Concentration of Small world-Networks and application of spectral algorithms

Laala Zeyneb [a], Belguerna abderrahmane [b]

Department of Mathematics and Computer Science, University of Salhi Ahmed, Naama, 45000, Algeria

Abstract: Recent researches on statistical network analysis has strongly included the random matrix theory. The principal goal of random matrix theory is to provide a knowledge of many properties of matrices such as the statistics of matrix eigenvalues with elements taken randomly from various probability distributions. In this paper, we present some results on the concentration of the adjacency and laplacian matrices around the expectation under the small-world network. We also present some relevant network model that may be of interest to probabilists looking for new directions in random matrix theory as well as random matrix theory tools that may be of interest to statistician looking to verify the features of network algorithm. Application of some results to the community detection problem are discussed.

Keywords: Small-world network, Regularization, Concentration, Random matrix, Random Graph, Network Statistical Analysis.

1. Introduction (Times New Roman 10 Bold)

Many different types of networks play different roles in society for different purposes in the real world, such as social networks **(Jeong H. et al 2000)** that represent human social interactions, citation networks that represent the articles of various authors published in a specific field and their associated citations in other papers, technological networks that represent resource distribution, and biological networks that represent protein-protein interaction.

Furthermore, network evolution is a key study subject in network modelling **(Joseph A. and Yi B. 2016)**, centrality analysis **(Reka A. and Barabási 2002)**, information diffusion, link prediction **(Krivelevich M and Vu V. 2002)**, and community discovery **(Krivelevich M and Sudakov B 2003)**. Power-law degree distributions **(Kossinets G and Watts D.J. 2006)**, small world networks **(Liben-Nowell D and Kleinberg J 2007)**, and community structures are some of the primary elements identified in the social network. By revealing such hidden characteristics, analysts can investigate the social network's functionality. There are numerous definitions of community depending on the contexts relevant to diverse applications.

According to a few widely accepted definitions, they are characterized as a group of nodes with dense connections among themselves as opposed to sparsity outside the group. A community in a social network is a group of people who have common interests and information. Learning from network social dynamics can be utilized to find hidden communities **(Montoya J.M and Solé R.V 2002)**. Their detection contributes to a better understanding of social and functional activity in the social network. However, due to the rapid growth in the size of social networks, identifying subgroups in the network within a certain time frame is a significant problem. Real-world social networks tend to follow the power-law **(Clauset A. and Newman M.E.J. 2004)** in both degree distribution and community size distribution.

Many real-world systems, including social networks **(Newman M.E.J. 2000), (Newman M.E.J. 2004)**, economic networks **(Nobutaka S. and Ryuhei M. 2016)**, transportation systems **(Oliveira R. 2010), (Pastor-Satorras R. and Vespignani A. 2001)**, epidemic spread **(Pastor-Satorras R., and Vespignani A. (2001))**, metabolic networks **(Qin T. and Rohe K. 2013), (Reka A. and Barabási 2002)**, food web **(Rosen D. et al. 2011)**, and so on, contain small world networks.

Because of the discovery of these traits, researchers have been able to investigate, improve, and change small world networks for useful applications. For example, the propagation of sickness or epidemics in social networks has been extensively studied, and small world characteristics have been used to develop ways for isolating and quarantining afflicted societies **(Stephen A. and Toubia O. 2009)**. Similarly, in online social groups **(Vu V.H. 2007)**, stock affiliation **(Wagner A. and Fell D. 2000)**, air transportation **(Rozenblat C. et al. 2008)**, and metabolic networks **(Watts D.J and Stogatz S.H 1994)**, small world attributes have been employed to analyze and improve decision-making processes.

Small world networks have two structural features: the small world effect and clustering. The small world effect states that any two nodes in a network can be connected by a short path. This concept is quantified by the average path length (APL) metric, which provides a value for each network reflecting the average distance

between any two nodes in the network. Because the typical APL values for every random network are low **(Zaidi F. 2012)**, random networks exhibit this feature as well.

