

# Rare events statistics of random walks on networks: localization and other dynamical phase transitions

Caterina De Bacco,<sup>1</sup> Alberto Guggiola,<sup>2</sup> Reimer Kühn,<sup>3</sup> and Pierre Paga<sup>3</sup>

<sup>1</sup>*LPTMS, Centre National de la Recherche Scientifique et Université Paris-Sud 11, 91405 Orsay Cedex, France.*

<sup>2</sup>*LPTENS, Unité Mixte de Recherche (UMR 8549) du CNRS et de l'ENS, associée à l'UPMC Université Paris 06, 24 Rue Lhomond, 75231 Paris Cedex 05, France.*

<sup>3</sup>*Department of Mathematics, King's College London*

Rare event statistics for random walks on complex networks are investigated using the large deviations formalism. Within this formalism, rare events are realized as typical events in a suitably deformed path-ensemble, and their statistics can be studied in terms of spectral properties of a deformed Markov transition matrix. We observe two different types of phase transition in such systems: (i) rare events which are singled out for sufficiently large values of the deformation parameter may correspond to *localized* modes of the deformed transition matrix; (ii) “mode-switching transitions” may occur as the deformation parameter is varied. Details depend on the nature of the observable for which the rare event statistics is studied, as well as on the underlying graph ensemble. In the present letter we report on the statistics of the average degree of the nodes visited along a random walk trajectory in Erdős-Rényi networks. Large deviations rate functions and localization properties are studied numerically. For observables of the type considered here, we also derive an analytical approximation for the Legendre transform of the large-deviations rate function, which is valid in the large connectivity limit. It is found to agree well with simulations.

Random walks are dynamical processes widely used to analyze, organize or perform important tasks on networks such as searches [1, 2], routing or data transport [3–5]. Their popularity is due to their cheap implementation, as they rely only on local information, such as the state of the neighborhood of a given node of the network. This ensures network scalability and allows fast data transmission without the need for large storage facilities at nodes, such as big routing tables in communication networks. These features make random walks an efficient tool to explore networks characterized by a high cost of information. Examples are sensor networks [6] where many signaling packets are needed to acquire wider networks status information. In peer-to-peer networks the absence of a central server storing file locations requires users to perform repeated local searches in order to find a file to download, and various random walk strategies have been proposed as a scalable method [7–9] in this context. Less attention has been paid to characterize rare events associated with random walks on networks. Yet the occurrence of a rare event can have severe consequences. In hide-and-seek games for instance [10], rare events represent situations where the seeker finds either most (or unusually many) of the hidden targets, or conversely none (or unusually few). In the context of cyber-security, where one is concerned with worms and viruses performing random walks through a network, a rare event would correspond to a situation where unusually many sensible nodes are successfully attacked and infected, which may have catastrophic consequences for the integrity of an entire IT infrastructure. Characterizing the statistics of rare events for random walks in complex networks and its dependence on network topology is thus a problem of considerable technological importance. A variant of this problem was recently analyzed for biased random walks in complex networks [11]. That paper addressed rare fluctuations in single node occupancy for an ensemble of independent (biased) walkers in the stationary state of the

system. By contrast, our interest here is in rare event statistics of *path averages*, or equivalently of time integrated variables. Rare event statistics of this type has been looked at for instance in the context of kinetically constrained models of glassy relaxation [12]; relations to constrained ensembles of trajectories were explored in [13] for Glauber dynamics in the 1d Ising chain. While these studies were primarily concerned with the use of large deviations theory as a tool to explore dynamical phase transitions in homogeneous systems, our focus here is on the interplay between rare event statistics and the heterogeneity of the underlying system.

In the present Letter we use large deviations theory to study rare events statistics for path averages of observables associated with sites visited along trajectories of random walks. Within this formalism, rare events are realized as typical events in a suitably deformed path-ensemble [12, 14]. Their statistics can be studied in terms of spectral properties of a deformed version of the Markov transition matrix for the original random walk model, the relevant information being extracted from the algebraically largest eigenvalue of the deformed transition matrix. Such deformation may direct random walks to subsets of a network with vertices of either atypically high or atypically low coordination. It also amplifies the heterogeneity of transition matrix elements for large values of the deformation parameter and we observe that, as a consequence, the eigenvector corresponding to the largest eigenvalue of the deformed transition matrix may exhibit a *localization transition*, indicating that rare large fluctuations of path averages are typically realized by trajectories that remain localized on small subsets of the network. Within localized phases, we also encounter a second type of dynamical phase transition related to *switching between modes* as the deformation parameter used to select rare events is varied. Our methods allow us to study the role that network topology and heterogeneity play in selecting these special paths, as well as

to infer properties of paths actually selected to realize extreme events.

