Concentration and Stability of Community-Detecting Functions on Random Networks

Weituo Zhang and Chjan C. Lim

Abstract. We propose a general form of community-detecting functions for finding communities—an optimal partition of a random network—and examine the concentration and stability of the function values using the bounded difference martingale method. We derive LDP inequalities for both the general case and several specific community-detecting functions: modularity, graph bipartitioning, and q-Potts community structure. We also discuss the concentration and stability of community-detecting functions on different types of random networks: sparse and nonsparse networks and some examples such as ER and CL networks.

1. Introduction

One of the main problems in network study is finding communities, that is, an optimal partition for a given network. A standard approach is to design a community-detecting function of network partitions that achieves its extremum when the partition is optimal, or conversely, the optimality of the partition is defined by a reasonably designed community-detecting function.

There are several important applications of this approach. One is to find the community structures of social networks. A well-known community-detecting function for this application is modularity [Newman 06]. Another application is called circuit partitioning in designing a computer system [Kirkpatrick et al. 83], in which circuits must be partitioned into groups so that the number of signals crossing the partition boundaries is minimized. The community-detecting function used in this application is a variation of the function used in the graph-bipartitioning problem [Wiethege and Sherrington 87]. In this paper, we propose a general form of community-detecting functions for which the functions mentioned above and many others are considered as specific cases.

Previous studies have shown that optimizing a community-detecting function is in general an NP-complete problem [Fu and Anderson 86]. Algorithms such as simulated annealing have been developed and tested on specific networks [Kirkpatrick et al. 83]. However, in theoretical studies, it is nearly impossible to obtain an exact solution for a single network. What we usually have is the average value of the community-detecting function over a random network ensemble [Wiethege and Sherrington 87, Wong and Sherrington 87]. This gap between the computational and theoretical points of view makes it necessary to study the concentration and stability of community-detecting functions.

Roughly speaking, the stability of a community-detecting function means that a change in the function is small when the corresponding network structure undergoes a small perturbation; and concentration means that the theoretically predicted average value of the community-detecting function becomes more precise as the system size grows larger. In this paper, we derive LDP-type inequalities to illustrate both aspects: the fluctuation of a community-detecting function in a fixed network ensemble and its asymptotic behavior as the system size goes to infinity. Here the inequalities of LDP (large deviation principle) type [Sanov 57] refer to those controlling the probabilities of rare events by an exponentially decaying function whose exponent is related to the rate function.

The concentration and stability of a community-detecting function, which will be further discussed in Section 3, are important from the following point of view: First, sometimes it is helpful to consider a given network as a sample randomly selected from a designed network ensemble. The ensemble should be easy to analyze. It should capture important features of the given network such as the average degree or the degree distribution, but neglect detailed structural information by randomization. In this scenario, concentration and stability estimate the departure of the function value of the given network from the ensemble average.

Second, for real-world networks such as social networks and the Internet, we have problems such as lack of information, uncertainty of the environment, and changing of the network over time. These networks should essentially be

considered as network ensembles. Since this ensemble is usually hard to analyze, simulation results of specific networks are, in turn, taken to estimate the ensemble average. In this scenario, the concentration and stability come from the problem itself. For example, in circuit partitioning, if we have already found an optimal circuit partition under a given signal flow configuration, then concentration and stability lead us to conclude that this partition is still "good" if it is not optimal under similar configurations.

Finally, the concentration and stability give a measure of how a specific network deviates from an ensemble. For example, we calculate the modularities for a given network G and the ER network ensemble. These two values by themselves do not indicate whether the communities are well defined in G. However, the LDP inequality from the concentration gives a bound on the probability that a network randomly picked from the ER network ensemble has smaller modularity than G by chance. If the probability is small enough, it is statistically significant that G is not chosen from the ER ensemble considering the community structure.

Although not essential to the analysis in this paper, the general form of a community-detecting function can be physically interpreted as a Hamiltonian of a spin system. We study concentration and stability through a classical approach invented by Spencer et al. [Shamir and Spencer 87]: consider an exploring process (edge by edge or node by node) of the network, construct a Doob martingale, and take advantage of martingale inequalities.

We apply our results to several special community-detecting functions: modularity, graph bipartitioning, and q-Potts community structure. We also discuss the problems on different classes of networks such as Erdős–Rényi (ER) and Chung–Lu (CL). Considering the asymptotic behavior of the number of edges as the network size grows, we also classify the network models into sparse and nonsparse. We study only "sparse" networks defined by a constant upper bound of the degree, for which we have proved a very general concentration result for all these problems. For nonsparse networks we derive our results only for specific cases.

2. Background

The underlying martingale inequality (Azuma's inequality) can be traced back to Chebyshev's inequality. Bernstein's inequality [Bernstein 24, Bernstein 37] is considered a modification of Chebyshev's inequality and gives an exponentially decreasing probability upper bound. The first version of the Hoeffding–Chernoff bound, which gives a very general probability upper bound for sums of i.i.d. random variables, was presented and proved in [Hoeffding 63]. In the same paper,

Hoeffding also proposed a slightly different version from the case in which the random variables are not necessarily identical but uniformly bounded, and this version is usually called Hoeffding's inequality. In [Azuma 67], the independent variables in Hoeffding's inequality are extended to martingale differences, resulting in Azuma's inequality. That is a great improvement for us, because the independence between the existence of edges is usually unavailable in network study.

Azuma's inequality is applied in [Rhee and Talagrand 87] to NP-hard optimization problems. The authors give stochastic models for the bin-packing problem and traveling salesman problem as two examples, and they assume a sequence of optimal solutions of a growing system to construct the martingale.

In [Shamir and Spencer 87], Azuma's inequality is applied to the problem of the chromatic number of a random graph. The study [McDiarmid 89] summarizes this technique on random graphs as the method of bounded differences and discusses some extensions such as isoperimetric inequalities. Then several statistics of random graphs are studied, such as average distance [Chung and Lu 02a], connected component size [Chung and Lu 02b], and number of triangles [Kim and Vu 04]. More generalized martingale inequalities [Godbole and Hitczenko 98, Kim and Vu 04] and models of random graphs [Chung and Lu 02a, Chung and Lu 02b] are considered in these studies.

Our study combines the above two types of applications of Azuma's inequality, considering both the optimization problem and the random graph factors. Compared with [Rhee and Talagrand 87], the stochastic model is replaced by a random network, and instead of an optimal solution sequence of increasing system sizes, we consider an arbitrary solution on a fixed but gradually uncovered random graph and finally try to find the optimal ones. The works [Shamir and Spencer 87, McDiarmid 89] are the most relevant to ours. We use almost the same technique in derivation. Other works mentioned above all consider some direct statistics of random networks without an optimization procedure, so they are conceptually simpler than ours.

