Supplementary Material

September 6, 2015

1 Related works and optimization functions

Recently, some optimization algorithms based on Potts models which used to detect community structure have attracted attention. Communities correspond to Potts model spin states, and the associated system energy indicates the quality of a candidate partition. Let A be the adjacency matrix of graph G and let σ_i denote the label of the community that node i belongs to. Furthermore, the Kronecker Delta function is defined by $\delta(\sigma_i, \sigma_j) = 1$ if $\sigma_i = \sigma_j$ and $\delta(\sigma_i, \sigma_j) = 0$, otherwise. Having the community membership labels σ, Reichardt & Bornholdt (RB) [1] proposed a generalized Hamiltonian as the core energy function,

$$H_{RB}(\{\sigma\}) = -\frac{1}{2}\sum_{i\neq j}(a_{ij} - \gamma_{RB}p_{ij})\delta(\sigma_i, \sigma_j). \tag{1}$$

where γ_{RB} is the resolution parameter, $p_{ij} \in \mathbb{R}$ is the random form of adjacent matrix $A = (a_{ij})$. In general, two typical null models for statistical tests can be considered: (i) an Erdős-Rnyi null model (RBER) in which all edges are equally likely to be connected and the corresponding energy function can be expressed as follows,

$$H_{RBER}(\{\sigma\}) = -\frac{1}{2}\sum_{i\neq j}(a_{ij} - \gamma_{RB}p)\delta(\sigma_i, \sigma_j). \tag{2}$$

and (ii) the configuration null model (RBCM) in which edge connection probabilities are based on the current graphs degree distribution with the following corresponding energy function,

$$H_{RBCM}(\{\sigma\}) = -\frac{1}{2}\sum_{i\neq j}(a_{ij} - \gamma_{RB}\frac{k_i k_j}{2m})\delta(\sigma_i, \sigma_j). \tag{3}$$

where k_i denotes the degree of node i. It worth to mention that Modularity measure is special case of RBCM when $\gamma_{RB} = 1$.

Using a more general probabilistic model, Hofman & Wiggins [2] introduced the generalized energy function and then optimized it using a variational Bayes approach,

$$E_{HW}(\{\sigma\}) = -\frac{1}{2}\sum_{i\neq j}(W_L a_{ij} - W_G)\delta(\sigma_i, \sigma_j) + \frac{1}{2}\sum_{\mu=1}^{K}\pi_\mu \sum_{i=1}^{n}\delta(\sigma_i, \mu). \tag{4}$$

where $W_G = \log\frac{1-p_{out}}{1-p_{in}}$, $W_L = \log\frac{p_{in}}{p_{out}} + W_G$, $p_{in}(p_{out})$ is the probability that two nodes are connected when they are in the same (different) community, π_i is the prior probability of community. Recently, Ronhovde & Nussinov [3] introduced an energy function of the following form that is a local and resolution-limit free model.

$$E_{RN}(\{\sigma\}) = -\frac{1}{2}\sum_{i\neq j}(W_{ij}a_{ij} - \gamma W_{ij}\overline{a}_{ij})\delta(\sigma_i, \sigma_j). \tag{5}$$

where $\overline{a}_{ij} = 1 - a_{ij}$ when $i \neq j$ and $\overline{a}_{ii} = 0$. In addition, $W = (w_{ij})$ is a general weight matrix that assigns a weight to each existing and missing edges. In their work they proposed to select $p_{in,\mu} \geq \frac{\gamma}{1+\gamma}$ where $p_{in,\mu}$ is the probability that two nodes inside community μ are connected.

Label propagation is another famous algorithm for community detection [4]. Briefly, the algorithm starts with randomly assigning a community label to each node. Then, each node updates its label by replacing it by the label most used by its neighbors. It has been shown that the label propagation method is equivalent to finding the local energy minima of a simple zero temperature kinetic Potts model [5], i.e.

$$E_{KPM}(\{\sigma\}) = -\sum_{i\neq j} a_{ij}\delta(\sigma_i, \sigma_j). \tag{6}$$

The other well-known optimization approaches used in community detection problem are Simulated Annealing (SA) [6], extermal optimization (DA) [7], expectation maximization [8], Bayesian inference [9], and variational Bayes [2]. For a comprehensive and comparative review on this topic we refer the reader to [10].

2 The algorithm framework

Algorithm 1 The algorithm of detecting the position of center and corresponding community

Require: Graph G with size n and volume m, the algorithm parameters, i.e. f_μ^+, f_μ^- and R_μ in Eq.(4) in the Main text.

