

Reinventing the Triangles: Rule of Thumb for Assessing Detectability

Jeremi K. Ochab
M. Smoluchowski Institute of Physics
Jagiellonian University
Cracow, Poland
Email:jeremi.ochab@uj.edu.pl

Abstract—Based on recent advances regarding eigenspectra of adjacency matrices, derived from the random matrix theory, the paper reintroduces the global clustering coefficient as a simple means of assessing if nonrandom modules can be detected in undirected, unweighted networks.

I. INTRODUCTION

Across disciplines one of the most relevant features of complex networks is their modular structure. It is crucial in social analysis, it defines or represents the function of biological networks, it affects the processes taking place on networks. The modules, also called communities, bring information: they allow to visualize and understand data, and to predict behavior.

Consequently, *community detection* is at the forefront of data mining techniques. Even though numerous—and at times surprisingly effective—algorithms of community detection have been appearing now for a decade, there remains a long-standing question of statistical significance of the modules discovered in networks. Several approaches were already described in [1] (see Sec. 14, *Significance of clustering* therein), but new techniques and theoretical advances, notably those stemming from the random matrix theory, have since been brought into play.

As described in the following section, the key problem with assessing significance is caused by the detectability limit. Our aim is to propose a fast and efficient rule-of-thumb to discriminate between networks in which there can or cannot hide significant communities.

II. RATIONALE—LIMITS OF DETECTABILITY

It has been known that community detection algorithms tend to find some clusters even in random graphs (RG), that they fail to detect clusters in graphs which are not fully random, and that in RG ensembles themselves there are instances of graphs in which communities appear solely by chance. In order to alleviate these problems one may apply bootstrap procedures [2], which significantly enhance the partitioning. Alternatively, the problem has been tackled with the aid of extreme value statistics [3], so that graph partitions can be assigned a statistical significance score.

However, Reichardt and Leone [4] have shown that a sharp phase transition exists between undetectable and detectable cluster structures. Such a divide means that below the transition community structure found by an algorithm is unreliable.

Further studies by Decelle et al. [5] confirm that the transition exists, in some cases however it is possible—although exponentially hard—to find true clusters below this limit.

In a series of papers that followed [6]–[8] Nadakuditi, et al., have demonstrated a connection between spectra of adjacency matrices of graphs and detectability. Namely, the eigenvalues of the adjacency matrix either belong to a bulk (a continuous spectrum) or are isolated; in graphs without a community structure the only isolated eigenvalue is the largest one; any new community introduces a new isolated eigenvalue. Conversely, whenever an eigenvalue associated with a given community hides into the bulk, it becomes undetectable. From this perspective, the knowledge of a graph’s eigenspectrum should be sufficient to answer the question of significance of clusters, and it seems that the spectral methods of community detection ought to be favored. That being said, one issue remains: calculating the spectra is computationally expensive.

Thus, the aim is to quantify detectability by means of a single parameter which can be computed fast and easily for any given network. We argue that global clustering coefficient can be applied for that purpose.

III. REMINDER—CLUSTERING COEFFICIENTS

It has been long years since the clustering coefficient has become widely used to benchmark empirical networks and new—small-world [9] and scale-free networks—against random Erdős-Rényi (ER) graphs. Unsurprisingly, one such number cannot describe a real complex network; later studies focused on degree distributions, path lengths, node correlations or quantities describing percolation, epidemic spread, and many others. Although the clustering coefficient(s) gradually ceased to be interesting to measure, databases such as Stanford Large Network Dataset Collection [10] still provide it, along the number of triangles, as a fundamental graph statistic. Indeed, our paper advocates this measure as a simple and fast (especially that often it is precalculated) benchmark for assessing the significance of community structure.

