On the Consequences of Power-Law Behavior in Partial Correlation Network Models

Matteo Barigozzi[‡] Christian Brownlees[†] Gábor Lugosi[†,◇]

January 2, 2017

Abstract

We study the cross-sectional dependence properties of a partial correlation network model with sparse power-law structure. We show that when the degree distribution of the network is power-law, the system exhibits a high degree of collinearity. More precisely, the largest eigenvalues of the inverse covariance matrix converge to an affine function of the degrees of the most interconnected vertices in the network. The result implies that the largest eigenvalues of the inverse covariance matrix are approximately power-law distributed, and that, as the system dimension increases, the eigenvalues diverge. As an empirical illustration we analyse a panel of stock returns of a large set of companies listed in the S&P500 and show that the covariance matrix of returns exhibits empirical features that are consistent with our power-law model.

Keywords: Partial Correlation Networks, Random Graphs, Power-Law

JEL: C39, C50, C55

[‡] Department of Statistics, London School of Economics and Political Science,
e-mail: m.barigozzi@lse.ac.uk;
[†] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE;
e-mail: christian.brownlees@upf.edu; gabor.lugosi@upf.edu.
[◇] ICREA, Pg. Llus Companys 23, 08010 Barcelona, Spain;
Christian Brownlees and Gábor Lugosi acknowledge financial support from Spanish Ministry of Science and Technology (Grant MTM2015-67304-P) and from Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075).

1 Introduction

The recent financial crises in the United States and Europe have boosted the interest in network analysis in economics and finance. These crises have forcefully shown that when the degree of interdependence in the system is excessive, a large fraction of firms may experience distress simultaneously, and the magnitude of the aggregate distress may be large relative to the entire system.

In the aftermath of these crises, several authors have proposed network estimation techniques for large panels of time series. These methods are typically applied to carry out inference on the degree of interdependence among firms on the basis of market data, such as stock prices. Contributions in this area include the work of Billio, Getmansky, Lo, and Pellizzon (2012), Diebold and Yılmaz (2014), Hautsch, Schaumburg, and Schienle (2014a,b) and Barigozzi and Brownlees (2016). An early influential contribution that analyzed the degree interconnectedness in the financial system in the US and Europe that predates the crisis is Hartmann, Straetmans, and de Vries (2007). This literature is often motivated by the claim that an excessive degree of interconnectedness poses threats to the stability of the entire system. However, the notion of excessive degree of interconnectedness is typically not explicitly formalized and it often remains an elusive concept.

The objective of this paper is to introduce a tractable model that allows one to study the consequences of high interconnectedness in large partial correlation networks. We achieve this by bridging together graphical models (Dempster, 1972; Lauritzen, 1996) with random graphs (Erdős and Rényi, 1960; Chung and Lu, 2006; van der Hofstad, 2015). We introduce a model for a random vector that induces a sparse partial correlation network structure among the variables in the system. A key element of the model is that the underlying network structure of the vector is determined by a random graph. In particular, in this work we focus on a sparse power-law graph, that is, a random graph in which the total number of linkages is proportional to the total number of vertices, and the degree distribution of the vertices follows a power-law. This class of random

graphs is often documented to replicate well the empirical characteristics of many real-world networks (Chung and Lu, 2006; van der Hofstad, 2015). We name this model the *power-law partial correlation network* model.

The main contribution of this work is to derive a number of properties of the power-law partial correlation network model that give insights on the consequences of an underlying power-law behavior. The key result of the paper concerns the spectrum of the covariance matrix of the model. We show that when the dimension of the system is large, the largest eigenvalues of the inverse covariance matrix (also known as the concentration matrix) converge to a positive affine transformation of the degrees of the most interconnected vertices. The result implies that when the underlying network has a power-law behavior the system can exhibit an excessive degree of collinearity. It is important to emphasize that the correlation structure induced by our model differs from, and in a way is complementary to, what is induced by a typical factor structure (Chamberlain and Rothschild, 1983).

As an empirical illustration, we analyze the covariance matrix of a large panel of stock returns of companies listed in the S&P500 index. One of the implications of our model is that the largest eigenvalues of the concentration matrix are approximately power-law distributed. This motivates us to fit a power-law distribution to the largest eigenvalues of the concentration matrix to estimate the power-law tail parameter. The highlight of this approach is that it allows us to carry out inference on the structure of interdependence in the system without having to estimate the underlying sparse graphical structure of the data (typically done using lasso-type techniques). Estimation results deliver a power-law coefficient tail parameter lower than three, which signals, from the standpoint of our model, the presence of highly influential assets in the S&P 500.

This paper is related to the literature on network estimation in econometrics and statistics which includes, among others, the work of Billio *et al.* (2012), Diebold and Yılmaz (2014), Hautsch *et al.* (2014a,b) and Barigozzi and Brownlees (2016). Our main message has analogies to the contribution of Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) that shows that because of network interdependence among firms, firm specific shocks can lead to aggregate system fluctuations.

The paper is structured as follows. Section 2 presents the framework and the main results. Sections 3 and 4 provide illustrations using simulated and real data. Concluding remarks follow in Section 5.

2 Framework

2.1 Partial Correlation Network Model

Consider an n-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ that is assumed to have mean zero and invertible covariance matrix $\boldsymbol{\Sigma} = (\mathbb{E}[Y_i Y_j])_{n \times n}$. The inverse covariance matrix $\boldsymbol{K} = \boldsymbol{\Sigma}^{-1}$, also known as the *concentration* or *precision* matrix, plays an important role in this paper. It is well known that the concentration matrix contains information on the partial correlation dependence structure among the components of the vector $\boldsymbol{Y}$ (Dempster, 1972; Lauritzen, 1996). Partial correlation is a linear measure of conditional dependence. It is defined as the correlation between Y_i and Y_j conditionally on all other variables in the system, that is

$$\rho^{ij} = \mathsf{Cor}(Y_i, Y_j | \{Y_k : k \notin \{i, j\}\}).$$

Partial correlations can be expressed as a function of the entries k_{ij} of the concentration matrix $\boldsymbol{K}$ by

$$\rho^{ij} = -\frac{k_{ij}}{\sqrt{k_{ii}\, k_{jj}}} \, .$$

Thus, the (i, j) entry of the concentration matrix $\boldsymbol{K}$ is zero if and only if Y_i and Y_j are conditionally uncorrelated given all other variables in the system. If the distribution of $\boldsymbol{Y}$ is Gaussian, absence of partial correlation is equivalent to conditional independence.

In this work we propose a model for the concentration matrix $\boldsymbol{K}$ that is a function of an underlying network $\mathcal{N}$. A network $\mathcal{N}$ is defined as a simple undirected graph $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{1, ..., n\}$ is the set of vertices and $\mathcal{E}$ is the set of edges. There are a number of matrices associated with a network that are relevant in this work: the *degree*, the

adjacency, and *Laplacian* matrices. The degree matrix $\boldsymbol{D}$ is defined as a diagonal matrix whose entry $D_{ii} = [\boldsymbol{D}]_{ii}$ is equal to the degree of vertex i, that is, the number of edges adjacent to vertex i. The degree of vertex i is also denoted by D_i. The adjacency matrix $\boldsymbol{A}$ is defined as a matrix whose entry $A_{i,j} = [\boldsymbol{A}]_{i,j}$ is one if i and j are connected by an edge and it is zero otherwise. Notice that the adjacency matrix is symmetric and its diagonal terms are zero. Finally, the Laplacian matrix $\boldsymbol{L}$ of a network is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$.

We now describe a simple model for a multivariate random vector whose partial correlation dependence structure is determined by a network.

Definition 1. (PARTIAL CORRELATION NETWORK MODEL) *We say that the random vector $\boldsymbol{Y}$ is generated by a partial correlation network model based on network $\mathcal{N}$ and scalars $\sigma^2 > 0$ and $\phi \geq 0$ if $\boldsymbol{Y}$ has a multivariate normal distribution with mean zero and concentration matrix*

$$\boldsymbol{K} = \frac{1}{\sigma^2}\boldsymbol{I} + \frac{\phi}{\sigma^2}\boldsymbol{L} \, .$$

The coefficients σ^2 and ϕ are called the variance and the network dependence parameters, respectively.

Our model establishes a link between the concentration matrix of the random vector $\boldsymbol{Y}$ and the Laplacian of the network $\mathcal{N}$. This is a natural model in many situations. The Laplacian matrix often arises in network analysis as its spectrum encodes several key properties of the network. For instance, the number of connected components of the network is equal to the number of zero eigenvalues of the Laplacian. Moreover, indices of centrality, *inter alia* eigenvector centrality, are associated with the largest eigenvalues of the Laplacian as well. We refer to Chung (1997) and Brouwer and Haemers (2011) for good general introductions to spectral graph theory. Importantly, the model ensures that the partial correlation between Y_i and Y_j is zero if and only if there is no link between i and j in the network $\mathcal{N}$. Moreover, the Laplacian is diagonally dominant with positive diagonal entries, which implies that $\boldsymbol{K}$ is a strictly diagonally dominant and therefore positive definite.

