.01 (2002), pp. 59–75.

[147] Park, Y., Priebe, C. E., and Youssef, A. 'Anomaly detection in time series of graphs using fusion of graph invariants'. *IEEE Journal on Selected Topics in Signal Processing* 7.1 (2013), pp. 67–75.

[148] Pastukhov, G., Veremyev, A., Boginski, V., and Prokopyev, O. A. 'On maximum degree-based γ-quasi-clique problem: Complexity and exact approaches'. *Networks* 71.2 (2018), pp. 136–152.

[149] Pattillo, J., Veremyev, A., Butenko, S., and Boginski, V. 'On the maximum quasi-clique problem'. *Discrete Applied Mathematics* 161.1-2 (2013), pp. 244–257.

[150] Pattillo, J., Youssef, N., and Butenko, S. 'On clique relaxation models in network analysis'. *European Journal of Operational Research* 226.1 (2013), pp. 9–18.

[151] Penrose, M. D. 'Random geometric graphs'. Oxford University Press, 2003.

[152] Pensky, M. and Zhang, T. 'Spectral clustering in the dynamic stochastic block model'. *Electronic Journal of Statistics* 13.1 (2019), pp. 678–709.

[153] Resnick, S. I. and Samorodnitsky, G. 'Asymptotic normality of degree counts in a preferential attachment model'. *Advances in Applied Probability* 48.A (2016), pp. 283–299.

[154] Shah, D. and Zaman, T. 'Rumors in a network: Who's the culprit?' *IEEE Transactions on Information Theory* 57.8 (2011), pp. 5163–5181.

[155] Spencer, J. H. and Florescu, L. 'Asymptopia'. American Mathematical Society, 2014.

[156] Stegehuis, C., Hofstad, R. van der, and Leeuwaarden, J. S. H. van. 'Scale-free network clustering in hyperbolic and other random graphs'. *Journal of Physics A: Mathematical and Theoretical* 52.295101 (2019).

[157] Tsybakov, A. B. 'Introduction to nonparametric estimation'. Springer-Verlag New York, 2009.

[158] Veremyev, A., Prokopyev, O. A., Butenko, S., and Pasiliao, E. L. 'Exact MIP-based approaches for finding maximum quasi-cliques and dense subgraphs'. *Computational Optimization and Applications* 64.1 (2016), pp. 177–214.

[159] Wang, D., Yu, Y., and Rinaldo, A. 'Optimal change point detection and localization in sparse dynamic networks'. 2018. arXiv: 1809.09602.

[160] Wang, H., Tang, M., Park, Y., and Priebe, C. E. 'Locality statistics for anomaly detection in time series of graphs'. *IEEE Transactions on Signal Processing* 62.3 (2014), pp. 703–717.

[161] Watts, D. J. 'Small Worlds: The Dynamics of Networks between Order and Randomness'. Princeton University Press, 1999.

[162] Watts, D. J. and Strogatz, S. H. 'Collective dynamics of 'small-world' networks'. *Nature* 393 (1998), pp. 440–442.

[163] Zeidanloo, H. R., Zadeh, M. J., Shooshtari, Amoli, P. V., Safari, M., and Zamani, M. 'A taxonomy of botnet detection techniques'. *Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology*. Vol. 2. 2010, pp. 158–162.

[164] Zhao, Z., Chen, L., and Lin, L. 'Change-point detection in dynamic networks via graphon estimation'. 2019. arXiv: 1908.01823.

Summary

The work in this thesis is centered around the analysis of community detection methods for inhomogeneous networks, as well the detection of other types of anomalies such as testing for the presence of a botnet. We study how existing community detection methods can be extended to a setting of inhomogeneous random graphs. This led to new insights on the properties of the random graphs we study, and we show how community detection methods can be extended to the inhomogeneous setting in an optimal manner. The insights from this project also sparked the interest to consider a related project about the detection of botnets. Lastly, we consider dynamically growing networks using the preferential attachment model. Below I will elaborate more on each of these projects, and the publications that originated from them.

One of the main questions we answer in this thesis is: "What is the smallest community that can theoretically be detected in an already inhomogeneous graph?". The initial work on this project resulted in novel insights about inhomogeneous graphs and, in particular, about the behavior of cliques in these graphs. Remarkably, the size of the largest clique is almost always the same, even in rather inhomogeneous random graphs. We also extended these results to quasi-cliques and show that also the quasi-clique is highly concentrated in dense inhomogeneous random graphs. We were also able answer our initial question and characterized the smallest community that can theoretically be detected in an inhomogeneous graph. We have done this by proposing and analyzing a scan test and showing that this scan test is optimal in the sense that it is not possible to detect any community that cannot also be detected by our scan test.

Communities are typically modeled as more densely connected subgraphs within a graph, but one is sometimes also interested in subgraphs that are different in other ways. For example, one could be interested in an anomaly such as a botnet that tries to mask its presence by not making too many connections. However, the connectivity structure or underlying geometry of a botnet is often still rather different, and this can be exploited to detect the presence of such a botnet. We formalized this idea and introduced two tests that can both detect such a botnet. Furthermore, we also show

that these tests are optimal in an asymptotic sense.

Many networks are dynamic and change over time, with some rules concerning the evolution or growth of the graph. For this we consider the preferential attachment model and study what happens when the attachment function changes after some time. In particular, we investigated when it is possible to detect that the attachment function has changed. We show that this is indeed possible, even when there are only few vertices with using the alternative attachment function.

About the author

Kay Bogerd was born in 1989 in Amsterdam, The Netherlands. He completed his secondary education at RSG Broklede in Breukelen in 2008, and then started his studies in mathematics and computer science at Utrecht University. After obtaining a bachelor's degree for both in 2012, he continued his studies at the same university by pursuing a master's degree in mathematics, with a specialization in probability and statistics. He obtained his master's degree in 2015.

In 2016 he started a PhD project at Eindhoven University of Technology under the supervision of Remco van der Hofstad and Rui Castro. His PhD research focused on methods to detect communities and anomalies in networks. The results of this research are presented in this dissertation and provided the basis for several scientific publications.

Kay will defend his PhD thesis at Eindhoven University of Technology on February 17, 2021.