*The model.* We consider a complex network with adjacency matrix  $A$ , with entries  $a_{ij} = 1$  if the edge  $(ij)$  exists,  $a_{ij} = 0$  otherwise. The transition matrix  $W$  of an unbiased random walk has entries  $W_{ij} = a_{ij}/k_j$  where  $k_j$  is the degree of node  $j$  and  $W_{ij}$  is the probability of a transition from  $j$  to  $i$ .

Writing  $\mathbf{i}_\ell = (i_0, i_1, \dots, i_\ell)$  a path of length  $\ell$ , quantities of interest are empirical path-averages of the form

$$\hat{\phi}_\ell = \frac{1}{\ell} \sum_{i=1}^{\ell} \xi_i, \quad (1)$$

where the  $\xi_i$  are quenched random variables associated with the vertices  $i = 1, \dots, N$  of the graph, which could be independent of, be correlated with, or be deterministic functions of the degrees  $k_i$  of the vertices. It is expected that the  $\hat{\phi}_\ell$  are for large  $\ell$  sharply peaked about their mean

$$\bar{\phi}_\ell = \frac{1}{\ell} \sum_{\mathbf{i}_\ell} P(\mathbf{i}_\ell) \sum_{i=1}^{\ell} \xi_i = \left\langle \frac{1}{\ell} \sum_{i=1}^{\ell} \xi_i \right\rangle \quad (2)$$

where  $P(\mathbf{i}_\ell)$  denotes the probability of the path  $\mathbf{i}_\ell$ .

The average (2) can be obtained from the *cumulant generating function*  $\psi_\ell(s) = \ell^{-1} \ln \sum_{\mathbf{i}_\ell} P(\mathbf{i}_\ell) e^{s \sum_{i=1}^{\ell} \xi_i}$  as  $\bar{\phi}_\ell = \psi'_\ell(s)|_{s=0}$ . Here, we are interested in rare events, for which the empirical averages  $\hat{\phi}_\ell$  take values  $\phi$  which differ significantly from their mean  $\bar{\phi}_\ell$ . Large deviations theory predicts that for  $\ell \gg 1$  the probability density  $P(\phi)$  for such an event scales exponentially with path-length  $\ell$ ,  $P(\phi) \sim e^{-\ell I(\phi)}$ , with a *rate function*  $I(\phi)$  which, according to the Gärtner-Ellis theorem [14] is obtained as a Legendre transform  $I(\phi) = \sup_s \{s\phi - \psi(s)\}$  of the limiting cumulant generating function  $\psi(s) = \lim_{\ell \rightarrow \infty} \psi_\ell(s)$ , provided that this limit exists and that it is differentiable. We shall see that the second condition may be violated, and that the derivative  $\psi'(s)$  may develop discontinuities at certain  $s$ -values, entailing that we observe regions where  $I(\phi)$  is strictly linear and only represents the convex hull of the true rate function [14].

In order to evaluate  $\psi_\ell(s)$ , we express path probabilities using the Markov transition matrix  $W$  and a distribution  $\mathbf{p}_0 = (p_0(i_0))$  of initial conditions as  $P(\mathbf{i}_\ell) = [\prod_{i=1}^{\ell} W_{i_{i-1}i_i}] p(i_0)$ , entailing that  $\psi_\ell(s)$  can be evaluated in terms of a deformed transition matrix  $W(s) = (e^{s\xi_i} W_{ij})$  as  $\psi_\ell(s) = \ell^{-1} \ln \sum_{i_0} [W^\ell(s)]_{i_0 i_0} p(i_0)$ . Using a spectral decomposition of the deformed transition matrix one can write this as

$$\psi_\ell(s) = \ln \lambda_1 + \frac{1}{\ell} \ln \left[ (\mathbf{1}, \mathbf{v}_1)(\mathbf{w}_1, \mathbf{p}_0) + \sum_{\alpha(\neq 1)} \left( \frac{\lambda_\alpha}{\lambda_1} \right)^\ell (\mathbf{1}, \mathbf{v}_\alpha)(\mathbf{w}_\alpha, \mathbf{p}_0) \right]. \quad (3)$$