We provide a number of properties of the model of Definition 1 that are straightforward to verify. We begin with the regression representation of the model. The i-th variable in the system can be expressed as

$$Y_i = \sum_{j \in N(i)} \frac{\phi}{\phi D_i + 1} Y_j + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2_{\epsilon i}) \,,$$

where $N(i)$ denotes the set of neighbors of vertex i and $\sigma^2_{\epsilon i} = \sigma^2/(1 + \phi D_i)$. Rearranging terms, we get

$$Y_i = \left(\frac{D_i}{D_i + 1/\phi} \right) \frac{1}{D_i} \sum_{j \in N(i)} Y_j + \epsilon_i = \frac{\psi_i}{D_i} \sum_{j \in N(i)} Y_j + \epsilon_i \,,$$

where $\psi_i = D_i/(D_i + 1/\phi)$. The last equation shows that conditionally on the other variables in the system, the realization of the i-th variable can be interpreted as a linear function of the average of the neighbors and a location-specific innovation. We believe that this "graphical model" is a quite natural model of multivariate Gaussian distributions with an underlying network structure.

The partial correlations implied by the model are

$$\rho^{ij} = \frac{A_{i,j}}{\sqrt{(D_i + 1/\phi)(D_j + 1/\phi))}}$$

(recall that $A_{i,j} = \mathbb{1}_{\{i \sim j\}}$ is the indicator of an edge joining vertex i and vertex j). Expansion of this formula with respect to ϕ gives insight to the partial correlation behavior of the model. For ϕ close to zero we have that the partial correlations are approximately constant in the sense that, as $\phi \to 0$,

$$\rho^{ij} = \phi A_{i,j} + O(\phi^2) \,.$$

On the other hand, when ϕ is large, the magnitude of the partial correlation between two variables is a decreasing function of the number of linkages of each variable, that is, as

$\phi \to \infty$,

$$\rho^{ij} = \frac{A_{i,j}}{\sqrt{D_i D_j}} + O(\phi^{-1}) \ .$$

The covariance matrix $\boldsymbol{\Sigma}$ of the system (see Proposition 1 in the Appendix) is

$$\boldsymbol{\Sigma} = \sigma^2(\boldsymbol{I} + \phi \boldsymbol{D})^{-1} + \frac{\sigma^2}{\phi} \sum_{k=1}^{\infty} \boldsymbol{W}_k \ ,$$

where $\boldsymbol{W}_k$ is what we call the *weighted walk matrix* of length k, defined as

$$[\boldsymbol{W}_k]_{i,j} = \sum_{w \in \mathcal{W}_k} \frac{\mathbb{1}_{\{w \text{ goes from } i \text{ to } j\}}}{\prod_{v \in \mathcal{V}(w)} (d(v) + 1/\phi)} \ ,$$

where $\mathcal{W}_k$ denotes the set of walks of length k (i.e., sequences of vertices $v_1, \ldots, v_{k+1}$ such that $v_i \sim v_{i+1}$ for $i = 1, \ldots, k$) in the network, $\mathcal{V}(w)$ denotes the set of vertices in walk w, $d(v)$ denotes the degree of vertex v and $\mathbb{1}$ is the indicator function. Notice that two variables are correlated if and only if there exists a path in the graph that connects them and that the higher the number of walks between two variables the higher their correlation is.

An important characteristics of our model is that it does not have a factor structure in the sense of, for example, Chamberlain and Rothschild (1983), Stock and Watson (2002a,b), and Bai (2003). According to their definitions, if $\boldsymbol{\Sigma}$ has an r-factor structure then its r largest eigenvalues diverge as the cross-sectional dimension increases at a rate $O(n)$, while the remaining $n-r$ stay bounded. However, in the partial-correlation network model it is straightforward to see that the eigenvalues of $\boldsymbol{\Sigma}$ are bounded from above. Indeed, let $\lambda_i^{\boldsymbol{M}}$ denote the i–th eigenvalue of a matrix $\boldsymbol{M}$ in a non-increasing order. Then, we have

$$\lambda_1^{\boldsymbol{\Sigma}} = \frac{1}{\lambda_n^{\boldsymbol{K}}} = \frac{\sigma^2}{1 + \phi \lambda_n^{\boldsymbol{L}}} = \sigma^2,$$

which follows from the fact that the smallest eigenvalue of the Laplacian of a network is zero. It is possible to consider our model as complementary to a factor model. Indeed, in the presence of a factor structure, a partial-correlation network structure can be defined for

the idiosyncratic component (the component not driven by the factors) without violating the standard assumptions of factor models.

2.2 Partial Correlation Network Model and Random Graphs

In this work we model the underlying network $\mathcal{N}$ of the partial correlation network model as a random graph. In particular, we work with an inhomogeneous random graph model known as the *Chung-Lu model* (Chung and Lu, 2006; Bollobás, Janson, and Riordan, 2007). The Chung-Lu model is a generalization of the classical Erdős-Rényi random graph (Erdős and Rényi, 1960) that allows one to model general degree distributions. In particular, the Chung-Lu model offers a flexible and versatile tool to model and analyze random graphs with a power-law degree distributions.

Definition 2. (CHUNG-LU RANDOM GRAPH) *Let $\boldsymbol{w} \in \mathbb{R}^n$ be a nonnegative weight vector such that*

$$w_M^2 = \max_i w_i^2 \leq \sum_{l=1}^n w_l \ . \tag{1}$$

A Chung-Lu random graph is a simple undirected graph $(\mathcal{V}, \mathcal{E})$ where the existence of an edge between vertex i and j $(i \neq j)$ is determined by an independent Bernoulli trial with probability

$$p_{i,j} = \frac{w_i w_j}{\sum_{l=1}^n w_l} \ .$$

The weights w_i are closely related to the expected degrees in the random graph, since

$$\begin{aligned} \mathbb{E}D_i &= \mathbb{E}\left(\sum_{j=1}^n A_{i,j}\right) = \sum_{j=1}^n p_{i,j} \\ &= \sum_{j:j\neq i}^n \frac{w_i w_j}{\sum_{l=1}^n w_l} = w_i \frac{\sum_{j:j\neq i}^n w_j}{\sum_{j=1}^n w_j} \approx w_i \ . \end{aligned}$$

Note, in particular, that by (1),

$$w_i \geq \mathbb{E}D_i \geq w_i \left(1 - \frac{w_i}{\sum_{j=1}^n w_j}\right) = w_i \left(1 - \frac{w_i}{n\overline{w}}\right) \ . \tag{2}$$

Also, $\mathbb{E}D_i \geq \mathbb{E}D_j$ if and only if $w_i \geq w_j$. By appropriately choosing the specification of the weight sequence w_i it is possible to replicate different types of random graph models proposed in the literature. In this paper we work with random graphs with specific sparsity properties. We focus on sparse graphs in which the average weight (i.e., roughly the average expected degree)

$$\overline{w} = \frac{1}{n}\sum_{l=1}^{n} w_i$$

is a positive constant $\overline{w} > 0$. This sparsity assumption allows us to analyze the implications of power-law degree distribution in networks in which the total number of linkages is constrained to be $O(n)$. Thus, our results are driven by the configuration of the linkages rather than by their sheer number.

Note that we recover the classical Erdős-Rényi model by taking $w_i = \overline{w}$ for all $i = 1, \ldots, n$. The Erdős-Rényi model is often acknowledged to be of limited power in modelling networks. In particular, the degree distribution of many networks appearing in practice seems to be much more heavy tailed than what would be implied by an Erdős–Rényi model. In order to overcome the limitations of this class of models, several authors have proposed different formulations of so-called power-law random graphs, that is, random graphs in which the degree distribution follows a power-law. In this work we find it convenient to focus on power-law random graphs derived from the Chung-Lu model. Other well-studied power-law random graphs include the configuration model and preferential attachment models, see van der Hofstad (2015) for a survey.

Definition 3. (POWER-LAW RANDOM GRAPH) *The power-law graph random graph with power-law parameter $\beta > 2$ is defined as a Chung-Lu random graph with weight sequence given by*

$$w_i = c\left(\frac{i + i_0 - 1}{n}\right)^{-1/(\beta-1)} \quad i = 1, \ldots, n\ ,$$

with

$$i_0 = n\left(\frac{c}{w_M}\right)^{\beta-1} \text{ and } c = \overline{w}\left(\frac{\beta-2}{\beta-1}\right)\ ,$$

where $\overline{w}$ is the average weight and $w_1 = w_M$ is the maximum weight.

It is straightforward to show (Chung and Lu, 2006) that the expected number n_k of vertices of degree k is power-law distributed, that is,

$$n_k \propto \frac{\Gamma(k-\beta+1)}{\Gamma(k+1)} \approx k^{-\beta}.$$

Power-law networks with $\beta \in (2,3)$ are particularly interesting as, empirically, this interval typically contains the vast majority of empirical estimates of the tail parameter obtained from many networks.

We define the power-law partial correlation network by associating the partial correlation model with the power-law random graph.

Definition 4. (POWER-LAW PARTIAL CORRELATION NETWORK) *The power-law partial correlation network is a partial correlation network model in which the underlying network $\mathcal{N}$ is a power-law random graph with parameters $\beta > 2$, $\overline{w} > 0$, and $w_M = n^\alpha$ with $\alpha \in (0, 1/2)$.*

One of our main findings is that the behavior of the largest eigenvalues of the concentration matrix is determined by the highest degrees in the network. The result is summarized in the following theorem.