3.   General Model

The random network model is represented by an ensemble $\Omega(N) = \{\Gamma(N), \mathcal{P}\}$, where $\Gamma(N)$ is the collection of all connected graphs with N nodes, and $\mathcal{P}$ is a probability measure on $\Gamma(N)$. Furthermore, $G \in \Omega(N)$ is considered a random network taking values in $\Gamma(N)$ with probability $\mathcal{P}(G)$. Its adjacency matrix is given by $A = [A_{ij}]_{N \times N}$, and the degree of each node is denoted by d_i $(i = 1, \ldots, N)$. Additionally, we require that the probabilities for any two nodes to be linked must be independent, i.e., $\{A_{ij} \mid i < j\}$ are $N(N-1)/2$ independent

random variables. A spin vector $\vec{s} = (s_1, \ldots, s_N)$ $(s_i = \{-1, 1\})$ assigns a spin to each node and indicates a partition that takes the nodes with the same spin direction (-1 or 1) in one group. The community-detecting function is given by

$$h_G(\vec{s}) = -\frac{1}{N} \sum_{i,j=1}^{N} F_{ij}(G) s_i s_j, \qquad (3.1)$$

where $\{F_{ij}\}$ are functions of the random network G and hence random variables. Many community-detecting problems can be considered special cases of this general model. For example, when $F_{ij}(G) = A_{ij}(G)$, equation (3.1) becomes the community-detecting function for the graph-bipartitioning problem [Wiethege and Sherrington 87]. Because the form of the community-detecting function can be considered the Hamiltonian of a spin system, we also call $h_G(\vec{s})$ the Hamiltonian in a later context. Let S be the spin configuration space including all possible spin states satisfying given constraints. We take $\vec{s}_0 \in S$ as the optimal configuration of h_G, and

$$H(G) = \min_{\vec{s} \in S} h_G(\vec{s}) = h_G(\vec{s}_0)$$

depends only on G and therefore is a property of the network itself rather than some specific configuration.

In this paper, we will discuss two properties of $H(G)$ with respect to the given network ensemble: *stability* and *concentration*, both in terms of the probability of large deviation $|H(G) - \langle H \rangle_G| > t$, where $\langle H \rangle_G$ is the ensemble average of $H(G)$, and t is a positive threshold. *Stability* refers to the property that the probability of large deviation is small, i.e., controlled by some exponentially decaying function of t. *Concentration*, as a stronger property than *stability*, refers to the property that the probability of large deviation is not only small, but also vanishing as the system size N tends to infinity.

We start from the following theorem, which is relatively weak but applicable to an arbitrary network ensemble.

Theorem 3.1. *If $\{F_{ij}\}$, $i, j = 1, \ldots, N$, are such that for any two networks $G, G' \in \Omega$ that differ only by one edge, $|F_{ij}(G) - F_{ij}(G')| \leq c$, where c is a constant independent of i, j, then for every real number t,*

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega| > t) \leq 2 \exp\left(-\frac{t^2}{c^2}\right),$$

where $\langle \cdot \rangle_\Omega$ is the ensemble average and c^2 is a constant independent of t and $\Omega(N)$.

To prove theorem 3.1, we need Azuma's inequality and a lemma.

Theorem 3.2. [Azuma 67] *Suppose $Y_K = \sum_{i=1}^{K} X_i$ is a martingale. Given the boundedness of each increment $|X_i| \le b_i$, the following inequality holds for every real number t:*

$$P(|Y_K - E[Y_K]| > t) \le 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{K} b_i^2}\right).$$

Lemma 3.3. *If the conditions in Theorem 3.1 hold, then $|H(G) - H(G')| \le c/N$.*

Proof. We first suppose that G and G' differ only at the edge between the nodes i_0 and j_0:

$$|h_G(\vec{s}) - h_{G'}(\vec{s})| \le \frac{1}{N}|F_{i_0 j_0}(G) - F_{i_0 j_0}(G')||s_0 s_0| \le \frac{c}{N}.$$

Without loss of generality, we assume that $H(G) \ge H(G')$ and that $H(G)$ is achieved at a specific configuration $\vec{s_0}$, that is,

$$h_G(\vec{s_0}) = H(G) \ge H(G') \ge h_{G'}(\vec{s_0}).$$

So $|H(G) - H(G')| \le |h_G(\vec{s_0}) - h_{G'}(\vec{s_0})| \le c/N$. $\qquad\square$

Proof of Theorem 3.1. Suppose $G \in \Omega(N)$ is a random graph whose node pairs are labeled from 1 to $N(N-1)/2$. Each node pair corresponds to one random variable $A_k = \mathbf{1}_{\{\text{node pair } k \text{ is linked}\}}$ $(k = 1, \ldots, N(N-1)/2)$. Therefore the random graph G can be considered a random process $\{A_1, \ldots, A_{N(N-1)/2}\}$, and its filtration is F_k $(k = 0, \ldots, N(N-1)/2)$.

Define $H_k = E[H(G)|F_k]$ $(k = 1, \ldots, N(N-1)/2)$, and $H_0 = E[H(G)|F_0] = \langle H \rangle_\Omega$. We observe that H_k $(k = 0, \ldots, N(N-1)/2)$ is a martingale by construction. We construct an auxiliary process $G' = \{A'_1, \ldots, A'_{N(N-1)/2}\}$ for which $A'_j = A_j$ when $j \ne k+1$ and $A'_j = 1 - A_j$ when $j = k+1$. Then G' shares the same filtration with G and represents the graph that differs from G only by the link between the node pair $k+1$. The increment of H_k is given by

$$|H_{k+1} - H_k| = |H_{k+1} - E[H_{k+1}|F_k]| = |E[H(G)|F_{k+1}] - E[H(G')|F_{k+1}]|$$
$$\le E[|H(G) - H(G')||F_{k+1}] \le \frac{c}{N}.$$

In the last inequality, $|H(G) - H(G')|$ as a random variable depending only on the realization of G is bounded by c/N according to Lemma 3.3, so its conditional expectation is also controlled by this bound. We apply Azuma's inequality to H_k, and obtain

$$P(|H(G) - \langle H \rangle_\Omega| > t) \le 2\exp\left(-\frac{t^2}{2(c/N)^2 N(N-1)/2}\right) \le 2\exp\left(-\frac{t^2}{c^2}\right).$$

The proof is complete.   □

The inequality by itself indicates only some level of stability of $H(G)$ over the ensemble Ω. To interpret this inequality as a concentration result instead of a single network ensemble, we need a random network model consisting of a sequence of network ensembles with different system sizes N.

The above general model does not specify the way to generate the ensemble sequence as $N \to \infty$, especially the growing speed of the number of edges. To keep the networks connected, the number of edges must grow at least linearly with respect to the number of nodes. We take network models whose numbers of edges grow linearly and superlinearly for sparse networks and nonsparse networks separately. There is no general concentration result for nonsparse networks, so we leave a discussion of this case to later sections for specific community-detecting functions. As to the sparse networks, there are various ways to define them as long as their numbers of edges grow linearly and we consider only one definition, K-bound networks.

4. K-Bound Networks

A network is called K-bound if the constant K is an upper bound on the network degrees. A network ensemble or a random network model is called K-bound if K is a uniform upper bound on the degrees over the network ensemble or the ensemble sequence.

A subtle point here is that the K-bound constraint may violate the independence of the links. However, the violation can be ignored, because in a random network with independent links, the probability distribution of the degree decays very fast (the probability that the degree exceeds K has order $O(e^{-K^2})$), and as K increases, the difference between the probability measures of the network ensembles with and without the K-bound constraint goes to zero very quickly. For K-bound networks, we have the following concentration result.