Ensure: The community membership matrix X;

1: For a given number of communities K
2: **repeat**
3: Calculate the top K eigenvector matrix $E_K = [e_1, e_2, ..., e_K]$ and initiate the community membership $X(0) = E_K$.
4: Update the position of center and corresponding community membership matrix X to minimize the Eq.(1) in the Main text.
5: **Until** exceeding the maximum number of iterations
6: Select the optimal number of communities K and corresponding community membership according to the maximum of Q defined in Eq.(4) in the Main text.

3 Distribution of community tightness.

In order to study the statistical properties of an arbitrary tightness score $S(x_1, ..., x_N)$ for nodes drawn independently from the distribution $P_0(x)$, the quality function is considered

$$\begin{array}{l} Z(\beta) = \Pi_{i=1}^N \int dx_i P_0(x_i) e^{\beta S(x_1,...,x_N)} \\ = \int dS p(S) e^{\beta S}. \end{array} \tag{7}$$

To study the computation procedure of $p(S)$, we consider the collection of all configurations of nodes set X with energy E, and $p(E)$ denotes the density of states as a function of E. If we replace the extensive energy with the intensive quantity, $E = Ne$, and use $p(E) = \frac{1}{N} p(e)$, there is

$$\begin{array}{l} \int p(E) e^{-\beta E} dE = \frac{1}{N} \int e^{-N\beta e + \log p(e)} \\ \simeq \frac{1}{N} e^{N sup_e (\log p(e)/N - \beta e)}. \end{array} \tag{8}$$

Next, if N is large enough, we use the saddle-point approximation and get

$$\log Z(\beta)/N = sup_e[logp(e)/N] - \beta e, \tag{9}$$

i.e. the normalised logarithm of the partition function, $\log Z(\beta)/N = -\beta f(\beta)$, is a Legendre transform (See the Section 5 in Supplementary Material) of the normalised logarithm of the probability, $\log p(e)/N$. Exploiting the duality of the Legendre transform, we get

$$\log p(e) \simeq -N sup_\beta[\beta f(\beta) + \beta e] = N[\beta_0 e - \beta_0 f(\beta_0)]. \tag{10}$$

with β_0 the saddle-point of the function in the squared brackets. Then, there is

$$\begin{array}{l} \log p(E) = \log p(e) + \log(\frac{1}{N}) \\ \simeq N[\beta_0 e - \beta_0 f(\beta_0)] + \log(\frac{1}{N}). \end{array} \tag{11}$$

Based on Eq.(11), given all configurations of nodes set $X = (x_1, ..., x_N)$ with a community tightness S, $p(S)$ denotes the density of states as a function of tightness S. Asymptotically for large N, we can extract this density function from $Z(\beta)$ based on Eq.(7) as

$$\log p(S) \simeq N\Omega(s) - \frac{1}{2}\log(gN). \tag{12}$$

Here $\Omega(s)$ is the entropy as a function of the tightness per element, i.e. $\Omega(s) = -\max_\beta[f(\beta) + \beta s]$. $\beta f(\beta) = -\log Z(\beta)/N$ is the free-energy density. We define the distribution of community tightness S as the probability $\int_S^{+\infty} p(S')dS'$, which can be used to find a score larger or equal to S. From analysis above, this is a typical p-value form and can be used to represent the statistical significance conveniently and directly.

4 The maximum entropy principle.

Assume that we are given a random variable x taking values $\{x_1, ..., x_n\}$, with an unknown probability distribution $p(x_i)$. Additionally, we are also given a prior information about the random variable: the expected value a of some property, here described by function $f(x_i)$,

$$E[f(x_i)] = \sum_{i=1}^{N} p(x_i) f(x_i). \tag{13}$$

The question is: what is the unbiased inference about the distribution $p(x_i)$? In other words, what is the distribution which does not reduce the amount of uncertainty about the random variable?

The *maximum entropy principle* states that the probability distribution should maximise the information entropy subject to the prior knowledge about the random variable. If there is no prior information, the solution is, quite intuitively, a uniform distribution assigning the same probability to every value of the random variable. In the presence of a constraint from Eq.(13) and given the normalisation constraint,

$$\sum_{i=1}^{N} p(x_i) = 1, \tag{14}$$

we can infer distribution $p(x_i)$ using the Lagrange multipliers. The solution is

$$p(x_i) = e^{-\lambda - \beta f(x_i)}, \tag{15}$$

where constants λ and β are inferred such that Eq.(13) and Eq.(14) are met. The solution can be written in an equivalent form as

$$p(x_i) = e^{-\beta f(x_i)} / Z(\beta), \tag{16}$$

where

$$\log Z(\beta) = \lambda, \tag{17}$$

$$Z(\beta) = \sum_{i} e^{-\beta f(x_i)} \tag{18}$$

and importantly

$$-\frac{\partial}{\partial \beta} \log Z(\beta) = \sum_{i} e^{-\beta f(x_i)} / Z(\beta) f(x_i) = E[f(x_i)] \tag{19}$$

Substituting x for X, a state of a physical system, and $f(x)$ for $H(X)$, the Hamiltonian of the system, we obtain the Boltzmann distribution. This shows that the Boltzmann distribution is the maximum entropy distribution for a system with a given observed energy value.