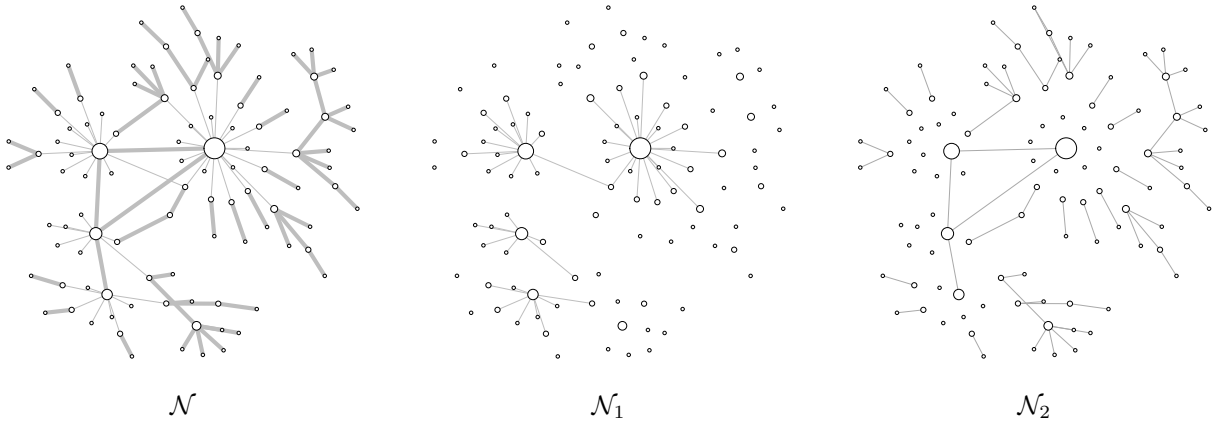
Theorem 1. *Let $\boldsymbol{Y}$ be generated by the power-law partial correlation network model with $\beta \in (2,3)$ and $\alpha < 1/(\beta-1)$. Let $\lambda_i^{\boldsymbol{K}}$ denote the i^{th} eigenvalue of the concentration matrix $\boldsymbol{K}$ ordered in descending order. Then for every $\epsilon > 0$ there exists constants $C, \delta > 0$ such that, with probability at least $1 - n\exp\left(-Cn^\delta\right)$, for all $i \leq n^{\min(\alpha, 1-\alpha(\beta-1))-\epsilon}$,*

$$\lambda_i^{\boldsymbol{K}} = \frac{1}{\sigma^2}\left[\phi w_i\left(1 + O(n^{-\delta})\right) + 1\right] \ . \tag{3}$$

Our result implies that the largest eigenvalues of the concentration matrix are closely related to the power-law distribution of the degrees. By a simple concentration argument, w_i may be replaced by either $\mathbb{E}D_i$ or D_i in (3).

A number of comments on the implications of Theorem 1 are in order. First, notice

Figure 1: Network Decomposition.



The figure provides an illustration of the network decomposition strategy we use to prove the main theorem of this work.

that as the system dimension n increases, the smallest eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$ of the model, which are the reciprocals of the largest eigenvalues of $\boldsymbol{K}$, converge to zero. Thus, the power-law partial correlation network is a model in which, in spite of the sparsity of the underlying network, the system experiences high collinearity among the variables. Second, the largest eigenvalues of the concentration matrix $\boldsymbol{K}$ are a noisy proxy (up to a linear transformation) of the largest degrees of the network. This suggests that these eigenvalues can be used to gain insight to the underlying degree distribution and, in particular, to the power-law coefficient β. This is appealing in that it suggests that we can learn the degree of the power-law exponent of the power-law partial correlation network simply by analyzing the spectrum of the concentration matrix and without having to learn the entire underlying network using, for instance, sparse network estimation techniques.

We detail the proof of this result in the following section and we sketch here the main steps of our argument. Central to our proof is a decomposition of the power-law network $\mathcal{N}$ into two networks $\mathcal{N}_1$ and $\mathcal{N}_2$. For illustration purposes, Figure 1 contains an example of our graph decomposition using a simulated power-law network with 100 vertices. Fix some $k < n$ and consider the k vertices $v_1, \ldots, v_k$ with largest expected degree in the network $\mathcal{N}$. (For instance, in Figure 1 the four vertices with highest degree are selected.)

The sub-network $\mathcal{N}_1$ has the same set of vertices as $\mathcal{N}$ and it consists of all edges adjacent to $v_1, \ldots, v_k$, except for all edges between these vertices. (Thus, $\mathcal{N}_1$ is a bipartite graph with bipartition $\{v_1, \ldots, v_k\}$, $\{v_{k+1}, \ldots, v_n\}$.) The network $\mathcal{N}_2$ (also defined on the same vertex set as $\mathcal{N}$) contains all remaining edges of $\mathcal{N}$.

Next we note that the Laplacian of $\mathcal{N}$ is equal to the sum of the Laplacians of $\mathcal{N}_1$ and $\mathcal{N}_2$. Thus, we can obtain bounds for the largest eigenvalues of the Laplacian of $\mathcal{N}$ using the eigenvalues of the Laplacians of $\mathcal{N}_1$ and $\mathcal{N}_2$. As it turns out, the largest eigenvalues of the Laplacian of $\mathcal{N}_1$ are closely related to the largest expected degrees in $\mathcal{N}$. This follows from general bounds for Laplacian eigenvalues in terms of degrees and probabilistic arguments.

Next we show that the largest eigenvalue of the Laplacian of $\mathcal{N}_2$ is small relative to the largest eigenvalues of the Laplacian of $\mathcal{N}_1$. Finally, the claim of the theorem follows by applying Weyl's inequality which allows us to establish that the largest eigenvalues of the Laplacian of $\mathcal{N}$ are close to the degrees of the largest degree vertices in the network, with high probability.

2.3 Proof of Theorem 1

Introduce $\ell = n^\gamma$ for some $\gamma < \min(\alpha, 1-\alpha(\beta-1))$. We prove that the ℓ largest eigenvalues of $\boldsymbol{K}$ satisfy the relation stated in the theorem.

We begin by noting that, with $w_1 = n^\alpha$, the weights of the power-law random graph become

$$w_i = cn^{1/(\beta-1)} \left(c^{\beta-1} n^{1-\alpha(\beta-1)} + i - 1\right)^{-1/(\beta-1)} .$$

When $i < c^{\beta-1} n^{1-\alpha(\beta-1)}$, then within the expression in parenthesis, the first term dominates, otherwise the second. In particular, we have

$$w_i \in \begin{cases} [n^\alpha/2^{1/(\beta-1)}, n^\alpha] & \text{if } i < c^{\beta-1} n^{1-\alpha(\beta-1)} \\ \left[c(n/(2i))^{1/(\beta-1)}, c(n/i)^{1/(\beta-1)}\right] & \text{if } i \geq c^{\beta-1} n^{1-\alpha(\beta-1)} . \end{cases} \tag{4}$$

The proof is based on decomposing the power-law network $\mathcal{N}$ in the union of two

non-overlapping networks $\mathcal{N}_1$ and $\mathcal{N}_2$. We say that the networks $\mathcal{N}_1$ and $\mathcal{N}_2$ are non-overlapping if the networks are defined over the same vertex set and their edge sets are disjoint. That is, $\mathcal{N}_1 = (\mathcal{V}, \mathcal{E}_1)$, $\mathcal{N}_2 = (\mathcal{V}, \mathcal{E}_2)$ with $\mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$. The union $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2$ of non-overlapping networks $\mathcal{N}_1$ and $\mathcal{N}_2$ is $\mathcal{N} = (\mathcal{V}, \mathcal{E}_1 \cup \mathcal{E}_2)$. The Laplacian of the union of non-overlapping networks equals the sum of the Laplacians $\boldsymbol{L} = \boldsymbol{L}_1 + \boldsymbol{L}_2$.

As described above, the sub-network $\mathcal{N}_1$ consists of all edges adjacent to the vertices $v_1, \ldots, v_k$ of highest weight, except for those edges that join two vertices v_i, v_j such that $i, j \leq k$. We choose $k = n^\rho$ for some $\rho \in (1 - \alpha(\beta - 1), 1 - \alpha)$. $\mathcal{N}_2$ contains all remaining edges of $\mathcal{N}$ so that $\mathcal{N}_1$ and $\mathcal{N}_2$ are non-overlapping and $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2$.

In order to estimate the largest eigenvalues of the Laplacian $\boldsymbol{L}$ of the power-law network $\mathcal{N}$, we write $\boldsymbol{L}$ as the sum of the Laplacians $\boldsymbol{L}_1$ of $\mathcal{N}_1$ and $\boldsymbol{L}_2$ of the residual network $\mathcal{N}_2$, $\boldsymbol{L} = \boldsymbol{L}_1 + \boldsymbol{L}_2$. It follows from Weyl's inequality that the eigenvalues of $\boldsymbol{L}$ satisfy, for all $i = 1, \ldots, n$,

$$\lambda_i^{\boldsymbol{L}_1} + \lambda_n^{\boldsymbol{L}_2} \leq \lambda_i^{\boldsymbol{L}} \leq \lambda_i^{\boldsymbol{L}_1} + \lambda_1^{\boldsymbol{L}_2} \, . \tag{5}$$

Noting that the smallest eigenvalue of the Laplacian of any equals zero, we have $\lambda_n^{\boldsymbol{L}_2} = 0$, and therefore, in order to prove the theorem we need to study the behavior of $\lambda_i^{\boldsymbol{L}_1}$ and $\lambda_1^{\boldsymbol{L}_2}$.