Theorem 4.1. *If Ω is a K-bound ensemble, $\{F_{ij}\}$, $i, j = 1, \ldots, N$, are such that for any two networks $G, G' \in \Omega$ that differ only by one edge, $|F_{ij}(G) - F_{ij}(G')| \leq c$, where c is a constant independent of i, j, then for every real number t,*

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega| > t) \leq 2 \exp\left(-\frac{Nt^2}{8K^2c^2}\right). \tag{4.1}$$

Proof. Suppose $G \in \Omega(N)$ is a random graph with nodes labeled $1, 2, \ldots, N$. Let G_k denote the subgraph containing nodes from 1 to k and also considered as a

filtration. Define $H_k = E[H(G) \mid G_k]$, $k = 1, \ldots, N$, and $H_0 = E[H(G) \mid G_0] = \langle H \rangle_\Omega$. Then H_k $(k = 0, \ldots, N)$ is a martingale by construction. Suppose that G'' is a network that differs from G only in the $(k+1)$th node, i.e., all the different elements between the adjacency matrices of G and G'' correspond to the edges linked to the $(k+1)$th node. Then G_k'' is the corresponding filtration for G. Note that the degree is at most K, and therefore, G and G'' can differ by at most $2K$ edges:

$$|H_{k+1} - H_k| = |H_{k+1} - E[H_{k+1}|F_k]| = \max\{|E[H(G)|G_k] - E[H(G'')|G_k'']|\}$$
$$\leq 2K\,|E[H(G)|G_k] - E[H(G')|G_k]| \leq \frac{2Kc}{N}.$$

Here G' is as defined in the proof of Theorem 3.1. The difference $H(G) - H(G'')$ can be decomposed into at most $2K$ terms all of which have the same bound as $H(G) - H(G')$. Applying Azuma's inequality, we have

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega \,| > t) \leq 2\exp\left(-\frac{t^2}{2(2Kc/N)^2 N}\right) \leq 2\exp\left(-\frac{Nt^2}{8K^2c^2}\right).$$

$\square$

Remark 4.2. According to this inequality, the fluctuation of $H(G)$ about its mean has order $O(N^{-1/2})$.

When we are given a network $G_0 = \{A_{0ij}\}$, one way to generate a network ensemble is to suppose that the given network undergoes a certain perturbation. Suppose the perturbation is denoted by $\delta G(p_0, p_1)$, where p_0, p_1 are two probabilities. The network ensemble $G = \{A_{ij}\}$ generated by G_0 and $\delta G(p_0, p_1)$ satisfies the following:

(a) If $A_{0ij} = 1$, then $A_{ij} = 0$ with probability p_0 and $A_{ij} = 1$ with probability $1 - p_0$.

(b) If $A_{0ij} = 0$, then $A_{ij} = 1$ with probability p_1 and $A_{ij} = 0$ with probability $1 - p_1$.

To conserve the average degree of the original network G_0, we additionally require

$$p_0 m = p_1 \left(\frac{N(N-1)}{2} - m\right),$$

where m is the total number of edges in G_0. Since $m < NK/2$, the requirement implies

$$p_1 \leq \frac{K}{N-1-K} \, p_0.$$

If G_0 satisfies the K-bounded degree condition, we have the following result.

Theorem 4.3. *If Ω is a K-bound ensemble generated by a given K-bound network G_0 and a small perturbation δG, $\{F_{ij}\}$, $i,j = 1, \ldots, N$, are such that for any two networks $G, G' \in \Omega$ that differ only by one edge, $|F_{ij}(G) - F_{ij}(G')| \leq c$, where c is a constant independent of i, j, then for every real number t,*

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega| > t) \leq 2 \exp\left(-\frac{Nt^2}{4\left[c^2\left(2K^2p_0^2 + Kp_0\right) + Kct/3\right]}\right). \quad (4.2)$$

To prove Theorem 4.3, we need a variant form of Azuma's inequality:

Theorem 4.4. [Chung and Lu 06] *Suppose $Y_K = \sum_{i=1}^{K} X_i$ is a martingale. Then given the boundedness of each increment $|X_i| \leq M$, $\mathrm{Var}(X_i) = v_i^2$, the following inequality holds for every real number t:*

$$\mathcal{P}(|Y_K - E[Y_K]| > t) \leq 2 \exp\left(-\frac{t^2}{2\left(\sum_{i=1}^{K} v_i^2 + Mt/3\right)}\right).$$

Proof. The proof is almost the same as that for Theorem 4.1. We have only to replace Theorem 3.2 in that proof by Theorem 4.4 and find the upper bound of $\mathrm{Var}(X_i)$.

Let the random variable L_i denote the number of changes between $G_0 = \{A_{0ij}\}$ and $G = \{A_{ij}\}$ ($i < j$) related to the node i. Since L_i can be considered a sum of 0-1 random variables, we have

$$\mathrm{Var}(L_i) = Kp_0(1-p_0) + (N-1-K)p_1(1-p_1)$$
$$\leq Kp_0 + (N-1-K)p_1 \leq 2Kp_0$$

and

$$E[L_i] = Kp_0 + (N-1-K)p_1 \leq 2Kp_0.$$

Since $|X_i| \leq L_i c/N$, we have

$$\mathrm{Var}(X_i) \leq \frac{c^2}{N^2} E[L_i^2] = \frac{c^2}{N^2}(E^2[L_i] + \mathrm{Var}(L_i)) = \frac{c^2}{N^2}\left(4K^2p_0^2 + 2Kp_0\right),$$

and Theorem 4.4 becomes a corollary of Theorem 4.3. Compared with Theorem 4.1, Theorem 4.3 gives a much sharper concentration around the average value whenever p_0 is small.   □

In the following sections, we will prove the concentration for several different community-detecting functions on both the nonsparse and sparse subclasses of Erdős–Rényi (ER) and Chung–Lu (CL) random networks. For all the nonsparse cases, we can use Theorem 3.1 together with a "divide and conquer" technique described in the next section to obtain the concentration. However, when the community-detecting function has additional properties such as $\langle H(G)\rangle_\Omega \simeq O(N)$ as in the graph-bipartitioning case, we can scale t in Theorem 3.1 by N to obtain a better concentration inequality without using the "divide and conquer" technique. For the K-bounded degree subclasses of ER and CL networks, we will use Theorems 4.1 and 4.3 along with technical estimates specific to each case.

5. Modularity

Modularity is one of several effective criteria for the detection of community structures in random networks [Newman 06]. In this section, the theorems presented above are used to obtain LDP results on the modularity functional over the nonsparse Erdős–Rényi subclass $\mathrm{ER}[p, N]$, and where there is a uniform bound on the degrees of G over the sparse $\mathrm{ER}[Np, K]$ subclass (and the Chung–Lu $\mathrm{CL}[N, \varpi, K]$ subclass, whose properties are given in the next section). The nonsparse cases here and in the next section are treated with a "divide and conquer" technique because without prior knowledge of properties of their respective ensemble average, $\langle H(G)\rangle_\Omega$, we have no recourse to the shorter method of rescaling t in theorem 3.1 by N.