Since there has been some confusion about what precisely is the clustering coefficient, let us define it: by *local clustering coefficient* C_l of node l we mean

$$C_l = N_{l,\Delta} / \binom{k_l}{2}, \quad (1)$$

where $N_{l,\Delta}$ is the number of triangles formed by the edges (l,m) , (l,n) and (m,n) , where m and n are nodes adjacent to l ; k_l is the degree of l (i.e., the number of edges incident to l), and the denominator is the number of all pairs of edges incident to node l .

The average or *mean local clustering coefficient* often found in the literature is simply

$$C = \frac{1}{N} \sum_{l=1}^N C_l, \quad (2)$$

where N is the size of the network (i.e., the number of nodes). As argued by Bollobás [11], it “is often not very informative”, therefore we do not use it further. Rather, we choose to work with what we will refer to as *global clustering coefficient* (GCC), customarily defined as

$$C = \frac{3 \times \text{number of triangles}}{\text{number of pairs of adjacent edges}}. \quad (3)$$

These definitions have been further extended to the cases of directed [12] and weighted [13], [14] networks.

As mentioned above, these coefficients have been used as a way of comparison with Erdős-Rényi graphs, for which it is known to be

$$C_{ER} = \frac{\langle k \rangle}{N-1}, \quad (4)$$

where $\langle \cdot \rangle$ denotes averaging, and $\langle k \rangle$ is simply the average degree. For uncorrelated graphs with a given degree distribution the formula [15]–[18]

$$C_{uc} \approx \frac{1}{N} \frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle^3} \quad (5)$$

is widely used.

As an example of a scale-free network, Bollobás [11] derives and proves a corrected (after [19]) estimate on the clustering coefficient for Barabási-Albert (BA) networks

$$C_{BA} = \frac{m-1}{8} \frac{(\log N)^2}{N}, \quad (6)$$

where m is the number of edges that a new node appears with in the process of network growth. We use BA graphs in Fig. 3 as an example of unclustered scale-free networks to show behavior of various estimates of C .

IV. OBSERVATIONS

A. GCC vs Mixing Parameter

The critical observation is that for ensembles of benchmark graphs with tunable strength of community structure the theoretically predicted limit of detectability coincides with the region where GCC saturates, as shown in Fig. 1.

Each of the points in the plot represents a single graph. There are two benchmark ensembles used: Newman-Girvan (NG) [20], whose graphs consist of 4 equally-sized communities (ER type with the probability of intercommunity linking a pair of nodes p_{out} and intracommunity linking probability p_{in}),

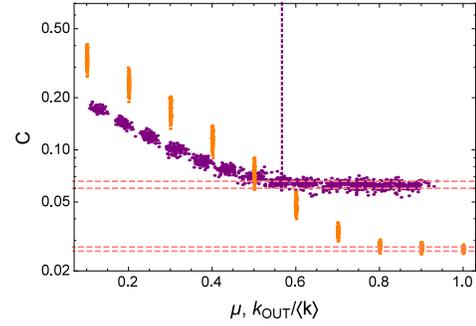


Fig. 1. Log plot of GCC versus the mixing parameter (μ for LFR, and $k_{OUT}/\langle k \rangle$ for NG networks; the ensembles are described in text). Dashed horizontal lines indicate mean \pm standard deviation of GCC for respective samples of random graphs. The dotted vertical line shows the theoretical detectability limit for NG graphs as given by [6]. The undetectability regime overlaps with the region where GCC curves saturate.

mean degree $\langle k \rangle = 16$, and $N = 256$ nodes; Lancichinetti-Fortunato-Radicchi (LFR) [21], with $\langle k \rangle = 20$, maximal degree $k_{MAX} = 50$, power-law degree distribution with exponent $\gamma = 2$, power-law community size distribution with exponent $\gamma_c = 1$, allowed range of community sizes 20 – 100 nodes, and the network size $N = 1000$ (we retain the parameters used originally in [22] for the sake of comparison). In the case of LFR graphs, the parameter quantifying community strength is the mixing parameter μ , which is a fraction of links of a node pointing outside of its community. In the case of NG graphs, this role is taken by $k_{OUT}/\langle k \rangle$, where k_{OUT} is the number of links pointing outside of a node’s community; it is simpler than μ , since the degree distribution is well peaked around its mean value. For each mixing parameter value, 100 networks were sampled (clouds of points visible in the plot).