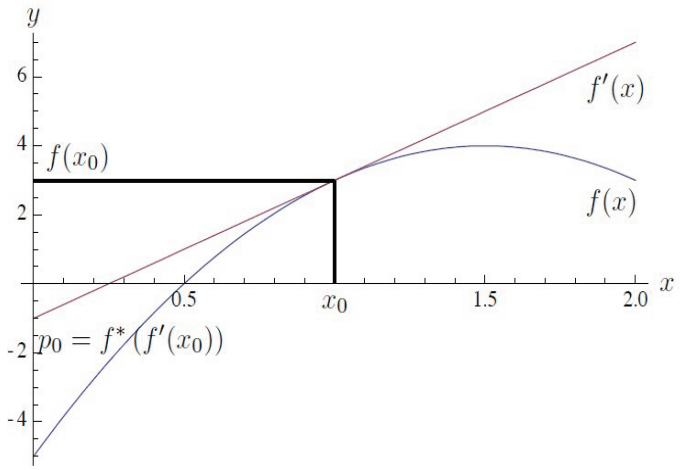


Figure 1: Legendre transform of a function. Function $f(x)$ (blue line) can be described by a set of points of the form $(x_0, f(x_0))$. A dual representation is achieved by using a function tangent to $f(x)$ at x_0, (red line). The new set of points is of the form $(f'(x_0)$, $f*(f'(x_0)))$, where $f*(y)$ is the Legendre transform of $f(x)$ and point $f*(f'(x_0))$ is an intercept of the tangent with y-axis.

5 The Legendre transform.

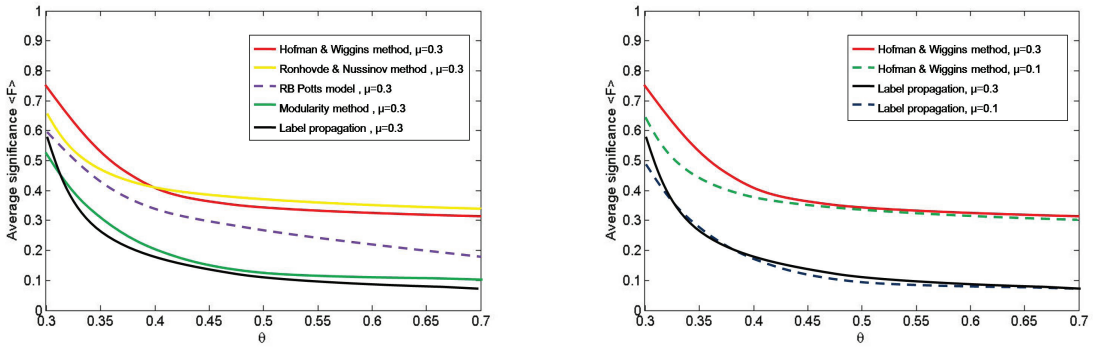
The Legendre transform of a real valued and differentiable function is an operation which gives a new, dual function f^*. The idea behind the transformation is that information about a functional relation, $(x_0, f(x_0))$, can be be equivalently expressed by another set of points of the form $(f'(x_0), p_0)$, where p_0 is an intercept of the line tangent to $f(x)$ at point x_0 and $f'(x) = \partial f(x)/\partial x$ is the derivative of function $f(x)$ over x, see Fig.1 for an illustration. The Legendre transform is formally defined as

$$f'(x) = \sup_x [xy - f(x)]. \tag{20}$$

To find a supremum of $(xy - f(x))$ with respect to x, we solve

$$\frac{\partial}{\partial x}(xy - f(x)) = 0, \tag{21}$$

which is met by $y = \frac{\partial}{\partial x} f(x) = f'(x)$. The intercept of the tangent to function $f(x)$ at x_0 is then $f^*(f'(x_0))$, so the point $(x_0, f(x_0))$ is now mapped to a point $(f'(x_0), f^*(f'(f(x_0)))$.