In the modularity problem,

$$F_{ij}(G) = \frac{N}{4m}\left(A_{ij} - \frac{d_i d_j}{2m}\right).$$

Therefore, the Hamiltonian (3.1) is

$$h_G(\vec{s}) = -\frac{1}{4m}\sum_{1\le i\ne j\le N}\left(A_{ij} - \frac{d_i d_j}{2m}\right)s_i s_j.$$

Here m is the total number of edges in the network, d_i is the degree of node i, and A_{ij} is the adjacency matrix. The Hamiltonian is more complicated than the previous two cases, since $F_{ij}(G)$ depends not only on the local information A_{ij} but also on the global information such as d_i, m. Suppose G' has only one

more edge $(i_0 j_0)$ than G. To apply Theorems 3.1 and 4.1, we need to estimate $|F_{ij}(G) - F_{ij}(G')|$ as follows:

$$\frac{4m}{N}|F_{ij}(G) - F_{ij}(G')| \tag{5.1}$$

$$= \left| \sum_{1 \le i \ne j \le N} \left[(A'_{ij} - A_{ij}) - \left(\frac{d'_i d'_j}{2(m+1)} - \frac{d_i d_j}{2m} \right) \right] s_i s_j \right|$$

$$+ \left(\frac{4m}{N} \right) \left(\frac{m+1}{m} - 1 \right) |F_{ij}(G')|$$

$$\le \frac{1}{m} |A_{i_0,j_0} - A'_{i_0,j_0}| + \left| \sum_{1 \le i \ne j \le N} \frac{d'_i d'_j - d_i d_j}{2(m+1)} \right|$$

$$+ \left| \left(\frac{1}{2(m+1)} - \frac{1}{2m} \right) \sum_{1 \le i \ne j \le N} d_i d_j \right| + \left(\frac{1}{m+1} \right) \left| A'_{ij} - \frac{d_i d_j}{2(m+1)} \right|.$$

Next, we estimate the four terms in (5.1). The first term is $|A_{i_0,j_0} - A'_{i_0,j_0}| = 1$. The second term is as follows:

$$\left| \sum_{1 \le i \ne j \le N} \frac{d'_i d'_j - d_i d_j}{2(m+1)} \right|$$

$$= \frac{1}{2(m+1)} \left| \sum_{1 \le j \ne i_0 \le N} (d'_{i_0} - d_{i_0}) d_j \right.$$

$$\left. + \sum_{1 \le i \ne j_0 \le N} (d'_{j_0} - d_{j_0}) d_i + (d'_{i_0} - d_{i_0})(d'_{j_0} - d_{j_0}) \right|$$

$$\le \frac{1}{2(m+1)} (2m + 2m + 1) < 2.$$

For the third term, we have

$$\left| \left(\frac{1}{2(m+1)} - \frac{1}{2m} \right) \sum_{1 \le i \ne j \le N} d_i d_j \right| \le \frac{1}{2m(m+1)} \left| \left(\sum_{i=1}^N d_i \right) \left(\sum_{j=1}^N d_j \right) \right|$$

$$\le \frac{1}{2m(m+1)} (2m)^2 < 2.$$

For the last term, since $d_i, d_j \le m$, we obtain

$$\left(\frac{1}{m+1} \right) \left| A'_{ij} - \frac{d_i d_j}{2(m+1)} \right| \le \left(\frac{1}{m+1} \right) \left| \frac{m^2}{2(m+1)} - 1 \right|$$

$$\le \max \left\{ \frac{m^2}{2(m+1)^2}, \frac{1}{m+1} \right\} < \frac{1}{2}.$$

With the above inequalities, we conclude that

$$|F_{ij}(G) - F_{ij}(G')| \leq \frac{11N}{8m} \leq \frac{11N}{8m^*},$$

where m^* is the minimum value of m required by connectivity. This completes the proof of the technical estimates needed in the application of Theorems 3.1 and 4.1 below.

For general ER networks without additional nonsparseness or bounded degree properties and also for sparse ER networks without assumptions on, for instance, the scale of the ensemble average, $\langle H(G) \rangle$, Theorem 3.1 gives

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega| > t) \leq 2 \exp\left(-\frac{t^2}{\sigma^2 (N/m^*)^2}\right) \leq 2 \exp\left(-\frac{t^2}{\sigma^2}\right),$$

where $\sigma = 11/8$, which shows that direct application of Theorem 3.1 alone does not allow us to obtain the concentration.

However, for a sparse ER network with bounded degree, using Theorem 3.2, we prove the following.

Theorem 5.1. *In* $\mathrm{ER}[Np = \lambda]$ *and* $\mathrm{CL}[N, \vec{w}]$ *network ensembles* Ω *with uniform upper bound* K *of the degree, the optimal modularity* $H(G)$ *satisfies the concentration inequality*

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega| > t) \leq 2 \exp\left(-\frac{Nt^2}{\frac{25}{2}(\frac{N}{m^*})^2 K^2}\right) \leq 2 \exp\left(-\frac{Nt^2}{\frac{25}{2} K^2}\right).$$

In some sense, the category of networks with degree upper bound is an extreme case of sparse networks. Next, we consider the other extreme case, the nonsparse networks, using the following "divide and conquer" method. The network is nonsparse if the following hold:

(a) $\langle m \rangle_{\Omega(N)} = pN^\alpha/2$ (p, α are two constant parameters).

(b) $P(m - \langle m \rangle_{\Omega(N)} < -Nt) \leq \exp\left(-t^2/\lambda^2\right)$, where m is the total number of edges and p is a constant independent of N.

The meaning of property (b) is revealed in the following lemma.

Lemma 5.2. *If in some network, all the edges are independent, then there exists a constant* λ *independent of* N *such that the number of edges* m *satisfies*

$$P(m - \langle m \rangle_{\Omega(N)} < -Nt) \leq \exp\left(-\frac{t^2}{\lambda^2}\right).$$

Proof. We have $m = \sum_{1 \leq i < j \leq N} \mathbf{1}_{\{A_{ij}=1\}}$, where $\mathbf{1}_{\{A_{ij}=1\}} \in [0,1]$ are independent random variables. According to Hoeffding's inequality [Hoeffding 63],

$$\mathcal{P}\left(\frac{m - \langle m \rangle}{N(N-1)/2} < -t\right) \leq \exp\left(-\frac{2t^2 N^2 (N-1)^2}{4 \sum_{1 \leq i < j \leq N} 1}\right).$$

Substituting t by $2t/(N-1)$, we get

$$\mathcal{P}(m - \langle m \rangle < -Nt) \leq \exp\left(-\frac{2t^2 N^2}{N(N-1)/2}\right) \leq \exp\left(-\frac{t^2}{\lambda^2}\right),$$

where $\lambda = \frac{1}{2}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

An example of a nonsparse network is an ER network $\mathrm{ER}(p, N)$ with constant probability p for any two nodes to be linked. Replacing t in (b) by $\mu N^{\alpha-1}/2$, where μ is independent of N, we get

$$\mathcal{P}(m < (p - \mu)N^\alpha/2) \leq \mathcal{P}\big(m < \big(\langle m \rangle_{\Omega(N)} - \mu N^\alpha/2\big)\big) \leq \exp\big(-\mu^2 N^{2(\alpha-1)}\big).$$

Finally, we split $P(|H(G) - \langle H \rangle_\Omega| > t)$ into two cases according to m, and get the concentration result

$$
\begin{aligned}
P(|H(G) &- \langle H \rangle_\Omega| > t) \\
&= P(m < (p - \mu)N^\alpha/2)P(|H(G) - \langle H \rangle_\Omega| > t | m < (p - \mu)N^\alpha/2) \\
&\quad + P(m \geq (p - \mu)N^\alpha/2)P(|H(G) - \langle H \rangle_\Omega| > t | m \geq (p - \mu)N^\alpha/2) \\
&\leq \exp\big(-\mu^2 N^{2(\alpha-1)}\big) \cdot 2\exp\big(-t^2\big) + 1 \cdot 2\exp\left(-\frac{t^2}{(N/(p-\mu)N\alpha)^2}\right) \\
&\leq 2\exp\big(-\mu^2 N^{2(\alpha-1)} - t^2\big) + 2\exp\big(-t^2 N^2 (p-\mu)^2\big).
\end{aligned}
$$

When $1 < \alpha \leq 2$, the inequality shows the concentration. For an $\mathrm{ER}[N, p]$ network $\alpha = 2$, we have the following theorem.