For these ensembles, the fully random connections form for $k_{OUT}/\langle k \rangle \approx 12/16$ for NG and $\mu = 1$ for LFR. In both cases, community detection algorithms fail much earlier [22]. The theoretical position of detectability limit as calculated in [6] is in this case $k_{OUT}/\langle k \rangle \approx 9/16$; unfortunately, the calculations for LFR benchmark are much harder to obtain.

The reason why it was possible to find the detectability limit theoretically, is the fact that we already know the structure of NG networks. In the limit of large network size and, what is more important, large average degree, Nadakuditi et al. have shown how to calculate the eigenspectra of adjacency matrices of stochastic block model [6], configuration model [7], and a generalization of the two [8], which allows to find the detectability limits for networks with arbitrary degree distributions and community structure. Nevertheless, in practice we do not know the structure in the first place, hence we do not know the full eigenspectrum, and so we cannot decide whether a given network has any communities within detectability regime or not. Now, let us try to deduce that from the value of GCC.

B. GCC vs Eigenspectrum

The reason that the global clustering coefficient can be an index of detectability is fairly straightforward. Let us rewrite

(3) as follows

$$C = \frac{\sum_{l=1}^N N_{l,\Delta}/N}{(\langle k^2 \rangle - \langle k \rangle)/2} = \frac{\sum_{l=1}^N A_{ll}^3/N}{\langle k^2 \rangle - \langle k \rangle} = \frac{\sum_{l=1}^N \lambda_l^3/N}{\langle k^2 \rangle - \langle k \rangle}, \quad (7)$$

where A_{ij} is an element of the adjacency matrix \mathbf{A} , λ_i is its eigenvalue, and where we have used $\sum_{l=1}^N \binom{k_l}{2} = N(\langle k^2 \rangle - \langle k \rangle)/2$ and $N_{l,\Delta} = A_{ll}^3/2$. The second equality in (7) comes from the fact that $(\mathbf{A}^n)_{ij}$ encodes the number of paths of length n between nodes i and j ; the third equality comes from the fact that thanks to the cyclic property of the trace of a matrix, it is invariant to orthogonal transformations.

Now, the simple fact that clustering coefficients are connected to the third moment of the eigenspectrum was known [23]. Yet, as explained in Sec. II, in undirected and unweighted networks communities appear as isolated eigenvalues to the right of the bulk of the spectrum. This means that the spectrum becomes more skewed to the right; in other words, the third moment of the spectrum—and consequently GCC—increases. Such a behavior is consistent with the observation in Sec. IV-A.

C. Estimating Baseline GCC

The forthcoming issue is how to calculate the reference value of GCC, i.e., the value for a network without communities. One straightforward approach is to randomize the network under study, while keeping constant other of its characteristics that can affect GCC, mainly the degree distribution (or even degree sequence). Notwithstanding, this method is computationally costly, since a whole ensemble of randomized graphs is needed to make the comparison with the original empirical network meaningful.

Another approach is to use a theory to reproduce the eigenspectrum of graphs with a given degree distribution, calculate its third moment, and obtain GCC. Indeed, it can be accomplished, e.g., following [8]. Under assumptions on the community structure, e.g., as obtained from a community detection algorithm, the isolated eigenvalues can also be separately calculated. For large scale-free graphs, however, solving numerically the resulting equation for a specified degree sequence can become increasingly infeasible, as the number of terms (and solutions) also increases with the number of distinct degrees in the network.