(a) (b)

Figure 2: The performance of significance $\langle F\rangle$ on LFR benchmark network and each point in curves is obtained by testing 100 times. (a) For all five algorithms, the $\langle F\rangle$ index decreases with the increasing of mix parameter θ. When $\theta \geq 0.5$ on average (no significant community), $\langle F\rangle$ is near 0.3 which is similar with GN network. (b) The value of $\langle F\rangle$ corresponding to $\mu = 0.3$ will be larger than $\mu = 0.1$ for the Hofman & Wiggins method and Label propagation method.

An important property of the Legendre transform is its duality: function f is also a Legendre transform of $f*$,

$$f(x) = \sup_y [xy - f^*(x)]. \tag{22}$$

As we will show later in this chapter, the intensive entropy and the intensive free energy of a system are in such a dual relation,

$$\omega(e) = \sup_\beta [\beta e - \beta f(\beta)]. \tag{23}$$

6 Experiments

A. LFR benchmark network. We also test the index on the more challenging LRF benchmark presented by Lancichinetti, Fortunato and Radicchi [11]. In the LFR benchmark, each node is given a degree took from a power law distribution with an exponent γ, and the sizes of the communities are took from a power law distribution with an exponent β. Moreover, each node shares a fraction $1-\theta$ of its links with other nodes of its community and a fraction θ with other nodes in the network, θ is the mixing parameter.

In this network, the average degree $k = 20$, maximum degree is 50 and $P(k) \propto k^\gamma$.

Maximum and minimum community sizes are 50 and 20 respectively. The significance score changes when we adjust the value of θ in LFR benchmark, and numerical results in the LFR-benchmark are shown in Fig.2(a). It can be observed that with the augment of θ, F decreases for all five optimization methods when $\mu = 0.3$. Same as GN network, the $\langle F \rangle$ values corresponding to Hofman & Wiggins method is largest at the begining, and the Label propagation method is the lowest. However, the $\langle F \rangle$ values corresponding to Ronhovde & Nussinov method will exceed Hofman & Wiggins method when when θ is larger than 0.4. Furthermore, when θ larger than 0.32, the $\langle F \rangle$ value corresponding to Label propagation method is close to Modularity optimization method. In addition, from Fig.2(b), it can be observed the value of $\langle F \rangle$ corresponding to $\mu = 0.3$ will larger than $\mu = 0.1$ when we take the Hofman & Wiggins method and Label propagation method as examples.

B. Stochastic block model. Finally, we consider the famous stochastic block model which used to detect community structure by Decelle and Zhang et al [12] [13] [14]. In this benchmark, each node i has a hidden label $t_i \in \{1, ..., q\}$, specifying which of q groups it is a member of. These labels are chosen independently, where y_a is the probability that a given node has label $a \in \{1, ..., q\}$ (normalized so that $\sum_{a=1}^{q} y_a = 1$). For each pair of nodes i, j with $i < j$, we put an edge between i and j independently with probability p_{t_i,t_j}, leaving them unconnected with probability $1 - p_{t_i,t_j}$. Our goal is to learn the parameters $q, \{y_a\}, \{p_{ab}\}$ of the block model, as well as the true group assignments $\{t_i\}$. Special cases of this model have often been considered in the literature. Here, a special case is considered, i.e. the planted partitioning, when $y_a = 1/q, c_{ab} = c_{out}$ for $a \neq b$ and $c_{aa} = c_{in}$ with $c_{in} > c_{out}$, is a classical problem in computer science and has been used as a benchmark for community detection. Here, $\varepsilon = c_{out}/c_{in}$ is the parameter used to control the fuzziness of generated network.

To verify the performance on sparse stochastic block model with low average degree, we generate a large network with $N = 5000$ nodes and $q = 10$ groups with average degree

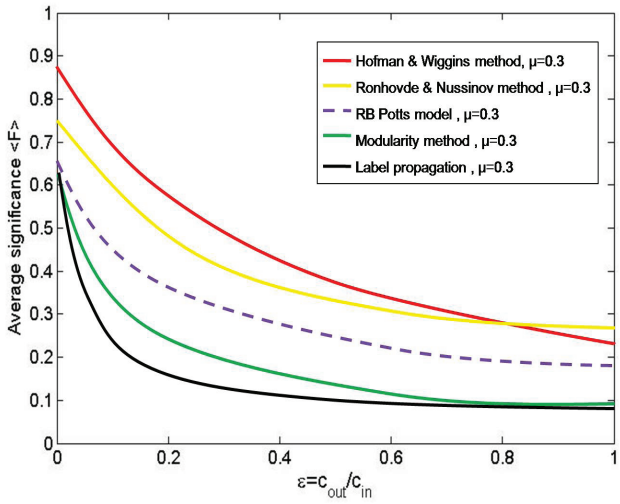


Figure 3: The performance of social significance $\langle F\rangle$ on stochastic block model. In this example, there are $N = 5000$ nodes and $q = 10$ groups. The average degree $c = 8$ and parameter $\varepsilon = c_{out}/c_{in}$ is used to control the fuzziness of generated network. Each point in curves is obtained by testing 100 times. With the increasing of ε, the $\langle F\rangle$ index decreases. For all algorithm, when the corresponding $\langle F\rangle$ is nearly larger than 0.3 on average($\varepsilon \approx 0.4$), there exists significant community structure which may detectable.