Theorem 5.3. *In a nonsparse $\mathrm{ER}[N, p]$ network ensemble Ω, the optimal modularity $H(G)$ satisfies the concentration inequality*

$$P(|H(G) - \langle H \rangle_\Omega| > t) \leq 2\exp\big(-\mu^2 N^2 - t^2\big) + 2\exp\big(-t^2 N^2 (p-\mu)^2\big). \quad (5.2)$$

6. Concentration of Modularity on the Chung–Lu Network $\mathrm{CL}_\infty[N, \varpi, \beta]$, $\mathrm{CL}[N, \varpi, K]$

Scale-free or power-law random graphs including Barabási–Albert (BA), Molloy–Reed (MR), and Chung–Lu (CL) models are frequently encountered in the study of random networks that arise in social and ecological problems. For brevity,

we will use the connected random graphs generated by the Chung–Lu model $CL(N, \varpi)$ [Chung and Lu 02b], which is based on fixed expected degrees sequence ϖ. The Chung–Lu model is easier to work with than the Molloy–Reed [Molloy and Reed 95], Newman–Strogatz–Watts [Watts and Strogatz 98], and Barabási–Albert [Albert and Barabási 02] formulations, because it specifies more information at the level of each node j in the graph $G(N)$. It is based on working with subsets $CL[N, \varpi]$ of the scale-free random graphs specified by a fixed (deterministic) expected degree sequence of weights

$$\varpi = (w_1, \ldots, w_N), \quad w_j = E[\deg(j)].$$

The average degree in $CL[N, \varpi]$ is given by

$$d(\varpi) = \frac{1}{N} \sum_{l=1}^{N} w_l.$$

In the growth process of CL random networks (cf. [Albert and Barabási 02]), a new node v_i is added at time $i \leq N$, and m' random and independent edges are then added between this node v_i and those already present. Thus, for node i added, the probability of adding a link to node j is $w_j / \sum_l w_l$. The average degree of the new node v_i, i.e., the second-order degree, is given by

$$\bar{d}(w) = \sum_{j=1}^{n} w_j \frac{w_j}{\sum_l w_l},$$

whence it is easily shown via an application of Cauchy–Schwarz that

$$\bar{d}(\varpi) - d(\varpi) \geq 0.$$

A key property of $CL[N, \varpi]$ for the proof of concentration of modularity below is the probability of an edge existing between an arbitrary pair of nodes i and j or that independent random variables $A_{ij} = 1$ (where $A(G)$ is the adjacency matrix of the random graph G) are given by [Chung and Lu 02a, Chung and Lu 02b]

$$\Pr\{A_{ij}(G(N)) = 1\} = \frac{w_i w_j}{\sum_{l=1}^{N} w_l}.$$

In each subclass $\Omega = CL[N, \varpi]$, the average number of edges m is

$$\langle m \rangle_\Omega = \left(\frac{N(N-1)}{2} + N \right) \frac{1}{N^2} \sum_i^N \sum_j^N \frac{w_i w_j}{\sum_k w_k} = \frac{N(N+1)}{2N^2} \sum_{i=1}^{N} w_i$$

$$= \frac{d(\varpi)}{2}(N + 1),$$

where the average degree d may depend on N through the weights w_j.

There are no known concentration results on the whole class $\mathrm{CL}[N, \varpi]$; this class includes sparse random networks that are not uniformly bounded in the degree of the nodes. Using the "divide and conquer" technique and Theorem 3.1 twice, we prove LDP results for nonsparse Chung–Lu networks $\mathrm{CL}_\infty[N, \varpi, \beta]$ for which, in addition to the above properties, the average degree grows with N, that is, for $\beta > 0$ and $B > 0$ both independent of N,

$$d(\varpi) \geq BN^\beta.$$

Concentration of modularity in the subclass of bounded-degree Chung–Lu networks $\mathrm{CL}[N, \varpi, K]$ was proved in the previous section.

Next, by Lemma 5.2, this subclass $\Omega(N) = \mathrm{CL}_\infty[N, \beta]$ satisfies the nonsparse property (b) in the following:

(a) $\langle m \rangle_{\Omega(N)} \geq MN^\alpha$ (here $M > 0$, $\alpha > 1$ are two constant parameters).

(b) There exists a constant λ independent of N such that

$$\Pr(m - \langle m \rangle_{\Omega(N)} < -Nt) \leq \exp\left(-\frac{t^2}{2\lambda^2}\right)$$

for every real $t > 0$.

Property (a) holds by choosing $M = B/2$ and $\alpha = 1 + \beta$. Replacing t in (b) by μN^β gives

$$P(m < (M - \mu)N^{\beta+1}) \leq P\left(m < (\langle m \rangle_\Omega - \mu N^{\beta+1})\right) \leq \exp\left(-\frac{\mu^2 N^{2\beta}}{2\lambda^2}\right).$$

By conditioning $P(|H(G) - \langle H \rangle_\Omega| > t)$ according to m, we derive that for constants $\lambda, \beta > 0$ and $\mu < M = B/2$ that do not depend on N,

$$
\begin{aligned}
P(|H(G) &- \langle H \rangle_{\mathrm{CL}_\infty}| > t) \\
&= P(m < (M - \mu)N^{\beta+1})P\{|H(G) - \langle H \rangle_{\mathrm{CL}_\infty}| > t \mid m < (M - \mu)N^{\beta+1})\} \\
&\quad + P(m \geq (M - \mu)N^{\beta+1})P\left\{|H(G) - \langle H \rangle_{\mathrm{CL}_\infty}| > t \mid m \geq (M - \mu)N^{\beta+1})\right\} \\
&\leq \exp\left(-\frac{\mu^2 N^{2\beta}}{2\lambda^2}\right) \cdot 2\exp\left(-\frac{t^2}{2\sigma^2}\right) + 1 \cdot 2\exp\left(-\frac{t^2(M - \mu)^2 N^{2\beta}}{2}\right) \\
&\leq 2\exp\left(-\frac{\mu^2 n^{2\beta}}{2\lambda^2} - \frac{t^2}{2\sigma^2}\right) + 2\exp\left(-\frac{t^2(M - \mu)^2 n^{2\beta}}{2}\right).
\end{aligned}
$$

We used the following applications of Theorem 3.1: for every real $t > 0$,

$$P\{|H(G) - \langle H\rangle_{\mathrm{CL}_\infty}| > t \mid m < (M - \mu)N^{\beta+1}\} \le 2\exp\left(-\frac{t^2}{2\sigma^2}\right),$$

$$P\{|H(G) - \langle H\rangle_{\mathrm{CL}_\infty}| > t \mid m \ge (M - \mu)N^{\beta+1}\} \le 2\exp\left(-\frac{t^2(M-\mu)^2 N^{2\beta}}{2}\right),$$

with respectively

$$c^2\frac{N(N+1)}{2} = \frac{25N(N+1)}{32m^2} \le \frac{25}{32}\left(1 + \frac{3}{N-1} + \frac{2}{(N-1)^2}\right) \le \frac{25}{16} \equiv \sigma^2,$$

and when $m^2 \ge (M - \mu)^2 N^{2(\beta+1)}$, then

$$c^2\frac{N(N+1)}{2} = \frac{25N(N+1)}{32m^2} \le \frac{25}{32}\frac{N(N+1)}{(M-\mu)^2 N^{2\beta+2}} \le \frac{1}{(M-\mu)^2 N^{2\beta}},$$

where the constant $c > 0$ comes from the technical estimate (5.1), which is valid for every pair of random networks G, G' differing by exactly one edge in any random ensemble Ω. This completes the proof of the following theorem.