Clustering, also known as transitivity **(Watts D.J and Stogatz S.H 1998)** or the fraction of transitive triples in a network **(Newman M.E.J. 2000)**, is the second characteristic of small world networks. It is the hypothesis that two nodes with a common neighbor are more likely to be connected to each other. This characteristic can be quantified using the clustering coefficient metric. When compared to a random network, this attribute is only present in real-world networks since in a random network, every node has an equal probability of connecting to any other node, hence transitive triples or triads are rarely created.

Many models that artificially construct networks with small world traits have been presented. These models are extremely useful because they can be used to build artificial networks with the necessary attributes and sizes that approximate real-world networks. These networks can then be used as a test of efficiency for a variety of experimental and empirical studies. Furthermore, these models give us insight into many real-world systems, as well as ways for identifying and analysing their structure.

A graph partitioning problem resembles community discovery. The majority of network partitioning algorithms are centred on improving a quality function. The first community discovery algorithm was proposed by Girvan and Newman, and it is based on a graph's hierarchical partitioning problem.

Modularity was used as the objective function for determining the quality of the generated division in their study **(Moore C. and Newman M.E.J. 1990).** Edges are deleted iteratively in order of their edge-betweenness value until the maximum modularity is reached in this algorithm. The number of shortest paths between two nodes that travel through an edge can be described as the edge's betweenness value.

A random graph is a type of graph in which nodes are connected by edges that are distributed randomly among them, but it does not resemble a real-world network. Watts and Strogatz devised a random graph generation model that aids in the production of random graphs with small world attributes such as average short path length and high clustering coefficient **(Scott J.O. 2000)**. A small global network has been discovered to exist between regular and random graphs. In small world networks, communities are regularly observed.

The ego network is a collection of nodes that includes a central actor and other nodes that are connected to it directly. They have qualities that are similar to those of a small-world network. When compared to the other nodes in the ego network, the central node may have the most influence. Exploring focal nodes may aid in the modelling of network influence propagation **(Kempe, D. et al. 1997).** Users can form groups on the social network based on a shared interest or a common event in their social lives. When friends are added or deleted dynamically in their social lives, group formation becomes a challenging process.

The notion of a small-world is resulting from the work of Stanley Milgram in 60's. Milgram's experiment was conceived to choose individuals from the Middle West to send a letter to a recipient from the Oust.

Milgram was surprised to find just the mean of chains achieve to recipient which was only **5.6.** This experience gave a myth that becomes in his version "*six degrees of separation*".

A small-world network is a mathematical graph in which most nodes are not neighbours, but most nodes can be reached by a small number of hops or steps from each other. A small-world network is defined as one in which the typical distance **L** between two randomly chosen nodes (the number of steps required) grows according to the logarithm of the network's node count **N**. That is

$$L \propto \log N$$

As a result, strangers are linked by a short chain of acquaintances in the formation of a social network, resulting in the tiny world phenomenon. The small-world effect, social networks, the underlying architecture of the Internet, wikis like Wikipedia, and gene networks are all depicted in numerous empirical graphs..

Cliques and near-cliques, which are sub-networks with connections to virtually any two nodes within them, are common in small-world networks. This is due to the distinguishing feature of a high clustering coefficient. Second, the majority of node pairs will be connected by at least one short path. This is because of the distinguishing feature of having a short mean-shortest path length. Small-world networks are frequently associated with a number of other characteristics. In most networks, there is an overabundance of hub-nodes with a large number of connections (known as high degree nodes). These hubs act as common connectors, bridging the short paths between other edges. Because many flights are routed through hub cities, the small-world network of airline flights has a short mean-path length (i.e., between any two cities, you'll probably have to take three or less flights).

This feature is frequently examined by examining the fraction of network nodes that have a specific amount of connections going into them (the degree distribution of the network). Networks with a higher predicted number of hubs will have a higher proportion of high degree nodes, and the degree distribution will be enriched at high degree values as a result. This is referred to as a fat-tailed distribution. Graphs with extremely varied topologies can be classified as small-world networks if they meet the two definitional requirements outlined above.