$c = 8$, which shown in Fig.3. Each point in curves is the result averaged by testing 100 times. When ε is close to 0, it can be observed the community structure is quite strong and the corresponding $\langle F\rangle$ value of all five algorithms are very high when $\mu = 0.3$. In contrast, when ε is increased close to 0.8, the network is nearly a fuzzy random one, and all $\langle F\rangle$ values are very low, near 0.1-0.3. Furthermore, we find that the $\langle F\rangle$ value of Hofman & Wiggins method will larger than all others when $\varepsilon < 0.81$, while lower than Ronhovde & Nussinov method when $\varepsilon > 0.81$. Specifically, we argue that for all algorithm when the corresponding $\langle F\rangle$ is nearly larger than 0.3 on average($\varepsilon \approx 0.4$), there exists significant community structure which may detectable [12]. Form the results, the F shows a great ability in characterizing the significant modular structure for optimization methods as we adjust the parameter ε.

C. Real network. Finally, we show significance can also be used to rank the real network partitions obtained by different algorithmic strategies. Zachary karate club network, Collage football network and Political books network are employed as the examples.

Tabel 1 presents the results estimated from three algorithms and we observed that they are coincided with the analysis in artificial networks. These observations are no evidence of overall superiority of one method over another, but an example of how to compare the significance and use the different partitioning algorithms on a given network.

Table 1: Comparison of various algorithms with $\langle F \rangle$ values.

Networks	Algorithms	Values of $\langle F \rangle$
Zachary network	Label propagation method	0.641
	Girvan-Newman algorithm	0.735
	RB Potts	0.827
Collage football network	Label propagation method	0.602
	Girvan-Newman algorithm	0.758
	RB Potts	0.831
Political books network	Label propagation method	0.581
	Girvan-Newman algorithm	0.698
	RB Potts	0.717

References

[1] J. Reichardt and S. Bornholdt. Statistical mechanics of community detection, *Physical Review E*, 74(1), 016110, (2006).

[2] J. M. Hofman and C. H. Wiggins. Bayesian Approach to Network Modularity, *Physical Review Letters*, 100(25), 258701, (2008).

[3] P. Ronhovde and Z. Nussinov. Local resolution-limit-free Potts model for community detection, *Physical Review E*, 81(4), 046114, (2010).

[4] U. N. Raghavan, R. Albert and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks, *Physical Review E*, 76(3), 036106, (2007).

[5] G. Tibély and J. Kertész. On the equivalence of the label propagation method of community detection and a Potts model approach, *Physica A: Statistical Mechanics and its Applications*, 387(19-20), 4982-4984, (2008).

[6] R. Guimera and L. A. Nunes Amaral. Functional cartography of complex metabolic networks, *Nature*, 433(7028), 895-900, (2005).

[7] J. Duch and A. Arenas. Community detection in complex networks using extremal optimization, *Physical Review E*, 72(2), 027104, (2005).

[8] M. E. J. Newman and E. A. Leicht. Mixture models and exploratory analysis in networks, *Proceedings of the National Academy of Sciences of the United States of America*, 104(23), 9564-9569, (2007).

[9] M. B. Hastings. Community detection as an inference problem, *Physical Review E*, 74(3), 035102, (2006).

[10] A. Lancichinetti and S. Fortunato. Community detection algorithms: A comparative analysis, *Physical Review E*, 80(5), 056117, (2009).

[11] A. Lancichinetti, S. Fortunato and F. Radicchi. Benchmark graphs for testing community detection algorithms, *Physical Review E*, 78(4), 046110-046115, (2008).

[12] A. Decelle, F. Krzakala, C. Moore and L. Zdeborová. Phase transition in the detection of modules in sparse networks, *Physical Review Letters*, 107(6), 065701, (2011).

[13] A. Decelle, F. Krzakala, C. Moore and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, *Physical Review E*, 84(6), 066106, (2011).

[14] P. Zhang and C. Moore. Scalable detection of statistically significant communities and hierarchies, using message-passing for modularity, *Proceedings of the National Academy of Sciences of the United States of America*, 111, 18144, (2014).