Theorem 6.1. *In a nonsparse* $\mathrm{CL}_\infty[N, B, \beta]$ *network ensemble* Ω, *the optimal modularity* $H(G)$ *satisfies the following concentration inequality: there exists* $\mu > 0$ *independent of* N *such that for every real* $t > 0$,

$$P(|H(G) - \langle H\rangle_\Omega| > t)$$
$$\le 2\exp\left(-\frac{\mu^2 N^{2\beta}}{2\lambda^2} - \frac{t^2}{2\sigma^2}\right) + 2\exp\left(-\frac{t^2(B/2 - \mu)^2 N^{2\beta}}{2}\right),$$

where $\lambda = 1/2$ *and* $\sigma = 5/4$.

7.   Graph Bipartitioning on ER and CL Networks

The graph-bipartitioning problem [Wiethege and Sherrington 87] is the simplest example in which $F_{ij}(G) = A_{ij}$ in (3.1):

$$h_G(\vec{s}) = -\frac{1}{N}\sum_{i,j=1}^{N} A_{ij}s_i s_j.$$

In this problem, each spin state corresponds to a two-group partition of the given graph, and the optimum gives the partition with the fewest intergroup links. In [Fu and Anderson 86], it was shown that this problem for an ER$[N, p]$ network (every pair of nodes are linked by an edge with probability $p > 0$, where p is independent of N) is equivalent to the problem for an infinite-range SK model

[Sherrington and Kirkpatrick 75], and the authors derived the average solution in the thermodynamic limit. The authors of [Banavar et al. 87] investigated another case of an $\text{ER}[Np = \lambda]$ network and gave an empirical average solution. However, without a concentration result, these solutions are only heuristic. Even if we accept the solution in the thermodynamic limit, the errors of the solutions applied to a finite system cannot be estimated. Our research completes those results.

In those studies, the constraint of zero magnetization $M = \sum_i s_i = 0$ is required to force an equal-size partition. With a nonzero fixed magnetization constraint $\sum_i s_i = c \neq 0$, we get the optimal partition with two given group sizes. Without any constraint on M, we get the overall optimal partition considering all possible group sizes. No matter which constraint we use, it only changes the spin configuration space and hence the definition of $H(G) = \min_{s \in S} h_G(s)$, so we have exactly the same concentration result and proofs in all of these cases.

For the most general random graph ensemble, we apply Theorem 3.1 as follows. Since

$$|F_{ij}(G) - F_{ij}(G')| = |A_{ij}(G) - A_{ij}(G')| \leq 1,$$

we have

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega| > t) \leq 2 \exp\left(-t^2\right),$$

which shows that the concentration is not obtained by this approach. To compare with the papers mentioned above, we consider the two types of ER network separately. For the first type, $\text{ER}[N, p]$, we assume in addition that $\langle H \rangle_\Omega$ is of order $O(N)$, which is consistent with the results from the replica method [Fu and Anderson 86], which allows us to scale t with the same order and have the following result.

Theorem 7.1. *In a nonsparse $\text{ER}[N, p]$ network ensemble Ω, the optimal graph partitioning $H(G)$ satisfies the concentration inequality*

$$P(|H(G) - \langle H \rangle_\Omega| > Nt) \leq 2 \exp\left(-N^2 t^2\right).$$

For the second type, $\text{ER}[Np = \lambda]$, for which Np is fixed as $N \to \infty$, we prove a theorem under the additional assumption of bounded degree. Consider the $\text{ER}[Np = \lambda]$ network ensemble generated with parameters N, p, excluding all samples whose maximum degree exceeds the bound $K > Np$, where K is independent of N. So as $N \to \infty$, the degree distribution of the network tends to a Poisson distribution with expected degree Np but with a cutoff at K. If K is

large enough, the network ensemble generated like this is almost the second type of ER. For this case, we use Theorem 4.1 to prove our next result.

Theorem 7.2. *In an* $\text{ER}[Np = \lambda]$ *ensemble* Ω *with uniform upper bound* K *of the degree, the optimal graph partitioning* $H(G)$ *satisfies the concentration inequality*

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega | > t) \leq 2 \exp\left(-\frac{Nt^2}{8K^2}\right).$$

For comparison with Theorems 7.1 and 7.2, the simulated concentration for both cases of ER model is shown in Figure 1. For the constant-p case, the fluctuation of $H(G)/N$ is roughly of order $O(N^{-1})$, while for the constant-Np case, the fluctuation of $H(G)$ is roughly of order $O(N^{-1/2})$.

As an application of the above theorems to a real-world problem, we consider the circuit partition optimization problem proposed in [Kirkpatrick et al. 83]. The objective function proposed there is equivalent to

$$h_G(\vec{s}) = -\frac{1}{N} \sum_{i,j=1}^{N} (A_{ij}/2 - \lambda)\, s_i s_j.$$

So $F_{ij} = A_{ij}/2 - \lambda$, where A_{ij} is the number of signals passing between circuits i and j, and λ is a balancing coefficient in the optimization [Kirkpatrick et al. 83].

In that study, A_{ij} is no longer a 0-1 random variable, but as long as the A_{ij} are uniformly bounded, which is quite reasonable in the problem considered, we can always normalize A_{ij} by the uniform upper bound, and all the theorems and proofs in this section still hold. The λ introduced in this problem represents a penalty on partitions with unequal groups sizes, making this model a special case of the modularity method.

8.   Concentration for q-Potts Community Structures on ER and CL Networks

This section is devoted to concentration results for the family of objective functionals derived from q-Potts models, which are introduced as a viable alternative to the modularity community-detection algorithm:

$$h_G(\vec{s}) = -\frac{J}{4m} \sum_{1 \leq i \neq j \leq N} A_{ij} \delta(s_i, s_j) + \frac{\gamma}{2m} f(\vec{n}_s), \tag{8.1}$$