First we bound from above the largest eigenvalue $\lambda_1^{\boldsymbol{L}_1}$ of the Laplacian $\boldsymbol{L}_1$ of the graph $\mathcal{N}_1$. Our main tool is a well-known bound of Anderson and Morley (1985) (Proposition 3 below) that implies that $\lambda_1^{\boldsymbol{L}_1}$ is at most $\max\left(D_i^{\mathcal{N}_1} + D_j^{\mathcal{N}_1}\right)$, where $D_i^{\mathcal{N}_1}$ denotes the degree of vertex v_i in $\mathcal{N}_1$ and the maximum is taken over all pairs (i, j) such that v_i and v_j are joined by an edge in $\mathcal{N}_1$.

Since $\mathcal{N}_1$ is bipartite, we have

$$\begin{aligned} \lambda_1^{\boldsymbol{L}_1} &\leq \max_{i \leq k} D_i^{\mathcal{N}_1} + \max_{i > k} D_i^{\mathcal{N}_1} \\ &\leq \max_{i \leq k} D_i + \max_{i > k} \sum_{j=1}^{k} A_{i,j} \, . \end{aligned}$$

In order to bound the right-hand side, first note that

$$\max_{i\le k} D_i \le \max_{i\le k} \mathbb{E}D_i + \max_{i\le k}(D_i - \mathbb{E}D_i) \le n^\alpha + \max_{i\le k}(D_i - \mathbb{E}D_i) \ .$$

By a simple Chernoff bound (see, e.g., Proposition 2 in the Appendix),

$$\mathbb{P}\left(\max_{i\le k}(D_i - \mathbb{E}D_i) > n^{3\alpha/4}\right) \le k\mathbb{P}\left(D_i - \mathbb{E}D_i > n^{3\alpha/4}\right) \le k\exp\left(-n^{\alpha/2}\right) \ .$$

On the other hand, note that for $i > k$, using (4),

$$\mathbb{E}\sum_{j=1}^{k} A_{i,j} \le \frac{kw_1w_i}{n\overline{w}} \le n^\alpha \frac{c^{1/(\beta-1)}n^{\rho+(1-\rho)/(\beta-1)-1}}{\overline{w}} \ .$$

(Note that $\beta > 2$ implies $\rho + (1-\rho)/(\beta-1) < 1$ and so $\mathbb{E}\sum_{j=1}^{k} A_{i,j} = o(n^\alpha)$.) Thus, again by Proposition 2,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{i>k}\sum_{j=1}^{k} A_{i,j} > 2n^\alpha \frac{c^{1/(\beta-1)}n^{\rho+(1-\rho)/(\beta-1)-1}}{\overline{w}}\right\} \\
&\le \ \mathbb{P}\left\{\max_{i>k}\left(\sum_{j=1}^{k} A_{i,j} - \mathbb{E}\sum_{j=1}^{k} A_{i,j}\right) > n^\alpha \frac{c^{1/(\beta-1)}n^{\rho+(1-\rho)/(\beta-1)-1}}{\overline{w}}\right\} \\
&\le \ n\exp\left(-n^\alpha \frac{c^{1/(\beta-1)}n^{\rho+(1-\rho)/(\beta-1)-1}}{\overline{w}}\right) \ .
\end{aligned}$$

Hence, since $\rho > 1-\alpha(\beta-1)$ implies $\alpha+\rho+(1-\rho)/(\beta-1)-1 > 0$, we have that, there exist positive constants C_1, ϵ_1 such that, with probability at least $1 - 2n\exp(-C_1 n^{\epsilon_1})$,

$$\lambda_1^{\boldsymbol{L}_1} \le n^\alpha\left(1 + C_1 n^{-\epsilon_1}\right) \ . \tag{6}$$

Recall that $\ell = n^\gamma$ for some $\gamma < \min(\alpha, 1-\alpha(\beta-1))$ We derive a lower bound for the ℓ-th largest eigenvalue $\lambda_\ell^{\boldsymbol{L}_1}$ of the Laplacian of $\mathcal{N}_1$. To this end, we use Proposition 4 that implies that

$$\lambda_\ell^{\boldsymbol{L}_1} \ge D_\ell^{\mathcal{N}_1} - \ell \ . \tag{7}$$

Since $\ell < k$, $D_\ell^{\mathcal{N}_1} = \sum_{i=k+1}^{n} A_{i,\ell}$. Thus, using (2) and (4)

$$\begin{aligned}
\mathbb{E}D_\ell^{\mathcal{N}_1} &= \mathbb{E}D_\ell - w_\ell \sum_{i=1}^{k} \frac{w_i}{n\overline{w}} \geq w_\ell \left(1 - \frac{w_\ell}{n\overline{w}} - k\frac{w_k}{n\overline{w}}\right) \\
&\geq w_\ell \left(1 - \frac{n^{\alpha-1}}{\overline{w}} - \frac{c}{\overline{w}} n^{(\rho(\beta-2)+1)/(\beta-1)-1}\right) . \qquad (8)
\end{aligned}$$

Since $\ell = n^\gamma$ with $\gamma < 1 - \alpha(\beta - 1))$, we have from the definition of the weights that

$$\begin{aligned}
w_\ell &= n^\alpha \left(1 + \frac{1}{c^{\beta-1}} n^{\gamma+\alpha(\beta-1)-1}\right)^{-1/(\beta-1)} \\
&\geq n^\alpha \left(1 + \frac{1}{(\beta-1)c^{\beta-1}} n^{\gamma+\alpha(\beta-1)-1}\right)^{-1} \\
&\quad \text{(by Bernoulli's inequality)} \\
&\geq n^\alpha \left(1 - \frac{1}{(\beta-1)c^{\beta-1}} n^{\gamma+\alpha(\beta-1)-1}\right) . \qquad (9)
\end{aligned}$$

Putting (8) and (9) together, we have that, for some positive constants C_1, ϵ_1,

$$\mathbb{E}D_\ell^{\mathcal{N}_1} \geq n^\alpha \left(1 - C_1 n^{-\epsilon_1}\right) .$$

By Proposition 2, $D_\ell^{\mathcal{N}_1} \geq \mathbb{E}D_\ell^{\mathcal{N}_1} - n^{3\alpha/4}$ with probability at least $1 - \exp(-\Omega(n^{\alpha/2}))$, and thus, combining these bounds with (7) and (6), we find that there exist positive constants C_0, ϵ_0 such that, with probability at least $1 - \exp(-C_0 n^{\epsilon_0})$, for all $i = 1, \ldots, \ell$,

$$n^\alpha \left(1 - C_0 n^{-\epsilon_0}\right) \leq \lambda_i^{\boldsymbol{L}_1} \leq n^\alpha \left(1 + C_0 n^{-\epsilon_0}\right) . \qquad (10)$$

Next, we bound the largest eigenvalue $\lambda_1^{\boldsymbol{L}_2}$ of the Laplacian of the residual network $\mathcal{N}_2$. To this end, once again we apply the Anderson-Morley inequality (Proposition 3 below) that implies that

$$\lambda_1^{\boldsymbol{L}_2} \leq 2 \max_{i=1,\ldots,n} D_i^{\mathcal{N}_2} ,$$

where $D_i^{\mathcal{N}_2}$ is the degree of vertex v_i in $\mathcal{N}_2$. By the definition of $\mathcal{N}_2$, we have that

$$\max_{i=1,\dots,n} D_i^{\mathcal{N}_2} \leq \max_{i=1,\dots,k} \sum_{j=1}^{k} A_{i,j} + \max_{i=k+1,\dots,n} D_i \ .$$

We proceed the same way as before. First we estimate the expected values of the random variables appearing in the bound and then use Proposition 2 and the union bound to conclude. For each $i = 1, \dots, k$,

$$\mathbb{E} \sum_{j=1}^{k} A_{i,j} = \sum_{j=1}^{k} p_{i,j} \leq k \frac{w_1^2}{n\overline{w}} = n^{\alpha} \frac{n^{\alpha+\rho-1}}{\overline{w}} \ .$$