Here the  $\lambda_\alpha$  are eigenvalues of  $W(s)$ , the  $\mathbf{v}_\alpha$  and  $\mathbf{w}_\alpha$  are the corresponding right and left eigenvectors,  $\mathbf{1} = (1, \dots, 1)$ , and the bracket notation  $(\cdot, \cdot)$  is used to denote an inner product. Eigenvalues are taken to be sorted in decreasing order  $\lambda_1 \geq |\lambda_2| \geq |\lambda_3| \dots \geq \lambda_N$ , with the first inequality being a

consequence of the Perron-Frobenius theorem [15]. This concludes the general framework. For the remainder of this Letter, we will restrict our attention to the case where  $\xi_i = f(k_i)$ .

For long paths, the value of the cumulant generating function is dominated by the leading eigenvalue  $\lambda_1 = \lambda_1(s)$  of the transition matrix  $W(s)$ , so  $\psi(s) = \log \lambda_1(s)$ . In the  $s = 0$  case, the eigenvalue problem is trivial, as the column-stochasticity of the transition matrix yields a left eigenvector  $w_i \equiv 1$  corresponding to the maximal eigenvalue  $\lambda_1 = 1$ . The associated right eigenvector is  $v_i \propto k_i$ . For nonzero  $s$ , such closed form expressions are in general not known. Performing a direct matrix diagonalization is quite daunting for large system sizes  $N$ , even if one exploits methods that calculates only the first eigenvalue [16]. Hence we are interested in fast viable approximations. Here we describe one such approximation expected to be valid for networks in which vertex degrees are typically large.

*Degree-based approximation.* We start by considering the left eigenvectors  $\mathbf{w}$  instead of the right eigenvectors, for which the eigenvalue equation can be written as

$$\lambda w_j = \frac{1}{k_j} \sum_{i \in \partial j} w_i e^{s f(k_i)}. \quad (4)$$

This system of equations can be simplified by considering a degree-based approximation for the first eigenvector, where one assumes that the values of  $w_i$  only depend on the degree of the node  $i$ :  $w_i = w(k_i)$ . If the average degree is large enough and the degree distribution is not too heterogeneous, we can write the eigenvalue equation (4) by appeal to the law of large numbers as

$$\lambda_1(s) w(k) = \sum_{k'} P(k'|k) w(k') e^{s f(k')} \quad (5)$$

where  $P(k'|k)$  is the probability for the neighbor of a node of degree  $k$  to have degree  $k'$ .

In an Erdős-Rényi (ER) ensemble [17], and more generally in any configuration model ensemble, we have  $P(k'|k) = P(k') \frac{k}{\langle k \rangle}$ . In this case the right-hand side of (5) does not depend on  $k$  and the  $w(k)$  are in fact  $k$ -independent. The eigenvalue equation then simplifies to

$$\lambda_1(s) = \left\langle \frac{k}{\langle k \rangle} e^{s f(k)} \right\rangle, \quad (6)$$

where the average is over the degree distribution  $P(k)$ . This approximation yields excellent results for large mean connectivities  $c = \langle k \rangle$  on ER graphs, and more generally for configuration models without low degree nodes. This is illustrated in figure 1, where we plot a comparison with numerical simulations for ER graphs with  $c = 30$ . In figure 1 and throughout the remainder of the paper simulation results are obtained as averages over 1000 samples.