where the spins $s_j \in (0, 1, \ldots, q - 1)$, $A_{ij} = 0, 1$, form the adjacency matrix of G, $\delta(s_i, s_j) = 1$ only if $s_i = s_j$ and is otherwise zero, $f(\vec{n}_s)$ is a function of the occupation numbers $\vec{n}_s = (n_0, n_1, \ldots, n_{q-1})$, $G(N, m)$ is a random graph with N
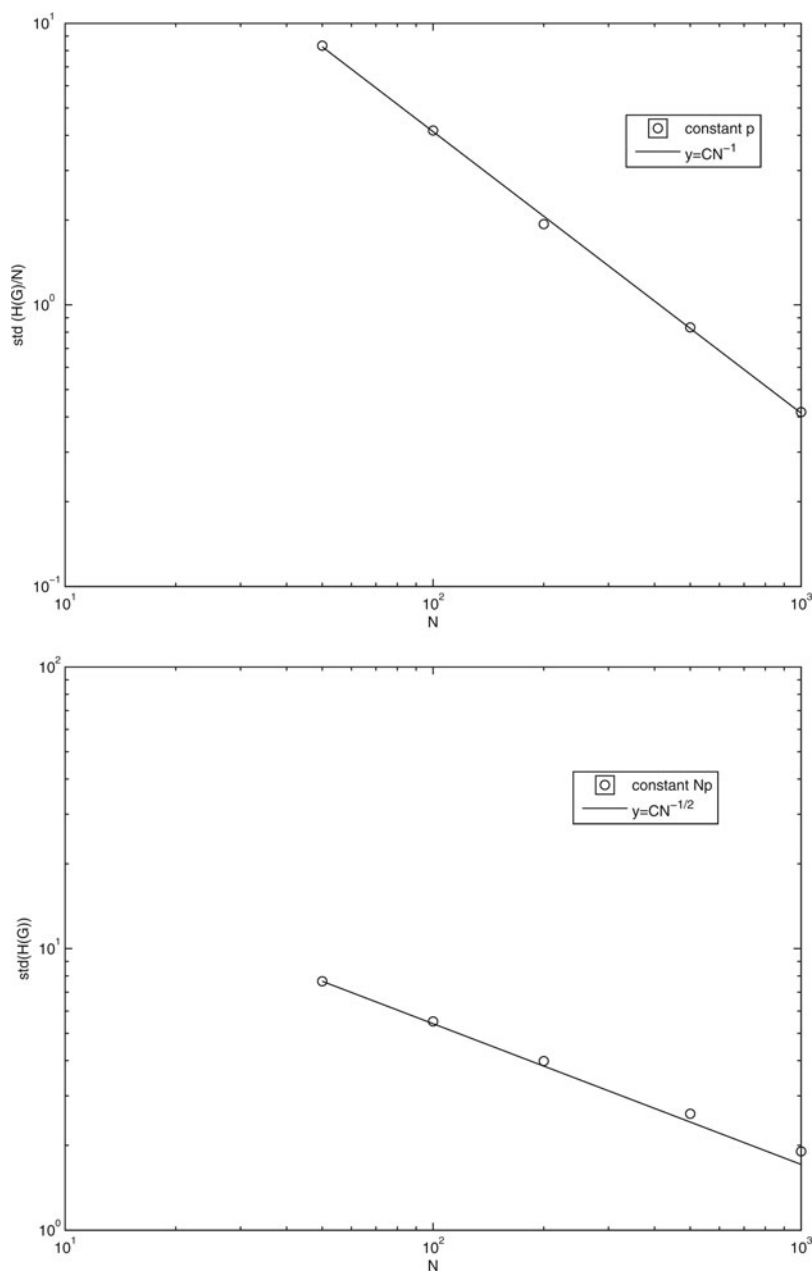
Figure 1. Each data point is obtained by optimization using simulated annealing on 100 realizations of the ER network. The first figure is for ER with constant $p = 0.05$, and the second figure is for ER with constant $pN = 5$.

nodes and m links, J is a ferromagnetic interaction energy, and γ is a parameter that determines the antiferromagnetic activity of the Hamiltonian. The particular example of these Hamiltonians studied in [Reichardt and Bornholdt 04] is

$$h_G(\vec{s}) = -\frac{J}{4m} \sum_{1 \leq i \neq j \leq N}^{N} A_{ij}\delta(s_i, s_j) + \frac{\gamma}{2m} \sum_{s=0}^{q-1} \frac{n_s(n_s - 1)}{2},$$

where for $\gamma > \gamma^*$, the optimum of h_G favors community structures that reflect the network topology of G. The threshold γ^* is fixed by requiring $h_G(\text{homogeneous}) > h_G(\text{diverse})$, which for two communities $c(n_1, m_1)$ and $c(n_2, m_2)$ can be rewritten thus:

$$\begin{aligned} &h_G(\text{homogeneous}) \\ &= -\frac{J}{4m}(m_1 + m_2 + m_{12}) + \frac{\gamma^*}{4m}N(N-1) \\ &= -\frac{J}{4m}(m_1 + m_2 + m_{12}) + \frac{\gamma^*}{4m}[n_1(n_1-1) + n_2(n_2-1) + 2n_1 n_2] \\ &> -\frac{J}{4m}(m_1 + m_2) + \frac{\gamma^*}{4m}[n_1(n_1-1) + n_2(n_2-1)] = h_G(\text{diverse}), \end{aligned}$$

which is, in turn, equivalent to the normalized value of the outlink density or intercommunity links density,

$$\frac{Jm_{12}}{2n_1 n_2} = \gamma^*,$$

since $m = m_1 + m_2 + m_{12}$ and $N = n_1 + n_2$.

The proofs of LDP over sparse Erdős–Rényi random graphs $\text{ER}(N, p)$ and scale-free Chung–Lu graphs $\text{CL}_\infty(N, \varpi)$ with expected degree sequence ϖ will be given for arbitrary f in (8.1), since they do not depend on the form of the function f. To begin the proof for $\text{ER}(N, p)$, based on the Azuma–Hoefding inequalities, we define, as before, $H(G) \equiv \min h_G = h_G(\vec{s}_0)$ and label the optimum $\vec{s}_0$. We need a technical estimate required in the application of Theorem 3.1.

Lemma 8.1. *There is a constant $c > 0$ independent of N such that for any two graphs $G(N, m')$, $G'(N, m'' = m' + 1)$ in the random ensemble $\Omega(N)$ that differ in exactly one edge,*

$$|H(G) - H(G')| < c.$$

Proof. For a fixed state $\vec{s} = (s_1, \ldots, s_N)$, since the second term f cancels, we obtain

$$\frac{4m'}{J} |h_G(\vec{s}) - h_{G'}(\vec{s})| = \left| \sum_{1 \leq i \neq j \leq N} \left[A_{ij} - A'_{ij} \right] \delta(s_i, s_j) \right|$$

$$= \left| \left[A_{i_0 j_0} - A'_{i_0 j_0} \right] \delta(s_{i_0}, s_{j_0}) \right| \leq 2,$$

implying that $|h_G(\vec{s}) - h_{G'}(\vec{s})| \leq J/4m$. A standard argument based on assuming (without loss of generality) $H(G) \leq H(G')$, that is, $h_G(\vec{s}_0) = \min h_G = H(G)$, $h_{G'}(\vec{s}_0^i) = \min h_{G'} = H(G')$, and $h_{G'}(\vec{s}_0) \geq \min h_{G'} \equiv H(G')$, so $h_G(\vec{s}_0) = H(G) \leq H(G') \leq h_{G'}(\vec{s}_0)$, implies the result

$$|H(G) - H(G')| < |h_G(\vec{s}_0) - h_{G'}(\vec{s}_0)| \leq \frac{J}{2}.$$

The rest of the proof is exactly the same as in previous sections. $\square$

Using the divide and conquer technique based on the nonsparse property valid for $\mathrm{ER}[p, N]$ and $\mathrm{CL}_\infty[N, B, \beta]$ and Theorem 3.1 together with Lemma 8.1, we prove the following.