(Recall that $\rho < 1 - \alpha$ and therefore $\mathbb{E} \sum_{j=1}^{k} A_{i,j} = O(n^{\alpha-\delta})$ for some $\delta > 0$.) On the other hand, for all $i > k$, using the fact that $\rho > 1 - \alpha(\beta - 1)$, we have

$$\mathbb{E} D_i \leq w_i \leq c \left(\frac{n}{k}\right)^{1/(\beta-1)} = c n^{(1-\rho)/(\beta-1)} \ ,$$

and therefore $\mathbb{E} D_i = O(n^{\alpha-\delta})$ for a positive δ. Using Proposition 2 and the union bound exactly as before, we obtain that, there exist positive constants C_2, ϵ_2 such that, with probability at least $1 - n \exp(-C_2 n^{\epsilon_2})$,

$$\lambda_1^{\boldsymbol{L}_2} \leq n^{\alpha-\epsilon_2} \ . \tag{11}$$

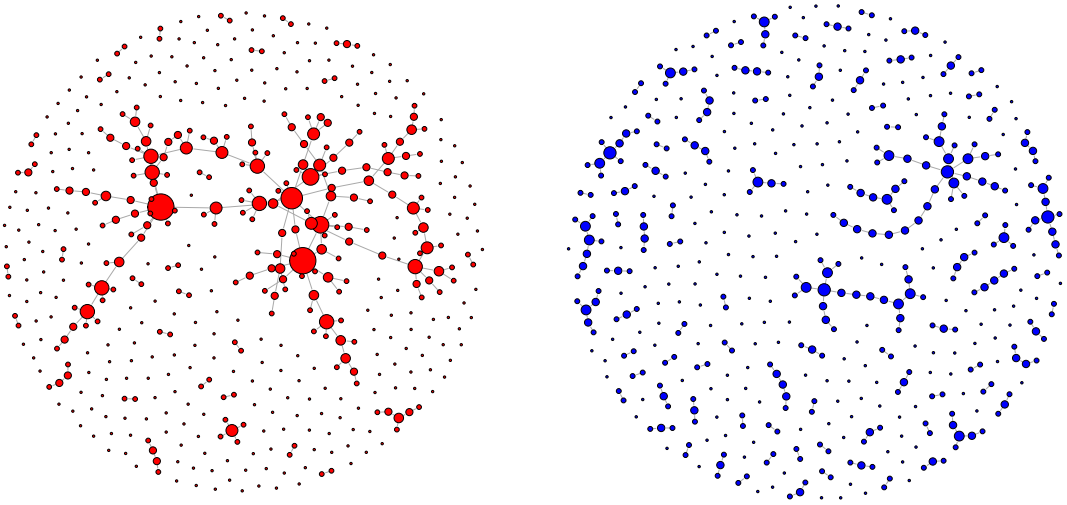
Finally, combining (5) with (10) and (11), we see that, there exist positive constants C_3, ϵ_3 such that, with probability at least $1 - n \exp(-C_3 n^{\epsilon_3})$, for all $i = 1, \dots, \ell$, the eigenvalues of the power-law network $\mathcal{N}$ satisfy

$$n^{\alpha} \left(1 - C_3 n^{-\epsilon_3}\right) \leq \lambda_i^{\boldsymbol{L}_1} \leq n^{\alpha} \left(1 + C_3 n^{-\epsilon_3}\right) \ .$$

This, together with (9), implies that, with probability at least $1 - n \exp(-C_3 n^{\epsilon_3})$, for all $i = 1, \dots, \ell$,

$$\left| \frac{\lambda_i^{\boldsymbol{L}}}{w_i} - 1 \right| \leq C n^{-\delta}$$

Figure 2: RANDOM GRAPH SIMULATION.



Realizations of a simulated power-law (left) and Erdős-Rényi (right) random graphs.

for some positive constants C, δ. The relationship between the eigenvalues of $\boldsymbol{K}$ and $\boldsymbol{L}$

$$\lambda_i^{\boldsymbol{K}} = \frac{1 + \phi \lambda_i^{\boldsymbol{L}}}{\sigma^2}$$

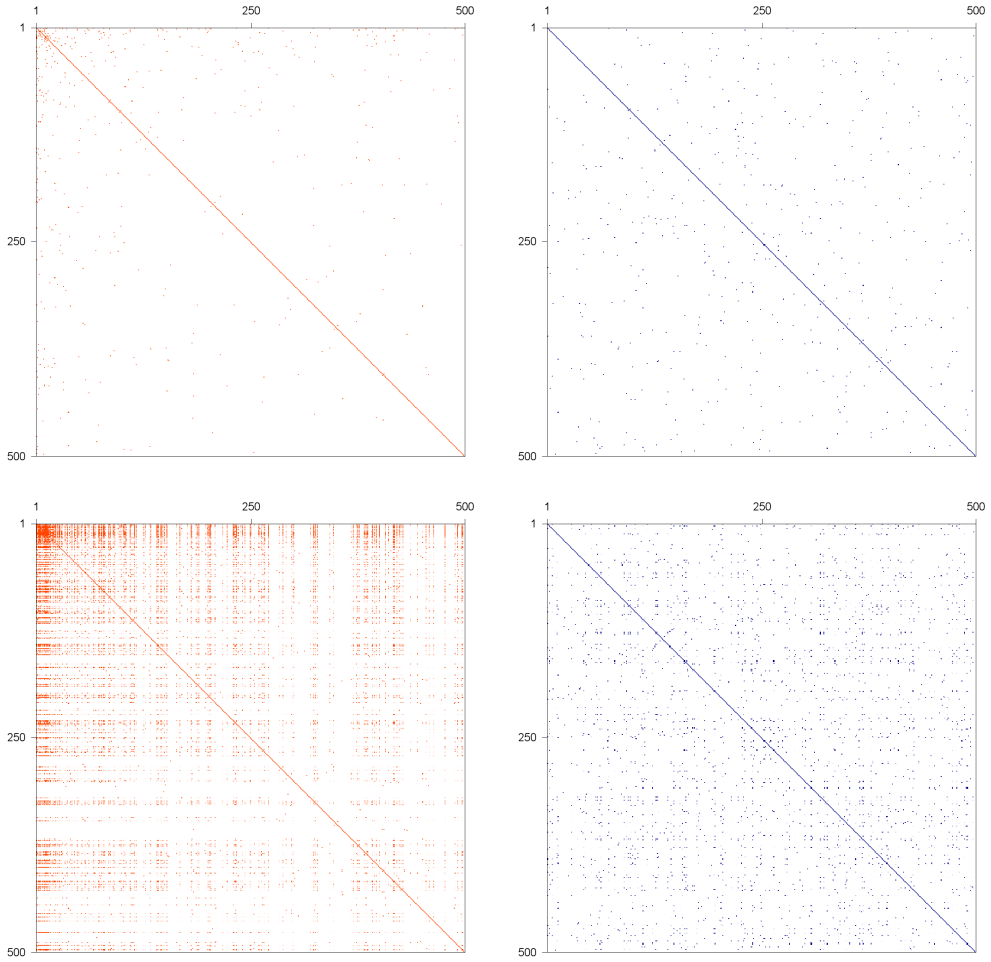
concludes the proof. □

3 Simulation Study

In this section we carry out a simulated illustration of the model introduced in the previous section to highlight some of its key properties, and we perform a Monte Carlo study to verify numerically the conclusion of Theorem 1.

We begin by showing a realization of the power-law partial correlation network model and comparing it with a partial correlation network model based on the Erdős-Rényi random graph. The average expected degree of both networks is $\overline{w} = 1$ and the size of the system is $n = 500$. The variance parameter σ^2 and the network dependence parameter ϕ are both set to one. For the power-law network we also set the power-law tail parameter β to 2.5 and the maximum expected degree w_M to $\lceil 500^{0.45} \rceil = 16$ (i.e., $\alpha = 0.45$).

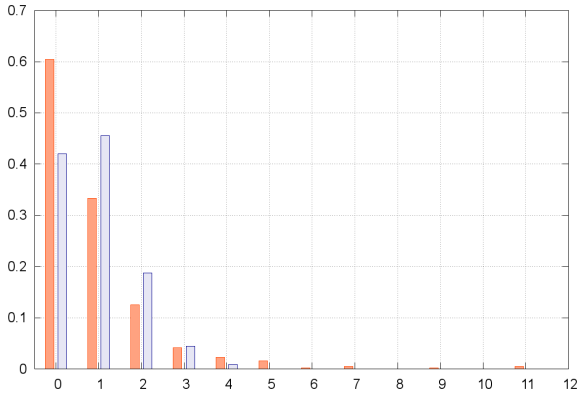
Figure 3: SPARSITY OF THE CONCENTRATION AND COVARIANCE MATRICES.



Concentration matrices in the upper panels and corresponding covariance matrices in the lower panels. Left column: power-law partial correlation model; right column: Erdős-Rényi model.

We display the realizations of the two random graphs in Figure 2. The plot of the two networks shows how the power-law model produces structures containing hubs of highly interconnected vertices, analogously to what is often encountered in the network analysis of economic and financial time series. Figure 3 shows the heat maps of the concentration and covariance matrices associated with the two networks. The power-law model delivers a more interdependent multivariate system, as it can be gauged by the inspection of the covariance matrix. In Figure 4 we report the degree distribution of the two simulated networks where it is evident that the power-law model is associated with heavier tails than the Erdős-Rényi.