*Eigenvector localization.* Because of the heterogeneity of the underlying system, one finds the random walk transition matrix to exhibit localized states, both for fast and slow relaxation modes [18], even in the undeformed system, although

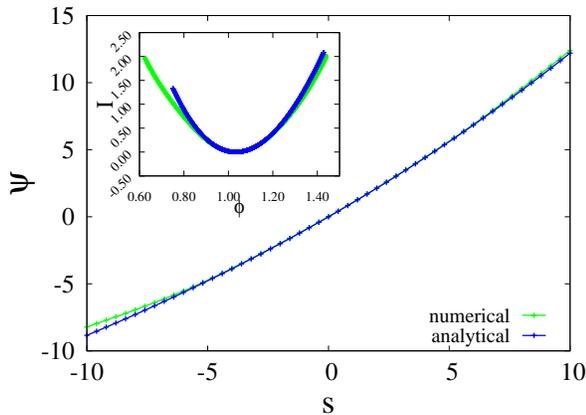


FIG. 1. (Colour online) Cumulant generating function  $\psi(s)$  for ER networks with  $c = 30$  and  $f(k_i) = k_i/c$ , comparing the large-degree approximation (6) (blue line) with results of a numerical simulation (green line). The inset shows the corresponding rate functions.

the eigenvector corresponding to the largest eigenvalue (the equilibrium distribution) is typically delocalized. However, given the nature of the deformed transition matrix, one expects the deformed random walk for large  $|s|$  to be localized around vertices where  $sf(k_i)$  is very large; hence we anticipate that in the deformed system, even the eigenvector corresponding to the largest eigenvalue *may* become localized for sufficiently large  $|s|$ . In order to investigate this effect quantitatively we look at the inverse participation ratio of the eigenvector corresponding to the largest eigenvalue  $\lambda_1$  of  $W(s)$ . Denoting by  $v_i$  its  $i$ -th component, we have

$$\text{IPR}[\mathbf{v}] = \frac{\sum_i v_i^4}{\left[\sum_i v_i^2\right]^2} \quad (7)$$

One expects  $\text{IPR}[\mathbf{v}] \sim N^{-1}$  for a delocalized vector, whereas  $\text{IPR}[\mathbf{v}] = O(1)$  if  $\mathbf{v}$  is localized.

*Results on random graphs.* We performed numerical simulations to evaluate  $\lambda_1(s)$  and the  $\text{IPR}[\mathbf{v}_1(s)]$  for several types of network, defined by their random graph topology. In the present letter we restrict ourselves to discussing results for ER networks. We found that other network ensembles such as scale-free random graphs give qualitatively similar results; we will report on these in an extended version of this letter.

We looked at various examples for the function  $f(k_i)$  but in the present letter we only report results for the normalized degree  $f(k_i) = k_i/c$ ; other deterministic types of degree-dependent functions exhibit similar behavior, thus focusing on the normalized degree is sufficient to capture the important aspects of this problem. We restrict our simulations to the largest (giant) component of the graphs, in order to prevent spurious effects of isolated nodes or small disconnected clusters (e.g. dimers) dominating  $\lambda_1(s)$  and the IPR for negative  $s$ , as these would represent trivial instances of rare events, where a walker starts, and is thus stuck on a small disconnected component of the graph. From here on, the network size given

must be understood as the size of the networks from which the giant component is extracted.

Fig. 2 shows the existence of two localized regimes for sufficiently large values of  $|s|$ , with IPRs on the localized side of both transitions increasing with system size. Results can be understood, as for large  $|s|$  the deformed random walk is naturally attracted to the nodes with the largest (resp. smallest) degrees for positive (resp. negative)  $s$ . Thus for large negative  $s$  the deformed walk tends to be concentrated at the end of the longest dangling chain, whereas for large positive  $s$  it will be concentrated at the site with the largest available coordination. On an ER network where the large-degree tail of the degree distribution decays very fast, such a high degree vertex is likely to be connected to vertices whose degrees are lower, even significantly lower, than that of the highest degree vertex in the network, which leads to IPRs approaching 1 in the large  $N$  limit. Conversely, for negative  $s$ , the deformed random walk will be attracted to the ends of dangling chains in the network, with the probability of escape from a chain decreasing with its length (with the length of the longest dangling chain increasing with system size). This can explain that IPRs initially saturate at  $1/2$  for large systems. Only upon further decreasing  $s$  to more negative values will the asymmetry of the deformed transition matrices, to and away from the end of a dangling chain, induce that further weight of the dominant eigenvector to become concentrated on the end-site, leading to a further increase of the IPR.