Small-world networks have been quantified by comparing clustering and path length of a given network to an equivalent random network with same degree on average. Watts and Strogatz (1998) are the first to demonstrate that the model of random graph of Erdös-Rényi that does not account for properties observed in natural networks. Their argument is based on the notion of local density, which they have been able to quantify by introducing the notion of grouping coefficient.

The Watts and Strogatz model, in network theory, is used for the construction of some small world networks. Genetically, it is a model of generation of random graphs with small average distances and high values of the clustering coefficient. The mathematical model takes the name of the mathematicians Duncan Watts and Steven Strogatz in the year 1998 in the journal of Nature.

Watts and Strogatz were intrigued by the fact that many networks in the real world display high levels of clustering, but small distances between most nodes. In the classical random graph model, the diameter is about $O(\log(N_v)$, and indicating "small-world" behavior, the clustering coefficient C is about N_v^{-1}, which suggests very little clustering.

Watts and Strogatz proposed starting with a graph with lattice structure and then randomly "rewiring" a small percentage of the edges to create a network graph with both of these properties. In this model, we start with a set of N vertices that are arranged in a periodic fashion and connect each vertex to r of its neighbours on each side. Then, for each edge, one end is moved independently and with probability p to be incident to another vertex, where the new vertex is chosen uniformly.

In this paper, we start with preliminaries and properties of such kinds small-world Network. Then, we present some results about the regularization for the concentration of graphs. Finally, we discuss numerical results.

2. Preliminaries

Let's first recall some basic notions. For an undirected finite graph, we note A the $n \times n$ adjacency matrix with vertex set V, |V|=n, with elements $A_{ij} = 1$ if there is an edge between the two vertices i and j, and 0 otherwise. The normalized Laplacian is defined by

$$L = I_n - D^{-1/2} A D^{-1/2}$$

where I_n is the $n \times n$ identity matrix, and $D = diag(d_i)$ is an $n \times n$ diagonal matrix with nodes degrees $d_i = \sum_{j \in V} A_{ij}$ on the diagonal.

The Laplacian matrix's eigenvalues and eigenvectors reflect some fundamental geometric characteristics of the graph G. We suggest in this paper that studying a generalized version of the Laplacian is important which is defined in **(Hafiz T. and Romain C. 2016)** by

$$L_\alpha \propto D^{-\alpha} B D^{-\alpha}$$

with $\alpha \in [0,1]$and B is the modularity matrix given by $B = A - \frac{dd^T}{dI_n}$. In particular, when $\alpha = 0, L_\alpha \propto B$ and when $\alpha = \frac{1}{2}$ $L_\alpha \propto I_n - L$.

3. Concentration of random graphs

The adjacency and Laplacian matrices, in particular, reflect the geometry of graphs. Random graphs concentration is defined as the concentration of these canonical random matrices around their mean. We will discuss here, the concentration of random graphs.

The distance between any two nodes in many real-world networks is relatively small, while the level of transitivity, or clustering, is relatively high.

The average path length in Erdös-Rényi random networks is short, which means that there are only a few edges connecting two nodes. The small-world property is shared by many real-world networks. A high degree of transitivity is one property of real-world graphs, such as social networks, that ER random graphs lack. Because

your friends are likely to be friends, the social network describing those friendships is likely to have a lot more triangles than the ER random network model predict. So in this paper, we are interesting in small-world networks.

We consider the adjacency matrix A which is symmetric with spectrum norm is defined as follows

$$\lambda(A) = \sup_{v \in R^n, ||v||=1} |v^T A v|$$

The following is likely the most well-known estimate for $\lambda(A)$as stated **by (Füredi Z. and Komlös J. 1981)**.

Theorem 1: For a random matrix A there is a positive constant $c = c(\sigma, K)$ (depending on $\sigma$and K but not on n) such that

$$2\sigma\sqrt{n} - cn^{\frac{1}{3}} \ln n \leq \lambda(A) \leq 2\sigma\sqrt{n} + cn^{\frac{1}{3}} \ln n,$$

holds almost surely

Theoretically and practically, sharpening this estimate is of great interest. **(Krivelevich M. and Vu V. 2022)** showed that $\lambda(A)$ is highly concentrated around its mean.