Figure 4: DEGREE DISTRIBUTION

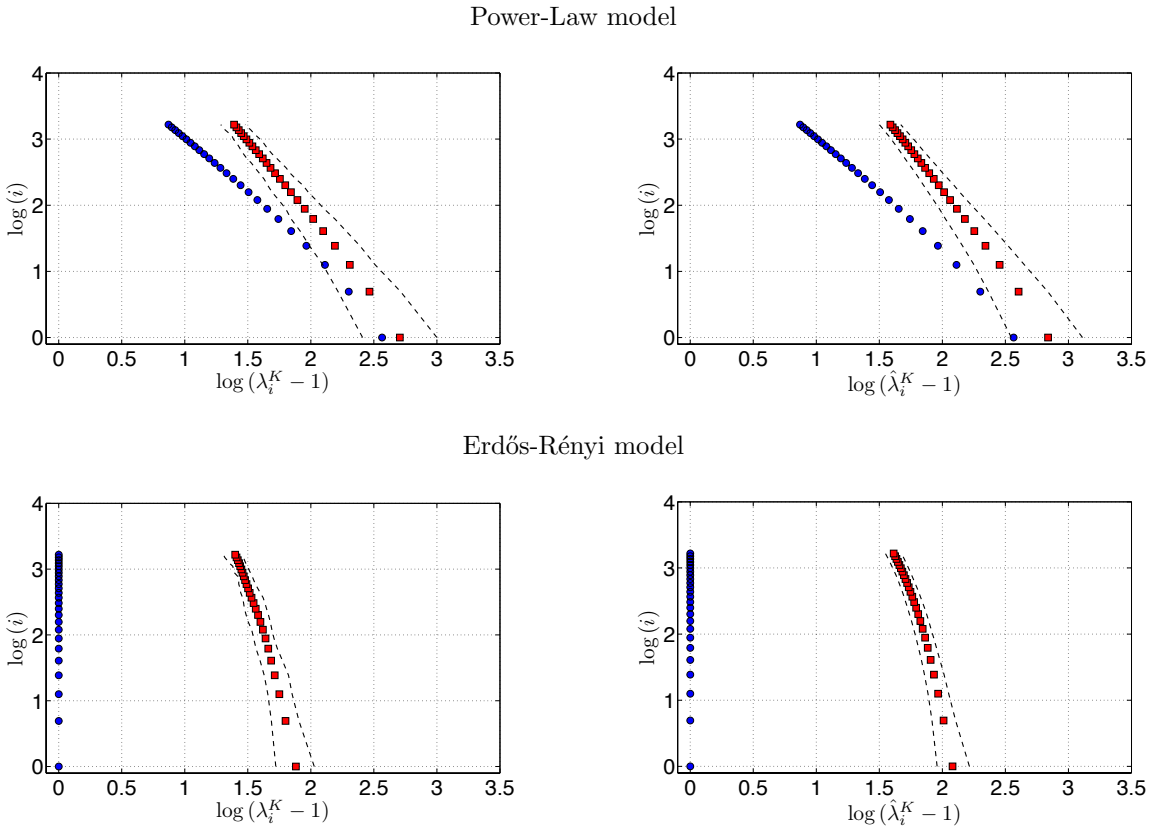


Degree distribution of the power-law (left) and the Erdős-Rényi (right) random graphs.

A Monte Carlo experiment is run to verify the conclusion of Theorem 1. The experiment is designed as follows. In each replication, we simulate the power-law partial correlation network model and draw 5000 random vectors from it. We then compute the eigenvalues of the concentration matrix as well as the eigenvalues of sample concentration matrix obtained from the sample of random vectors. The simulation exercise is replicated 10000 times. We also carry out the same exercise using a partial correlation network model based on the Erdős-Rényi random graph.

We report the results in Figure 5, where we plot the Monte Carlo averages of the largest eigenvalues minus one, that is $\lambda_i^{\boldsymbol{K}} - 1$, together with the largest expected degrees. Recall that in this current setting Theorem 1 implies that the largest eigenvalues of the power-law partial correlation network model concentration matrix are $\lambda_i^{\boldsymbol{K}} = D_i(1 + o_p(1)) + 1$. Notice that we plot adjusted eigenvalues and expected degrees using the so-called log-log rank-size plot, which is a plot of the log expected degree/eigenvalue versus its log rank. As predicted, when plotted on a log-log scale, the expected degrees lie approximately on a straight line with slope equal the negative of the power-law coefficient β. The plot shows that, as predicted by Theorem 1, the largest eigenvalues of the concentration matrix of the power-law partial correlation network model are close to the largest expected degrees of the power-law graph. Moreover, for large sample sizes the impact of estimation uncertainty in the estimation of the concentration matrix is marginal. In the Erdős-Rényi model on

Figure 5: EIGENVALUES BEHAVIOR IN SIMULATED DATA.



Log-log rank-size plot for the Monte Carlo averages of 25 largest eigenvalues $\log(\lambda_i^K - 1)$ (left, squares), estimated eigenvalues $\log(\widehat{\lambda}_i^K - 1)$ (right, squares) with 5% and 95% quantiles (dashed lines), and expected degrees w_i (circles) when assuming a power-law (top panel) or an Erdős-Rényi (bottom panel) partial correlation model.

the other hand, it is easy to see that the relationship between eigenvalues and expected degrees does not hold.

4 Empirical Illustration

In this last section we carry out an empirical study to assess to which extent real data presents empirical features that are consistent with the power-law partial correlation network model. In particular, we consider a large panel of daily stock returns of companies listed in the S&P500 index. The panel spans from 2006-01-03 to 2013-12-31 and contains all the companies that have been trading for more than 2000 days in the S&P500 throughout the sample period, which delivers a sample of 387 companies over 2013 time

Figure 6: PARTIAL CORRELATION NETWORK OF THE S&P500.

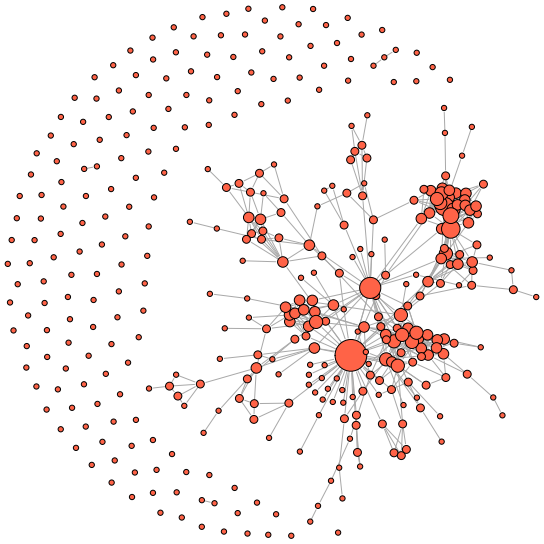
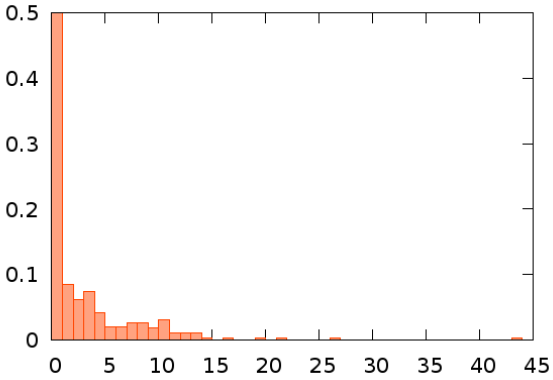


Figure 7: DEGREE DISTRIBUTION OF THE S&P 500



periods. The dependence structure of these data is clearly much more complex than the one implied by the power-law partial correlation network model and the analysis carried out in this section only aims at providing rough empirical validation for our model.

To illustrate the structure of the conditional correlations among the considered stock returns, we begin by estimating the partial correlation network structure of the stocks in the panel following the approach of Barigozzi and Brownlees (2016). Let r_{it} denote the return of stock i on day t and let r_{mt} denote the market return over the same period. We assume that the returns are generated according to the following CAPM type factor

model

$$r_{it} = a_i + b_i\, r_{mt} + \epsilon_{it}\ , \quad i = 1, \ldots, n, t = 1, \ldots, T\ , \tag{12}$$

where $\epsilon_t = (\epsilon_{it})_{i=1}^n$ are independent Gaussian vectors with zero mean and covariance $\boldsymbol{\Sigma}_\epsilon$. Denote the covariance matrix of the system as $\boldsymbol{\Sigma}$. Then concentration matrix is

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_\epsilon^{-1} - \frac{\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{b}\boldsymbol{b}'\boldsymbol{\Sigma}_\epsilon^{-1}}{1 + \boldsymbol{b}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{b}},$$

where $\boldsymbol{b}$ is the $n \times 1$ vector of factor loadings b_i. It is straightforward to see that, under the usual assumption of pervasive factors, that is, $\sum_{i=1}^n b_i^2 = O(n)$ or equivalently $\lambda_1^\Sigma = O(n)$, the matrix $\boldsymbol{\Sigma}^{-1}$ cannot be sparse and consequently the corresponding network will be fully interconnected. Therefore, as already noticed above, the model proposed in this paper can be seen as a model for the partial correlation structure of the stocks conditional on the market factor, r_{mt}, that is, the partial correlation structure of the idiosyncratic shocks ϵ_{it}, and, according to our notation, we define $\boldsymbol{K} = \boldsymbol{\Sigma}_\epsilon^{-1}$.