More importantly, and to our demise, the bulk of the spectrum thus obtained is strictly symmetric. Consequently, the whole third spectral moment of a random graph will come from the largest eigenvalue. Unfortunately, we observed that for ER, BA networks, and even Watts-Strogatz [9] networks with sufficiently high rewiring probability the third moment of the bulk is negative¹, see Fig. 2. Consequently, for a wide range of real-world networks we can expect

$$C < \frac{1}{N} \frac{\lambda_1^3}{\langle k^2 \rangle - \langle k \rangle} \quad (8)$$

¹It should be noted that in graphs which are bipartite, or even partly so—e.g., WS graphs with small p —negative isolated eigenvalues can appear and thwart our efforts. Hence, we assume only positive isolated eigenvalues.

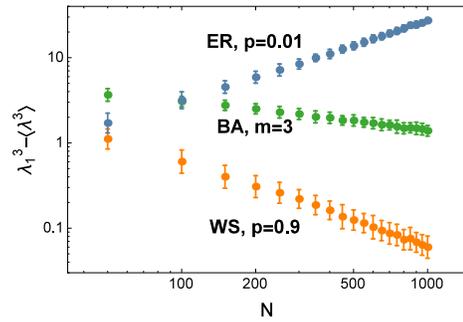


Fig. 2. Log-log plot of minus third moment of the bulk of eigenspectrum for Barabási-Albert, Erdős-Rényi, and Watts-Strogatz networks with the given parameters. Error bars are standard deviations for samples of 100 graphs each.

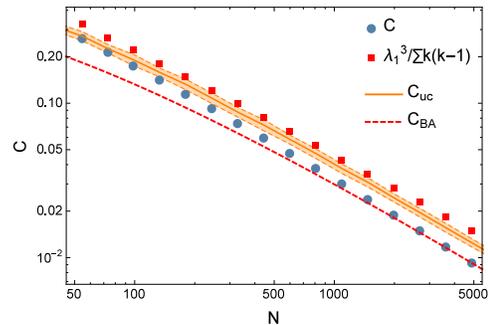


Fig. 3. Comparison of global clustering coefficients C (average over a sample of 100 BA graphs; standard deviations are of the size of the points), theoretical predictions obtained for uncorrelated networks with a given degree distribution C_{uc} (5) (shaded orange region), for BA (6) (red dashed line), and the contribution of the largest eigenvalue to GCC (red squares). When the network structure is unknown, the predictions are overestimated by more than the sample error of GCC.

to hold. (This inequality can also be seen as a lower bound for $\lambda_1 > \sqrt[3]{\text{number of triangles}}$ which is usually stronger than the known bound $\lambda_1 > \sqrt[3]{2 \max_l N_{l,\Delta}}$.)

It follows that the theoretical prediction on GCC as calculated using [8] is largely overestimated (i.e., an order of magnitude larger than the standard deviation of an ensemble with a given degree distribution). An example of such a behavior is shown for BA networks in Fig.3. The only theoretical prediction that could be useful for our purposes is (6); it is, however, a prediction specifically derived for this particular type of network. When analyzing real-world graphs, we lack such conveniences.

V. RESULTS

At this point we ought to recapitulate our aims, collect the pieces of information we have, and formulate the assumptions under which we will be able to proceed.

The aim is: based on the global clustering coefficient (3) to construct a criterion that decides whether

- 1) a given network has a significant community structure,
- 2) a given partition of the network is statistically significant,
- 3) a given community of the partition is statistically significant.

In this paper we will propose an answer to 1) and validate it, as well as suggest a way to develop a criterion for 2).

Let us list the (empirical) facts we have gathered up to this point:

- 1) GCC is proportional to the third moment of the eigen-spectrum,
- 2) (empirical) for graphs with a community structure GCC is greater than for random graphs, but it reaches the RG value before its connectivity is fully random, see Fig. 1,
- 3) (empirical) GCC is strictly lower than what can be predicted by the largest eigenvalue alone, see (8),
- 4) GCC for the uncorrelated graphs can be approximated by (5).