Theorem 2: For a random matrix A there is a positive constant c=c(K) such that for any t>0

$$P(|\lambda(A) - E(\lambda(A))| \geq ct) \leq 4e^{-t^2/32}$$

(Can M. et al. 2017) show based on studied of **(Wasserman S. and Faust K. 1994**) and **(Füredi Z. and Komlòs J. 1981)** that dense graphs concentrate well around the expected degree. And, they found that with a high probability we have

$$||A - E(A)|| = 2\sqrt{d}(1 + o(1)) \quad if \;\; d \gg log^4 n$$

Now, for sparse graphs, the concentration breaks down because the expected degree $d = ||E(A)||$ is bounded.

As stated in **(Krivelevich M and Dudakov B. 2003)**, an Erdös-Rényi graph from G(n,p) satisfies with high probability that

$$||A|| = (1 + o(1))\sqrt{d(A)} = (1 + o(1))\sqrt{\frac{\log n}{\log\log n}} \quad if\ d = O(1),$$

Here d(A) denotes the maximal degree of the graph, so in this case it will be $||A|| \gg ||E(A)|| = d$which indicates that sparse graphs do not concentrate. Nodes with a very high degrees make the norm A large in this case.

According to **(Can M. et al. 2017)**, we can see that the best method to concentrate the random matrix is the regularization. In other word, concentration is enforced via regularization.

4. Regularization

Following the work of **(Zaidi F. 2012)** in which the author proposed a new method for generating a network. The proposed model for generating small-world networks from random networks was presented in which they started by generating an Erdös-Rényi as the first step. The following step is to replace each node in the random network with a triad, which is a group of three nodes connected by three edges to form a clique of size 3. In the last step, an edge is added between the two triads that replace the original nodes in the initial random network, linking one of the randomly selected nodes from each triad for each edge in the random network. The resulting network has two structural properties that are typical of a small world network.

The addition of triads increases the network's overall clustering coefficient, which is a key characteristic of small-world networks. The newly created triads' network connectivity is based on the randomly generated network. We already know that random networks have a minor effect on the average path length in the created network **(Reka A. and Barabási 2002).** Note that the new number of nodes is N=k*n, with k number of cliques.

Definition 1 : The average path length of a graph G=(V,E) is defined by

$$\text{APL}=\frac{2}{n(n-1)}\sum_{\{i,j\}\subseteq V, i\neq j} dist(i,j)$$

Corollary 1: From **(Qin T. and Rohe K. 2013)** For a d-regular graph G of order n and diameter 3,

$$APL(G) \geq 3 - \frac{d(d+1}{n-1}$$

Theorem 3: Consider a random graph generated from the model proposed by **(Zaidi F. 2012)** and let $d = max_{ij}\, Np_{ij}$. For the high number of clustering coefficient. Then, the new adjacency matrix noted $A_N$of the new created graph concentrate.

$$\left|\left|A_N - E(A)\right|\right| = O(\sqrt{d})$$

Then the average path length concentrate around their mean.

Concentration of laplacian

The concentration of Laplacian in random graphs has been studied in many research's papers (**see (Chaudhuri K. et al. 2012), (Oliveira R 2010), (Qin T. and Rohe K. 2013), (Joseph A. and Yi B. 2016), and (Gao et al. 2015))**. Similar to adjacency matrix, the laplacian is known to concentrate in dense regime and fail for sparse.

The clustering coefficient and average path length values show that networks generated by this model generate small world networks. Because the metric clustering coefficient measures the presence of triads in networks, substituting triads for nodes in a random network statistically satisfies the properties of a small network with high clustering coefficient values, but larger size cliques may exist in real-world networks.