Theorem 8.2. *In the nonsparse* $\mathrm{ER}[N, p]$ *and* $\mathrm{CL}_\infty[N, B, \beta]$ *ensembles* Ω*, the optimal q-Potts functional $H(G)$ satisfies the following concentration inequalities: there exists $\mu > 0$ independent of N such that for real $t > 0$,*

$$P(|H(G) - \langle H \rangle_\Omega | > t) \leq 2 \exp\left(-\mu^2 N^2 - t^2\right) + 2 \exp\left(-t^2 N^2 (p - \mu)^2\right),$$

$$P(|H(G) - \langle H \rangle_\Omega | > t) \leq 2 \exp\left(-\frac{\mu^2 N^{2\beta}}{2\lambda^2} - \frac{t^2}{2\sigma^2}\right)$$
$$+ 2 \exp\left(-\frac{t^2 (B/2 - \mu)^2 N^{2\beta}}{2}\right),$$

with $\lambda = 1/2$ and $\sigma = J/2$.

Using Lemma 8.1 and Theorem 4.1, we prove the following result.

Theorem 8.3. *In* $\mathrm{ER}[Np, K]$ *and* $\mathrm{CL}[N, \varpi, K]$ *ensembles* Ω *with uniform upper bound K on the degree that is independent of N, the optimal q-Potts functional $H(G)$ satisfies the following concentration inequality: for every real number $t > 0$,*

$$\mathcal{P}(|H(G) - \langle H \rangle_\Omega | > t) \leq 2 \exp\left(-\frac{Nt^2}{8K^2}\right).$$

9.   Conclusion

In this paper, we have derived LDP-type inequalities for the optimal values of community-detecting functions on random networks to show their concentration and stability. There is no concentration for the most general case. We prove concentration for the general community-detecting function on K-bound sparse networks and an even sharper concentration when the network ensemble is generated by a given network and a small perturbation. Then we examine several specific cases. The three specific community-detecting functions we considered are modularity, graph bipartioning, and q-Potts community structure. The specific network types we considered are ER and CL networks, and each of them in the sparse (K-bound) and nonsparse cases. We proved concentration in these cases, which means that in these cases, the community-detecting functions are stable, especially when the system size is large enough.

References

[Albert and Barabási 02] R. Albert and A.-L. Barabási. "Statistical Mechanics of Complex Networks." *Reviews of Modern Physics* 74 (2002), 47–97.

[Azuma 67]   K. Azuma. "Weighted Sums of Certain Dependent Random Variables." *Tohoku Math. Jour.* 19 (1967), 357–367.

[Banavar et al. 87]   Jayanth R Banavar, David Sherrington, and Nicolas Sourlas. "Graph Bipartitioning and Statistical Mechanics." *J. Phys. A: Math. Gen.* 20 (1987), L1–L2.

[Bernstein 24]   S. N. Bernstein. "On a Modification of Chebyshev's Inequality and of the Error Formula of Laplace." *Ann. Sci. Inst. Sav. Ukraine, Sect. Math. 1* 4:5 (1924), 38–49.

[Bernstein 37]   S. N. Bernstein. "On Certain Modifications of Chebyshev's Inequality." *Doklady Akademii Nauk SSSR* 17:6 (1937), 275–277.

[Chung and Lu 02a]   F. Chung and L. Lu. "The Average Distances in Random Graphs with Given Expected Degrees." *Proceedings of National Academy of Sciences* 99 (2002), 15879–15882.

[Chung and Lu 02b] F. Chung and L. Lu. "Connected Components in Random Graphs with Given Expected Degree Sequences." *Annals of Combinatorics* 6 (2002), 125–145.

[Chung and Lu 06] F. Chung and L. Lu. "Concentration Inequalities and Martingale Inequalities: A Survey." *Internet Mathematics* 3 (2006), 79–127.

[Fu and Anderson 86] Y. T. Fu and P. W. Anderson. "Application of Statistical Mechanics to NP-Complete Problems in Combinatorial Optimisation." *J. Phys. A* 19 (1986), 1605–1620.

[Godbole and Hitczenko 98] A. Godbole and P. Hitczenko. "Beyond the Method of Bounded Differences." In *Microsurveys in Discrete Probability Princeton, NJ, 1997*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 41, pp. 43–58. American Mathematical Society, 1998.

[Hoeffding 63] Wassily Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables." *Journal of the American Statistical Association* 58:301 (1963), 13–30.

[Kim and Vu 04] J. H. Kim and V. H. Vu. "Divide and Conquer Martingales and the Number of Triangles in a Random Graph." *Random Structures and Algorithms* 24:2 (2004), 166–174.

[Kirkpatrick et al. 83] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi. "Optimization by Simulated Annealing." *Science* 220 (1983), 671–680.

[Lambiotte 10] R. Lambiotte. "Multi-scale Modularity in Complex Networks." In *Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 546–553, 2010.

[McDiarmid 89] C. McDiarmid. "On the Method of Bounded Differences." In *Surveys in Combinatorics*, London Math. Soc. Lectures Notes 141, pp. 148–188. Cambridge Univ. Press, 1989.

[Molloy and Reed 95] M. Molloy and B. Reed. "A Critical Point of Random Graphs with a Given Degree Sequence." *Random Structures and Algorithms* 6 (1995), 161–180.

[Newman 06] M. E. J. Newman. "Modularity and Community Structure in Networks." *Proc. Natl. Acad. Sci. USA* 103:23 (2006), 8577–8582.

[Reichardt and Bornholdt 04] J. Reichardt and S. Bornholdt. "Detecting Fuzzy Community Structures in Complex Networks with a Potts Model." *Phys. Rev. Lett.* 93 (2004), 218701.

[Rhee and Talagrand 87] Wansoo T. Rhee and Michel Talagrand. "Martingale Inequalities and NP-Complete Problems." *Mathematics of Operations Research* 12:1 (1987), 177–181.

[Sanov 57] I. N. Sanov "On the Probability of Large Deviations of Random Magnitudes." *Matem. Sbornik* 42:84 (1957), 11–44.

[Shamir and Spencer 87] E. Shamir and J. Spencer. "Sharp Concentration of the Chromatic Number on Random Graphs $G(n, p)$." *Combinatorica* 7:1 (1987), 121–129.

[Sherrington and Kirkpatrick 75] D. Sherrington and S. Kirkpatrick. "Solvable Model of a Spin-Glass." *Physics Review Letters* 35:26 (1975), 1792–1796.

[Watts and Strogatz 98] D. J. Watts and S. H. Strogatz. "Collective Dynamics of 'Small-World' Networks." *Nature* 393 (6684) (1998), 409–10.

[Wiethege and Sherrington 87] W. Wiethege and D. Sherrington. "Bipartitioning of Random Graphs of Fixed Extensive Valence." *J. Phys. A. Math. Gen. M* 20 (1987), L9–Lll.

[Wong and Sherrington 87] K. Y. M. Wong and D. Sherrington. "Graph Bipartitioning and Spin Glasses on a Random Network of Fixed Finite Valence." *J. Phys. A: Math. Gen.* 20 (1987), L793–L799.

Weituo Zhang, Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Amos Eaton 301, 110 8th Street, Troy, NY 12180, USA (zhangw14@rpi.edu)

Chjan C. Lim, Mathematical Sciences and NeST Center, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA (limc@rpi.edu)