As it is commonly assumed in empirical finance, the market factor is treated as observed and is here identified with the S&P500 index. In order to estimate the network we first obtain the idiosyncratic shocks $\widehat{\epsilon}_{it}$ as the least squares residuals of (12), and we then estimate the partial correlation network using the `space` algorithm, a LASSO estimation procedure proposed by Peng, Wang, Zhou, and Zhu (2009).[1]

We report the graph plot and degree distribution of the estimated network in Figure 6 and 7 respectively. The estimated network exhibits several features that are commonly encountered in power-law graphs. The network is made up of a giant component containing several hubs and a few small components. There is substantial heterogeneity in the number of connections of each vertex and the most interconnected vertices have a large number of connections relative to the total. Accordingly, the degree distribution of the network exhibits heavy tails. These results justify our assumption of an underlying

[1]This estimator depends on a tuning parameter which determines the degree of penalization of the LASSO procedure, which is here chosen on the basis of the BIC criterion, as suggested among others by Peng *et al.* (2009) themselves.

power-law partial correlation network.

Theorem 1 shows that in large multivariate systems a power-law partial correlation structure can lead to a high level of collinearity. In particular, Theorem 1 shows that the distribution of the largest eigenvalues of the concentration matrix is closely related to the power-law distribution of the degrees of the network. This motivates us to fit a power-law distribution to the largest eigenvalues of the concentration matrix in the attempt to get insight on how heavy tailed the eigenvalues distribution is. Notice that evidence of power-law behavior in the distribution signals high degree of collinearity in the data, irrespective of whether the data has been generated by our model or not. Given the above results and Theorem 1, we assume that the cumulative distribution function of λ_i^K is given by a power-law distribution with power-law exponent β, that is

$$\mathbb{P}(\lambda_i^K \geq \ell) = \left(\frac{\ell}{\ell_{\min}}\right)^{-(\beta-1)}, \qquad \ell > 0.$$

where $\ell_{\min} \equiv \lambda_{i_{\max}}^K$, that is this distribution holds for the largest $i_{\max}$ eigenvalues of $\boldsymbol{K}$. Different methodologies have been proposed in the literature to estimate the power-law parameter β. In this work we focus on the approach proposed in Clauset, Shalizi, and Newman (2009), which allows one to estimate simultaneously β and $i_{\max}$. The technique consists of computing the maximum likelihood estimator of β (the Hill estimator) for different values of $i_{\max}$ and then to find the threshold $i_{\max}^*$ that minimises the Kolmogorov-Smirnov distance between the empirical cumulative distribution of the data and the theoretical cumulative distribution obtained using different values of $i_{\max}$. Under appropriate regularity assumptions Embrechts, Klüppelberg, and Mikosch (1997) show that, for n and $i_{\max}^*$ large enough, $\sqrt{i_{\max}^*}(\widehat{\beta}_n - \beta) \sim N(0, \beta^2)$.

Obviously, the eigenvalues of the concentration matrix are not observed and need to be estimated from the data. Let $\widehat{\boldsymbol{\Sigma}}_\epsilon$ be an estimator of the covariance matrix of the idiosyncratic shocks obtained from a sample of size T and denote by $\widehat{\lambda}_i^{\widehat{\boldsymbol{\Sigma}}_\epsilon}$ its eigenvalues. Then the eigenvalues of the concentration matrix $\boldsymbol{K} = \boldsymbol{\Sigma}_\epsilon^{-1}$ are estimated as $\widehat{\lambda}_i^{\widehat{\boldsymbol{K}}} = (\widehat{\lambda}_{n-i+1}^{\widehat{\boldsymbol{\Sigma}}_\epsilon})^{-1}$. Given, $\widehat{\epsilon}_{it}$, we consider here two possible alternative estimators of $\boldsymbol{\Sigma}_\epsilon$: the

sample covariance matrix estimator and the shrinkage covariance matrix estimator of Ledoit and Wolf (2004).[2] The second estimator is particularly useful when the sample size T is large and of size comparable to n.

We report the estimation results of the $\widehat{\beta}_n$ parameter in Table 1. The left panel of Table 1 reports the estimates $\widehat{\beta}_n$ obtained using the methodology of Clauset *et al.* (2009) as well as the Hill estimator using fixed choices of the threshold $i_{\max}$. We report parameter estimates constructed on the basis of both the sample covariance and the regularised covariance matrix. Estimates exhibit a moderate degree of variation depending on the chosen estimation approach. Interestingly, all estimates hint at a tail parameter larger than 2 and smaller than 3, indicating a heavy-tailed power-law distribution.

As an additional investigation, we also carried out the same analysis on the constituents of the S&P1500 and report results in the right panel of Table 1. The results on the S&P1500 deliver results which are in line with the one based on the smaller sample.

Finally, for both datasets considered, we show in Figure 8 the size-rank log-log plot of the largest estimated eigenvalues of the concentration matrix based on the shrinkage covariance estimator jointly with the power-law fit (reported for different values of the intercept) based on the Clauset *et al.* (2009) procedure. Overall, we find the tail of the largest eigenvalues of the concentration matrix to be well described by a power-law parameter smaller than three.

5 Conclusions

Power-law networks are a class of random graphs able to reproduce several of the empirical stylized facts that can be observed in real world networks, in particular small world effects and power-law degree distribution. This motivates us to investigate the consequence of a power-law interconnectedness in partial correlation network models. To do this we introduce a tractable class of partial correlation network models with an underlying power-law network structure. Our central result concerns the behavior of the largest

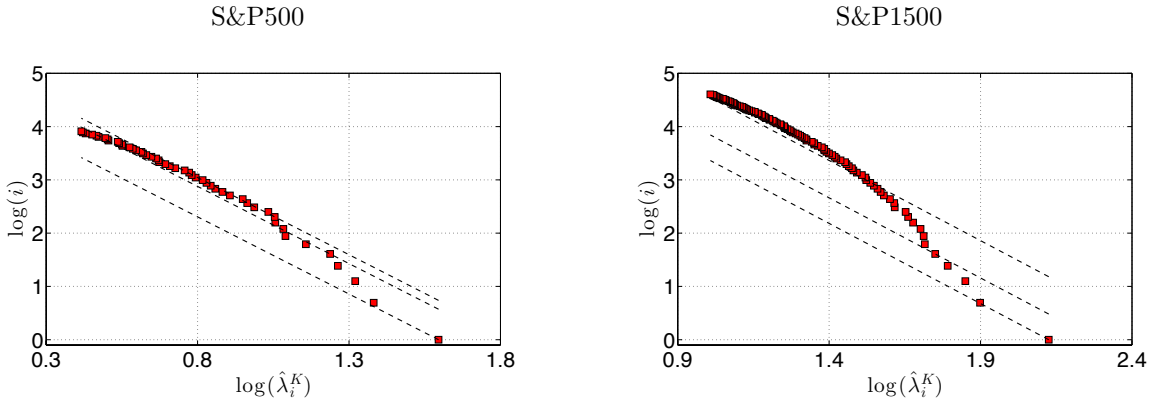
[2] Alternative regularised covariance estimators are given for example in Bickel and Levina (2008); Lam and Fan (2009); Abadir, Distaso, and Zikes (2014).

Table 1: POWER-LAW BEHAVIOR OF S&P PANELS.

S&P500 ($n = 387$)			S&P1500 ($n = 1234$)		
Threshold $i_{\max}$	Sample covariance	Regularised covariance	Threshold $i_{\max}$	Sample covariance	Regularised covariance
100	2.79 (0.28)	3.04 (0.30)	250	2.65 (0.17)	2.87 (0.18)
150	2.42 (0.20)	2.65 (0.22)	350	2.33 (0.12)	2.50 (0.13)
200	2.18 (0.15)	2.37 (0.17)	450	2.13 (0.10)	2.26 (0.11)
CSN	2.94 (0.31)	2.90 (0.26)	CSN	2.60 (0.16)	2.98 (0.20)
$i_{\max}$	90	122	$i_{\max}$	266	230

Values of the estimated power-law exponent $\widehat{\beta}_n$ with standard errors in parenthesis. CSN: Clauset *et al.* (2009) method.

Figure 8: POWER-LAW BEHAVIOR OF S&P PANELS.



Log-log rank-size plot for the 50 largest eigenvalues $\log(\widehat{\lambda}_i^K)$. Dashed lines: power-law fit for the largest log-eigenvalues with slope $-\widehat{\beta}_n$ given by the Clauset *et al.* (2009) method and for different values of the intercept.

eigenvalues of the concentration matrix of the model. We show that the largest eigenvalues of the concentration matrix converge to an affine function of the degree of the vertices with largest expected degree. The result implies that when the tails of the power-law distribution are heavy and the system dimension is large, the system exhibits a high degree of collinearity. As an empirical illustration, we analyze the covariance matrix of

a large panel of stock returns of companies listed in the S&P500 index and document that the data exhibits empirical features that are consistent with the power-law partial correlation network model.

References

Abadir, K. M., Distaso, W., and Zikes, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics*, **181**, 165–180.

Acemoglu, D., Carvalho, V., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, **80**, 1977–2016.

Anderson, W. and Morley, T. (1985). Eigenvalues of the Laplacian of a graph. *Linear and Multilinear Algebra*, **18**, 141–145.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, **71**, 135–171.

Barigozzi, M. and Brownlees, C. (2016). NETS: Network Estimation for Time Series. Technical report, Barcelona GSE.