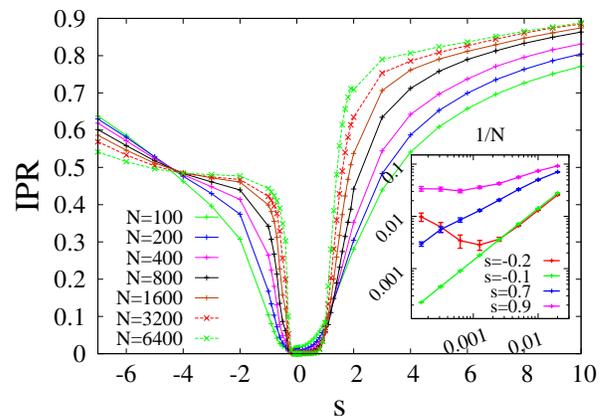


FIG. 2.  $\text{IPR}[\mathbf{v}]$  as a function of the deformation parameter  $s$  for ER graphs with  $c = 6$ , and  $f(k_i) = k_i/c$ . The inset exhibits the  $N^{-1}$ -scaling of IPRs for 4 different values of the deformation parameter  $s$ , chosen in pairs on either side of *two* localization transitions, one at negative, and one at positive  $s$ .

From the values of  $\lambda_1(s)$  we also derived the large deviation rate functions for path averages of the normalized degree  $f(k_i) = k_i/c$ , for various systems sizes and average connectivities. In fig. 3 we report  $I(\phi)$  for an ER network at a low connectivity of  $c = 3$ . While the right branch of  $I(\phi)$  is for large  $N$  well approximated by a parabola, our results show the emergence of a linear region on the left branch, which

becomes more pronounced as the system size is increased. This is a signature of a non-differentiable point of  $\psi(s)$  at a point  $s^*$  estimated to be at  $s^* = -0.060 \pm 0.002$ : at this point the Gärtner-Ellis theorem cannot be used to evaluate the rate function, and the linear branch only represents the convex envelope of the true  $I(\phi)$  [14]. The latter can either coincide with its convex envelope, or it can indeed be non-convex. However this information cannot be accessed by the theorem. The emergence of a jump-discontinuity in  $\psi'(s)$  is due to a level crossing of the two largest eigenvalues, where the system switches between two modes that correspond to the largest eigenvalue on either side of  $s^*$ . In finite systems the crossing is an ‘avoided crossing’ due to level repulsion, but the two largest eigenvalues become asymptotically degenerate at  $s^*$  in the  $N \rightarrow \infty$  limit, leading to a divergence of the correlation length  $\xi(s) = [\ln(\lambda_1(s)/\lambda_2(s))]^{-1}$  at  $s^*$ , in close analogy with phenomenology of second order phase transitions, the divergence being logarithmic in  $N$  in the present case.

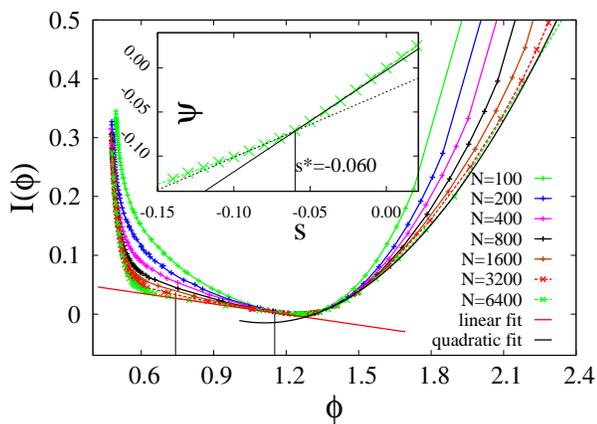


FIG. 3. Rate function  $I(\phi)$  for ER graphs with  $c = 3$ , and  $f(k_i) = k_i/c$  for system sizes ranging from  $N = 100$  to  $N = 6400$ . In the inset, we show  $\psi(s)$  in the vicinity of the non-differentiable point. For the largest system size, a linear fit of the convex envelope of the left branch and a quadratic fit of the right branch of  $I(\phi)$  are shown as well.

*Conclusions and future perspectives.* In this Letter we have analyzed rare events statistics for path averages of observables associated with sites visited along random walk trajectories on complex networks. Results are obtained by looking at spectral properties of suitably deformed transition matrices. The main outcome of our analysis is the possible emergence of two types of dynamical phase transitions in low mean degree systems: localization transitions which entail that large deviations from typical values of path averages may be realized by localized modes of a deformed transition matrix, and *mode-switching transitions* signifying that the modes (eigenvectors) in terms of which large deviations are typically realized may switch as the deformation parameter  $s$  and thus the actual scale of large deviations are varied. Results of numerical simulations consistently support these claims. We also developed an analytical approximation valid for networks in

which degrees are typically large.