After GCC, the degree distribution and its two moments, $\langle k \rangle$ and $\langle k^2 \rangle$, have been calculated, which is the only computational burden, we may proceed further. One has to keep in mind, nevertheless, following assumptions:

- 1) the network is not bipartite, see Footnote 1.,
- 2) the network is uncorrelated, so that (5) is legitimate,
- 3) $\lambda_1 \geq \langle k^2 \rangle / \langle k \rangle - 1$,
- 4) N is large, as always.

Let us note that for the Poisson degree distribution the Assumption 3) is always true, since $\lambda_1 \geq \sqrt{\langle k^2 \rangle} = \sqrt{\langle k \rangle^2 + \langle k \rangle} > \langle k \rangle = \langle k^2 \rangle / \langle k \rangle - 1$, which comes from Rayleigh's inequalities (see e.g. [24]). For scale-free networks, however, the second inequality is inverted and there may appear graphs breaking Assum. 3). (We did find such instances for BA networks but, surprisingly, not for LFR which are scale-free as well.)

A. Procedure for an unpartitioned graph

In the case of Aim 1), the way to proceed is embarrassingly simple: if

$$C \leq C_{uc} \quad (9)$$

is true, then the network is outside the detectability regime, i.e., there are no significant communities.

A remark is necessary: let us notice that the above equation is a conservative criterion. The condition that we should have checked is in fact (8). But it can be easily checked that the approximation C_{uc} corresponds to $\lambda_1 \approx \langle k^2 \rangle / \langle k \rangle - 1$.

However, since the true largest eigenvalue is unknown, we can utilize the well-known approximation by Chung et al. [25],

$$\lambda_1 = (1 + o(1)) \max \left\{ \frac{\langle k^2 \rangle}{\langle k \rangle}, \sqrt{k_{MAX}} \right\}, \quad (10)$$

which in most cases we are interested in yields $\langle k^2 \rangle / \langle k \rangle$. Consequently, the criterion (8) should be given by

$$C < \frac{1}{N} \frac{\langle k^2 \rangle^3}{\langle k \rangle^3 (\langle k^2 \rangle - \langle k \rangle)}. \quad (11)$$

By now, it should be clear why the Assumption 3) was listed before: as long as it is fulfilled the criterion given in (9) is stronger than (8).

Though (10) is proved to be the upper bound of λ_1 , in Fig. 3 the approximation would be even higher than the one

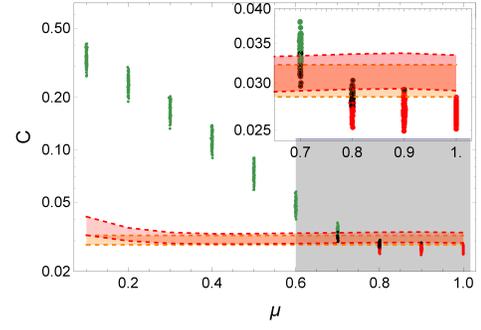


Fig. 4. Log plot of GCC for LFR $N = 1000, B$ benchmark, as in Fig. 1. The red shaded strip shows the contribution of true λ_1 to C ; the orange shaded strip shows the region between conditions (9) and (11). The red dots represent graphs fulfilling (9); the black dots fulfill (11), but not (9); the green dots fulfill neither. The grey region indicates where all known community detection methods gradually begin to fail. The inset magnifies the region of (un)detectability transition.

resulting from true value of λ_1 (red squares). This means that in some cases the factor $o(1)$ is of a magnitude which can affect our inference. Finally, the ordering of the theoretical predictions, as depicted in Fig. 3, will be of use to us.

Even though, condition (9) does not provide us with any error estimates, we still believe that the approximate position of λ_1 is given by (10). As a result we can use (11) as an upper bound for an error estimate. Thus, the rule-of-thumb is as follows:

- if condition (9) is true, there is no detectable community structure,
- if condition (9) is false, but (11) is true, either there is no detectable community structure or there are some detectable and some undetectable communities,
- if both are false, there is some detectable community structure.