Bickel, P. J. and Levina, E. (2008). High Dimensional Inference and Random Matrices. *The Annals of Statistics*, **36**, 2577–2604.

Billio, M., Getmansky, M., Lo, A., and Pellizzon, L. (2012). Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *Journal of Financial Economics*, **104**, 535–559.

Bollobás, B., Janson, S., and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, **31**, 3–122.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities:A Nonasymptotic Theory of Independence*. Oxford University Press.

Brouwer, A. E. and Haemers, W. (2008). A lower bound for the laplacian eigenvalues of a graph-proof of a conjecture by guo. *Linear Algebra and Applications*, **429**, 21312135.

Brouwer, A. E. and Haemers, W. (2011). *Spectra of graphs*. Springer.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, **51**, 1281–1304.

Chung, F. (1997). *Spectral graph theory*, volume 92. American Mathematical Soc.

Chung, F. and Lu, L. (2006). *Complex Graphs and Networks*. American Mathematical Society, Providence.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, **51**, 661–703.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, **28**, 157–175.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, **182**, 119–134.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*, volume 33. Springer.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **5**, 17–61.

Hagerup, T. and Rüb, C. (1990). A guided tour of Chernoff bounds. *Information Processing Letters*, **33**, 305–308.

Hartmann, P., Straetmans, S., and de Vries, C. (2007). Banking System Stability. A Cross-Atlantic Perspective. In *The Risks of Financial Institutions*, NBER Chapters, pages 133–192. National Bureau of Economic Research, Inc.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014a). Financial network systemic risk contributions. *Review of Finance*. available online.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014b). Forecasting systemic impact in financial networks. *International Journal of Forecasting*, **30**, 781–794.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**, 13–30.

Karp, R. (1988). *Probabilistic Analysis of Algorithms*. Class Notes, University of California, Berkeley.

Lam, C. and Fan, J. (2009). Sparsistency and Rates of Convergence in Large Covariance Matrix Estimation. *The Annals of Statistics*, **37**, 4254–4278.

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, **88**, 365–411.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial Correlation Estimation by Joint Sparse Regression Models. *Journal of the American Statistical Association*, **104**, 735–746.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**, 1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, **20**, 147–162.

van der Hofstad, R. (2015). *Random Graphs and Complex Networks*.

A Auxiliary Results

Proposition 1. (COVARIANCE OF THE PARTIAL CORRELATION NETWORK) *The covariance matrix of the partial correlation network model is*

$$\boldsymbol{\Sigma} = \sigma^2(\boldsymbol{I} + \phi\boldsymbol{D})^{-1} + \frac{\sigma^2}{\phi}\sum_{k=1}^{\infty}\boldsymbol{W}_k \,,$$

where $\boldsymbol{W}_k$ is defined as

$$[\boldsymbol{W}_k]_{i,j} = \sum_{w\in\mathcal{W}_k} \frac{\mathbb{1}_{\{w \text{ goes from } i \text{ to } j\}}}{\prod_{v\in\mathcal{V}(w)}(d(v)+1/\phi)} \,, \tag{A-1}$$

where $\mathcal{W}_k$ denotes the set of walks of length k in the network, $\mathcal{V}(w)$ denotes the set of vertices in walk w, $d(v)$ denotes the degree of vertex v and $\mathbb{1}$ is the indicator function (and convergence is defined in spectral norm). We call $\boldsymbol{W}_k$ the weighted walk matrix of length k.

Proof. Define $\boldsymbol{C} = \boldsymbol{I} + \phi\boldsymbol{D}$ and notice that $\boldsymbol{C}$ is a diagonal matrix with positive diagonal entries. Then

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{K}^{-1} = \sigma^2(\boldsymbol{C} - \phi\boldsymbol{A})^{-1} \\
&= \sigma^2(\boldsymbol{C}^{1/2}(\boldsymbol{I} - \phi\,\boldsymbol{C}^{-1/2}\boldsymbol{A}\boldsymbol{C}^{-1/2})\boldsymbol{C}^{1/2})^{-1} \\
&= \sigma^2\boldsymbol{C}^{-1/2}(\boldsymbol{I} - \phi\,\boldsymbol{C}^{-1/2}\boldsymbol{A}\boldsymbol{C}^{-1/2})^{-1}\boldsymbol{C}^{-1/2}
\end{aligned}$$

where $\boldsymbol{C}^{1/2}$ and $\boldsymbol{C}^{-1/2}$ denote diagonal matrices with diagonal equal respectively to the square root and inverse square root of the diagonal of $\boldsymbol{C}$. Notice that the eigenvalues of $\phi\,\boldsymbol{C}^{-1/2}\boldsymbol{A}\boldsymbol{C}^{-1/2}$ are smaller than one in absolute value. (To see this, note that $\|\phi\,\boldsymbol{C}^{-1/2}\boldsymbol{A}\boldsymbol{C}^{-1/2}\| \leq \phi\|\boldsymbol{A}\|/\|\boldsymbol{C}\|$ and that $\|\boldsymbol{C}\| \geq \phi D_{\max}$ and $\|\boldsymbol{A}\| \leq D_{\max}$ where $D_{\max}$ denotes the highest degree in the network.) Thus, we can apply the von Neumann series identity to $(\boldsymbol{I} - \phi\,\boldsymbol{C}^{-1/2}\boldsymbol{A}\boldsymbol{C}^{-1/2})^{-1}$ and get

$$\begin{aligned}
\boldsymbol{K}^{-1} &= \sigma^2\boldsymbol{C}^{-1/2}\left(\boldsymbol{I} + \sum_{k=1}^{\infty}(\phi\,\boldsymbol{C}^{-1/2}\boldsymbol{A}\boldsymbol{C}^{-1/2})^k\right)\boldsymbol{C}^{-1/2} \\
&= \sigma^2\boldsymbol{C}^{-1} + \sigma^2\sum_{k=1}^{\infty}\boldsymbol{C}^{-1/2}(\phi\,\boldsymbol{C}^{-1/2}\boldsymbol{A}\boldsymbol{C}^{-1/2})^k\boldsymbol{C}^{-1/2} \\
&= \sigma^2\boldsymbol{C}^{-1} + \sigma^2\sum_{k=1}^{\infty}\phi^k\underbrace{\boldsymbol{C}^{-1}\boldsymbol{A}\,\boldsymbol{C}^{-1}\boldsymbol{A}\,\ldots\,\boldsymbol{C}^{-1}\boldsymbol{A}}_{k \text{ times}}\boldsymbol{C}^{-1} \,.
\end{aligned}$$

Recall that the (i,j) element of $\boldsymbol{A}^k$ is equal to the number of distinct walks of length k from i to j, which can be expressed as

$$\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_{k-1}=1}^{n} A_{i,i_1}A_{i_1,i_2}\cdots A_{i_{k-1},j}.$$

Analogously, it is straightforward to see that the (i,j) entry of $(\boldsymbol{C}^{-1}\boldsymbol{A})^k\boldsymbol{C}^{-1}$ is equal to

$$\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_{k-1}=1}^{n}\frac{A_{i,i_1}}{\phi D_i+1}\frac{A_{i_1,i_2}}{\phi D_{i_1}+1}\cdots\frac{A_{i_{k-1},j}}{\phi D_{i_{k-1}}+1}\frac{1}{\phi D_j+1}.$$

We obtain the statement of the proposition by applying the last equation and rearranging terms. □

Proposition 2. *(See Hoeffding, 1963; Karp, 1988; Hagerup and Rüb, 1990.) Let $X_1,\ldots,X_n$ be independent random variables, taking their values from $[0,1]$. If $m=\mathbb{E}S_n$ with $S_n=\sum_{i=1}^n X_i$ then for any $t\geq m$,*

$$\mathbb{P}\{S_n\geq t\}\leq\left(\frac{m}{t}\right)^t e^{t-m}\ .$$

In particular, for all $u\leq m$,

$$\mathbb{P}\{S_n\geq\mathbb{E}S_n+u\}\leq e^{-u^2/m}\ .$$

On the other hand, for all $u\leq m$,

$$\mathbb{P}\{S_n\leq\mathbb{E}S_n-u\}\leq e^{-u^2/(2m)}\ .$$

Proposition 3. *(Anderson and Morley, 1985.) Let $\mathcal{G}$ be a graph with edge set $\mathcal{E}$. Then the maximum eigenvalue of the graph Laplacian $\boldsymbol{L}_{\mathcal{G}}$ satisfies*

$$\lambda_1^{L_{\mathcal{G}}}\leq\max\{D_i+D_j|(i,j)\in\mathcal{E}\}\,.$$

Proposition 4. *(Brouwer and Haemers, 2008, see also Brouwer and Haemers, 2011.) Let $\mathcal{G}$ be any graph on n vertices that is not the union of the complete graph K_m and $n-m$ isolated vertices for any $m\leq n$. Then, for all $i=1,\ldots,n$,*

$$\lambda_i^{L_{\mathcal{G}}}\geq D_i-i+2\ ,$$

where $\lambda_i^{L_{\mathcal{G}}}$ is the i-th largest eivenvalue of the Laplacian of $\mathcal{G}$ and D_i is the i-th largest degree of the vertices of $\mathcal{G}$.