Our work opens up the perspective to study a broad range of further interesting problems. On a technical level, one would want to implement more powerful techniques, such as derived in [19], to obtain the largest eigenvalue in the present problem class for larger system sizes. Then there is clearly the need to systematically study the dependence of the phenomena reported here on the degree statistics, and on the nature of the observables for which path averages are looked at. We have gone some way in this direction and will report results in an extended version of the present paper. In particular one might wish to look at observables which, rather than being deterministic functions of the degree, are only statistically correlated with the degree, or at observables taking values on *edges between* nodes [13, 14]. This could be of interest in applications such as traffic or information flows on networks subject to capacity constraints on edges. Moreover, given the nature of the mode-switching transition observed in the present letter, it is clearly conceivable that *several such transitions* could be observed in a single system, depending of course on the nature of the observables studied and on the topological properties of the underlying networks. Finally, critical phenomena associated with the localization transition and with mode-switching transitions also deserve further study. We believe that this list could go on.

This work was supported by the Marie Curie Training Network NETADIS (FP7, grant 290038).

- [1] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman. Search in power-law networks. *Phys. Rev. E*, 64:046135, 2001.
- [2] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas. Optimal network topologies for local search with congestion. *Phys. Rev. Lett.*, 89(24):248701, 2002.
- [3] S. D. Servetto and G. Barrenechea. Constrained random walks on random graphs: routing algorithms for large scale wireless sensor networks. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 12–21. ACM, 2002.
- [4] B. Tadić, S. Thurner, and G. J. Rodgers. Traffic on complex networks: Towards understanding global statistical properties from microscopic density fluctuations. *Phys. Rev. E*, 69(3):036102, 2004.
- [5] B. Tadić and S. Thurner. Information super-diffusion on structured networks. *Physica A: Stat. Mech. and its App.*, 332:566–584, 2004.
- [6] H. Tian, H. Shen, and T. Matsuzawa. Randomwalk routing for wireless sensor networks. In *Parallel and Distributed Computing, Applications and Technologies, 2005. PDCAT 2005. Sixth International Conference on*, pages 196–200. IEEE, 2005.
- [7] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker. Making gnutella-like p2p systems scalable. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 407–418. ACM, 2003.
- [8] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and replication in unstructured peer-to-peer networks. In *Proceed-*

- ings of the 16th international conference on Supercomputing, pages 84–95. ACM, 2002.
- [9] N. Bisnik and A. Abouzeid. Modeling and analysis of random walk search algorithms in p2p networks. In *Hot topics in peer-to-peer systems, 2005. HOT-P2P 2005. Second International Workshop on*, pages 95–103. IEEE, 2005.
- [10] K. Sneppen, A. Trusina, and M. Rosvall. Hide-and-seek on complex networks. *Europhys. Lett.*, 69(5):853, 2005.
- [11] V. Kishore, M. S. Santhanam, and R. E. Amritkar. Extreme events and event size fluctuations in biased random walks on networks. *Phys. Rev. E*, 85:056120, 2012.
- [12] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duivendijk, and F. van Wijland. First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *J. Phys. A*, 42:075007, 2009.
- [13] R. L. Jack and P. Sollich. Large deviations and ensembles of trajectories in stochastic models. *Progr. of Theor. Phys. Supplement*, 184:304–317, 2010.
- [14] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478(1):1–69, 2009.
- [15] F. R. Gantmacher. *Applications of the Theory of Matrices*. Interscience, New York, 1959.
- [16] Cornelius C. Lanczos. *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*. United States Governm. Press Office, 1950.
- [17] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–ñ61, 1960.
- [18] R. Kühn. Spectra of random stochastic matrices and relaxation in complex systems. *Europhys. Lett.*, 109:60003, 2015.
- [19] Y. Kabashima, H. Takahashi, and O. Watanabe. Cavity approach to the first eigenvalue problem in a family of symmetric random sparse matrices. *J. Phys. Conf. Ser.*, 233:012001, 2010.