In this model, larger cliques can be used instead of triads. The resulting network clearly has small-world properties. This concept can be expanded to replace cliques of varying sizes in order to build a small world network, and the values do not have to be constant. Depending on how and why this artificially generated network will be used, this can be a range between two constants or a random distribution **(Zaidi F. 2012).** Domain knowledge can also be used to determine the size of these cliques, which can vary greatly between domains.

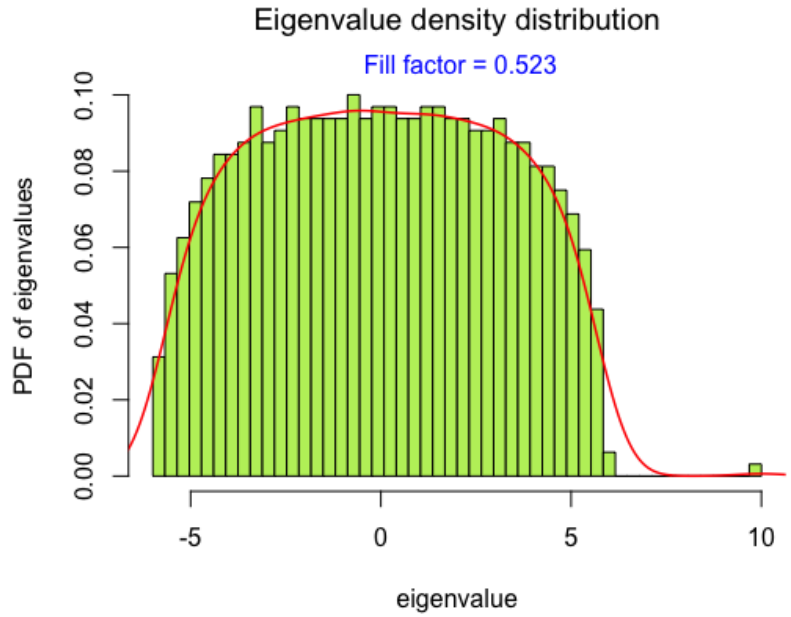
Theorem 4: (Concentration of the regularized Laplacian)

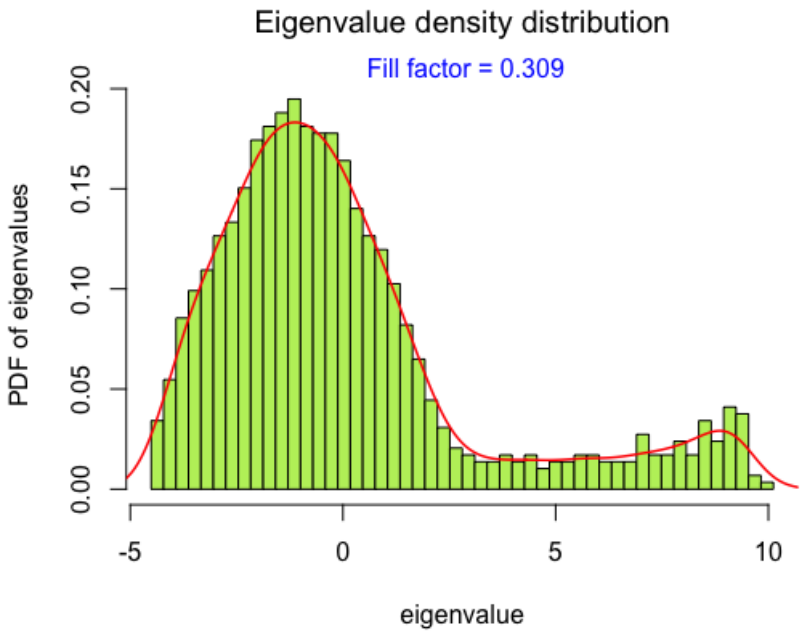
Consider a random network generated first by the Erdös-Rényi model, after modification and replacing each node by a clique size equal to five. The Laplacian matrix associated to the new crated adjacency matrix concentrates;

$$\left|\left|L(A_{5n}) - L\big(E(L_{5n})\big)\right|\right| = O(\frac{1}{\sqrt{d}})$$