To validate above criteria we again resort to LFR benchmark graphs, since all of the state of the art community detection methods have been tested on them, including Infomap [26], the Louvain method [27] or OsloM [28]. Tests on these graphs, with exactly the same parameters, have shown [2], [22] that all known methods begin to decrease their accuracy at $\mu \approx 0.6$, and fail after crossing $\mu \approx 0.7 - 0.8$. As illustrated in Fig. 4, this is in accord with the criteria given above.

To address the Aim 2) above, let us assume that we already have obtained a partition of the graph into q clusters. For instance, it might be a result of any of the algorithms we have mentioned, or it might be the planted partition, e.g., in LFR benchmarks (note that by construction these graphs have a partition planted even for $\mu = 1$). Following [8], it is possible to calculate the q discrete eigenvalues that are theoretically predicted to appear in such a partitioned graph. We are then able to compute their contribution to C in (3), just as we did for λ_1 , and accordingly revise the criteria (9)-(11). Note, that instead of using approximation (10) for λ_1 in this Section, we may use [8] for the largest eigenvalue only. Depending on the computational resources one has, it might be feasible

to simply calculate λ_1 numerically—although the theoretical approximations seem sufficient, an improvement on the error estimate of roughly 20% can be expected (at least on graphs akin to LFR), as can be seen in the inset of Fig. 4.

VI. CONCLUSION

We have proposed a simple method to deduce whether it is possible to detect communities in a given network. Since the only computations needed are: calculating triangles and the degree distribution, we claim this method is fast and computationally inexpensive. The state of the art algorithm for an exact triangle enumeration runs in $O(E^{1.41})$, where E is the number of edges in the graph. Faster approximate algorithms do exist [29], whose accuracy of around 95% are sufficient, although introduce an additional error. Although there exist algorithms of community detection linear in time, assessing statistical significance of the partitions obtained usually involves bootstrapping which is not favorable.

The error bound in procedure we suggested is rather heuristic; however, the approximations of λ_1 (10) and GCC (5) can be improved, e.g., if we have ansatz on degree-degree correlations, which in some cases are known [30]. Still, if the assumption 3) holds, the criteria are conservative and on the safe side.

The natural extension of the scheme presented here is: directed and weighted networks, for which generalized versions of clustering coefficients have been proposed [12]–[14]. Initial computations on directed LFR benchmarks show that analogous criteria are plausible. For weighted networks, however, the clustering coefficients behave nonmonotonously as a function of the mixing parameter, which introduces additional complications.

Further systematic comparison with other methods is needed and, more importantly, attacking real-world networks.

ACKNOWLEDGMENT

My thanks go to Santo Fortunato for the discussions at the early stage of the study and a week's stay at BECS, Aalto University and to Zdzisław Burda for consultations throughout. I also thank an Anonymous Reviewer for useful insights about bounds for the principal eigenvalue. The work has been funded by Grant No. DEC-2013/09/N/ST6/01419 of the National Science Centre of Poland.

REFERENCES

- [1] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 35, pp. 75 – 174, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0370157309002841>
- [2] A. Lancichinetti and S. Fortunato, "Consensus clustering in complex networks," *Scientific reports*, vol. 2, 2012.
- [3] A. Lancichinetti, F. Radicchi, and J. J. Ramasco, "Statistical significance of communities in networks," *Phys. Rev. E*, vol. 81, p. 046110, Apr 2010. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.81.046110>
- [4] J. Reichardt and M. Leone, "(Un)detectable cluster structure in sparse networks," *Phys. Rev. Lett.*, vol. 101, p. 078701, Aug 2008. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.101.078701>