5. Numerical Application

To conclude our discussion of regularization, we conduct a numerical experiment to demonstrate the effect of regularization. Using computer simulation we consider a small world network generated from Watts-Strogatz network with n=3000 nodes, with average path length in the interval [3,5] and probability equal to 0.05. At the same time we generate another random graph (Erdös-Rényi) with same number of nodes. We develop this network by adding cliques, we started replacing each node by 3 nodes, and then 5 nodes respectively to compare the results. The figures below present the histogram of the spectrum of the generated small-world network and the Erdös-Rényi graph after regularization. If we compare between the figures, we can see clearly the concentration of the new graph which is the network after regularization and that means the regularization can affect the concentration of the graphs.





6. Application to community detection

Another key feature of graphs is the possibility of dividing them into cohesive and dense modules or sub-graphs that are weakly connected to each other. The traditional approach for detecting communities in a graph is to minimize the cutting rate (i.e. the number of inter-community edges) by performing a partitioning, with the main drawback of having to choose the number of communities beforehand.

The presence of community structures or clusters is another important feature of networks and the concentration of the random graphs has an important use in the statistical analysis of the graph, particularly in community detection.

For testing the network, we have used the proposed method for the some real network data, such as the famous dolphin's networks data and football college network. To generate a new adjacency matrix, we start replacing each node of the network with a triad, then we applied two spectral methods to find communities based on Bethe-Hessian matrix. It gives a correct estimation of the number of network.

7. Summary and discussion

Recently, modelling the networks and analyse their statistics properties become a more interesting topic due to the rapid changes of the world, and thus the whole world has become linked together. Understanding the structure of the network can help to analyse and predict some situation in the networks and the community structure is one of the important tool in analysing. We present an important property, the concentration in which some authors prove that it will be around the mean. The small world in which is the more realistic network can concentrate after the proposed regularization. We can conclude that the good regularization can affect the concentration of random graphs in a positive way.

References (APA)

Bakshy, E.; Rosenn, I.; Marlow, C.; Adamic, L. The role of social networks in information diffusion.
In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012;
pp. 519–528.

Bilke S., and Peterson C. (2001). Topological properties of citation and metabolic networks. Rev. E, 64:036106.

Borgatti, S.P.; Cross, R. A relational view of information seeking and learning in social networks. Manag. Sci. 2003, 49, 432–445.

Can M. Le, Elizaveta Levina and Roman Vershynin. (2017). Random Structures & Algorithms. 51.

Carrington, P.J.; Scott, J.; Wasserman, S. Models and Methods in Social Network Analysis; Cambridge University Press: New York, NY, USA, 2005; Volume 28.

Chaudhuri K., Chung F. and Tsiatas A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. Jpirnal of Machine Learning Research Workshop and Conference Proceedings.

Clauset, A.; Newman, M.E.J.; Moore, C. Finding community structure in very large networks. Phys. Rev. E 2004, 70, 066111.

Discazeaux C., Rozenb;at C., Koeing P.-Y. and Melançon G. (2007). Territorial and topological levels in worldwide air transport network. In 15th European Colloquium on Theoritical and Quantitative Geography, Montreux. Switzerland.

Feige U. and Ofek E. (2005). Spectral techniques applied to sparse random graphs. Wiley InterScience.

Freeman, L.C. Centrality in social networks conceptual clarification. Soc. Netw. 1978, 1, 215–239.

Füredi Z. and Komlòs J. (1981). The eigenvalues of random symmetric matrices. Combinatorica. Pages: 233-241.

Gao C. Ma Z., Zhang Y. and Zhou H. H. (2015). Achieving optimal misclassification proportion in stochatic block model. arXiv:1505.03772.

Guimera R., Mossa S., Turtschi A., and Amaral L.A.N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities'global roles. Proceedings of the national acdemy of sciences of the USA, 102(22) pp: 7794-7799.

Hafiz Tiomoko Ali and Romain Couillet. (2016). Random matrix improved community detection in heterogeneous networks. In signals, Systems and computers, 50th Asilomar Conference on, pages 1385-1389. IEEE.