- [5] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Inference and phase transitions in the detection of modules in sparse networks," *Phys. Rev. Lett.*, vol. 107, p. 065701, Aug 2011. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.107.065701>
- [6] R. R. Nadakuditi and M. E. J. Newman, "Graph spectra and the detectability of community structure in networks," *Phys. Rev. Lett.*, vol. 108, p. 188701, May 2012. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.108.188701>
- [7] —, "Spectra of random graphs with arbitrary expected degrees," *Phys. Rev. E*, vol. 87, p. 012803, Jan 2013. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.87.012803>
- [8] X. Zhang, R. R. Nadakuditi, and M. E. J. Newman, "Spectra of random graphs with community structure and arbitrary degrees," *Phys. Rev. E*, vol. 89, p. 042816, Apr 2014. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.89.042816>
- [9] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [10] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," <http://snap.stanford.edu/data>, Jun. 2014.
- [11] B. Bollobás and O. M. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of graphs and networks: from the genome to the internet*, S. Bornholdt and H. G. Schuster, Eds. Weinheim: Wiley-VCH, 2003, pp. 1–34.
- [12] G. Fagiolo, "Clustering in complex directed networks," *Phys. Rev. E*, vol. 76, p. 026107, Aug 2007. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.76.026107>
- [13] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 11, pp. 3747–3752, 2004.
- [14] J. Saramäki, M. Kivela, J.-P. Onnela, K. Kaski, and J. Kertész, "Generalizations of the clustering coefficient to weighted complex networks," *Phys. Rev. E*, vol. 75, p. 027105, Feb 2007. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.75.027105>
- [15] M. E. J. Newman, "Random graphs as models of networks," in *Handbook of graphs and networks: from the genome to the internet*, S. Bornholdt and H. G. Schuster, Eds. Weinheim: Wiley-VCH, 2003, pp. 35–68.
- [16] M. Boguñá and R. Pastor-Satorras, "Class of correlated random networks with hidden variables," *Phys. Rev. E*, vol. 68, p. 036112, Sep 2003. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.68.036112>
- [17] Z. Burda, J. Jurkiewicz, and A. Krzywicki, "Network transitivity and matrix models," *Phys. Rev. E*, vol. 69, p. 026106, Feb 2004. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.69.026106>
- [18] S. N. Dorogovtsev, "Clustering of correlated networks," *Phys. Rev. E*, vol. 69, p. 027104, Feb 2004. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.69.027104>
- [19] K. Klemm and V. M. Eguíluz, "Growing scale-free networks with small-world behavior," *Phys. Rev. E*, vol. 65, p. 057102, May 2002. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.65.057102>
- [20] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [21] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Phys. Rev. E*, vol. 78, p. 046110, Oct 2008. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.78.046110>
- [22] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E*, vol. 80, p. 056117, Nov 2009. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.80.056117>
- [23] I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek, "Spectra of 'real-world' graphs: Beyond the semicircle law," *Phys. Rev. E*, vol. 64, p. 026704, Jul 2001. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.64.026704>
- [24] P. Van Mieghem, *Graph spectra for complex networks*. Cambridge: Cambridge University Press, 2010.
- [25] F. Chung, L. Lu, and V. Vu, "Spectra of random graphs with given expected degrees," *Proceedings of the National Academy of Sciences*, vol. 100, no. 11, pp. 6313–6318, 2003.
- [26] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.
- [27] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks,"

Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, p. P10008, 2008. [Online]. Available: <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008>

- [28] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding statistically significant communities in networks," *PLoS ONE*, vol. 6, no. 4, p. e18961, 04 2011. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0018961>
- [29] C. E. Tsourakakis, M. N. Kolountzakis, and G. L. Miller, "Approximate triangle counting," *CoRR*, vol. abs/0904.3761, 2009. [Online]. Available: <http://arxiv.org/abs/0904.3761>
- [30] P. L. Krapivsky and S. Redner, "Organization of growing random networks," *Phys. Rev. E*, vol. 63, p. 066123, May 2001. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.63.066123>