Holme P. and Kim B. J. (2002). Growing scale-free networks with tunable clustering. Physical Review E, 65:026107.

Jeong H., Tomber B. Albert R., Oltvai Z. N., and BarabásiA.-L. (2000). the large-scale organization of metabolic networks. Nature, 407:651-654.

Joseph A. and Yi B. (2016). Impact of regularization on spectral clustering. Ann. Statist., 44(4) pp: 1765-1791.

Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the spread of influence through a social network.
In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data
Mining, Washington, DC, USA, 24–27 August 2003; pp. 137–146.

Kirman A. (1997). The economy as evolving network. Journal of evolutionary economics, Springer, 7(4) pp: 339-353.

Krivelevich M. and Vu V. (2002). Approximating the independence number and the chromatic number in expected polynomial time. J. Comb. Optim. 6(2), Pages: 143-155.

Krivelevich M. and Sudakov B. (2003). The largest eigenvalue of sparse random graphs. Combin. Probab. Comput., 12, Pages: 61-72.

Kossinets, G.; Watts, D.J. Empirical analysis of an evolving social network. Science 2006, 311, 88–90.

Liben-Nowell, D.; Kleinberg, J. The link-prediction problem for social networks. J. Assoc. Inf. Sci. Technol. 2007, 58, 1019–1031.

Montoya J. M and Sol\'{e} R. V. (2002). Small-world patterns in food webs. Journal of Theo. Biol.214:405-412.

Moore C. and Newman M. E. J. (1999). Epidemics and percolation in small-world networks. Phys. Rev. E, 61:5678-5682.
Newman, M.E.J.; Watts, D.J. Scaling and percolation in the small-world network model. Phys. Rev. E 1999, 60, 7332.
Newman M. E. J. (2000). Models of small world. Journal of Statistical Physics. 101:819-841.
Newman, M.E.J. Detecting community structure in networks. Eur. Phys. J. B-Condens. Matter Complex Syst. 2004, 38, 321–330.
Nobutaka Shimizu and Ryuhei Mori. (2016). Average shortest path length of graphs of diameter 3. arXiv:1606.05119v1.
Oliveira R. (2010). Concentration of the adjacency matrix and the laplacians in random graphs with independent edges. arXiv:0911/0600.
Pastor-Satorras R., and Vespignani A. (2001). Epidemic spreading in scalefree networks. Phys. Rev. Lett., 86(14):3200-3203.
Qin T. and Rohe K. (2013). Regularized spectral clustering under the degree-corrected stochastic block model. In advances in Neural Information Processing Systems, pages: 3120-3128.
Reka A. and Barabási. (2002). Statistical mechanics of complex networks. Rev. Mod. Phys., 74:47-97
Rosen D., Barnett G. A., and Kim J.-H. (2011). Social networks and online environments: when science and practice co-evolve. Social Netw. Analys. Mining, 1(1):27-42.
Rozenblat C., Melançon G., and Koeing P.-y. (2008). Continental integration in multilevel approach of world air transportation (2000-2004). Networks and spacial economics.
Scott J. O. (2000). Social network analysis: a handbook. SAGE publications.
Solis R. (2009). Visualizing stock-mutual fund relatioships through social network analysis. Global Journal of Finance and Banking Issues. 3(3).
Stephen, A.T.; Toubia, O. Explaining the power-law degree distribution in a social commerce network.
Soc. Netw. 2009, 31, 262–270.
Vu, V.H. (2007). Spectral norm of random matrices. Combinatorica 27, pages: 721-736.
Wagner A. and Fell D. (2000). The small world inside large metabolic networks.
Watts D. J. and Strogatz S. H. (1998). Collective dynamics of small-world networks. Nature.
Wasserman S. and Faust K. (1994). Social network analysis: methods and applications. Cambridge university press, Cambridge.
Zaidi F. (2012). Small world networks and clustered small world networks with random connectivity. Social Network Analysis and Mining, Springer.