A corrected normalized mutual information for performance evaluation of community detection

Darong Lai[1, 2*], **Christine Nardini**[3, 4]

1. School of Computer Science and Engineering, Southeast University, Nanjing, China
2. Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, China.
3. Group of Clinical Genomic Networks, Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Shanghai, PR China
4. University of Chinese Academy of Sciences, Beijing, PR China

* Email: darong.lai@gmail.com

Abstract. Normalized mutual information (NMI) is a widely used metric for performance evaluation of community detection methods, recently proven to be affected by finite size effect. To overcome this issue, a metric called relative normalized mutual information (rNMI) has been proposed. However, we show here that rNMI is still a biased metric and may lead, under given circumstances, to erroneous conclusions. The bias is an effect of the so called *reverse finite size effect*. We discuss different strategies to address this issue, and then propose a new metric the *corrected normalized mutual information* (cNMI), symmetric and well normalized, in the form of empirical calculation and closed-form expression. The experiments show that cNMI not only removes the finite size effect of NMI but also the reverse finite size effect of rNMI, and is hence more suitable for performance evaluation of community detection methods and for other approaches typical of the more general clustering context.

I. Introduction

Community detection in a complex network corresponds to the partition of the network into groups of nodes with much denser connections within each group than among such groups. This represents an important and challenging problem in network structural analysis, applied in a variety of fields including sociology, biology, computer science, to name a few. Plenty of different community detection methods have thus been designed [1] and, as a consequence, testing and comparing performances is of great importance and usefulness in facilitating end-users, especially those outside of the fields of physics and computer science, to effectively improve or apply community detection methods in their own area.

Performance testing and comparison of different community detection methods is by no means a trivial task due to the different motivations and intended applications. As a result, determining if one community detection method is better than another is usually done in a simplified way by employing canonical benchmarks like computer generated synthetic networks and real-world networks where node classification is

known. Benchmark networks are generated with arbitrarily designed built-in community structures or selected with already known correct community structure (termed *known partition* of the network). The accuracy is usually computed in terms of similarity by comparing the solution delivered by the method (termed *computed partition* of the network) with the known partition of the network. The larger the similarity the better the performance of the method on the given benchmark is. Synthetic networks with built-in community structure generated by statistical models like stochastic block model [2, 3], planted partition model [4] or its variant proposed by Lancichinetti, Fortunato, and Radicchi (denoted as LFR model) [5], and real-world networks as karate club network [6], football network [7], politics network [8] to name a few, are routinely used as benchmark networks. With these well-known benchmarks, performance testing thus involves defining a metric to establish how similar the computed partition is to the known partition.

A metric designed for this purpose must have a number of desirable properties. First, it should equal 1 when the computed partition is identical to the known partition. Second, when the computed partition is a random permutation of the known partition, the value of their similarity should approach zero. Third, it should be symmetric, i.e. the similarity computed by comparing the computed partition with the known partition should be the same as that computed by comparing the known partition with the computed partition.

Various such similarity metrics have been previously proposed. The *Rand* index, based on pair counting scheme, defines the similarity of two partitions as the ratio of the number of vertex pairs correctly classified in both partitions by the total number of vertex pairs [9]. The *Jaccard* index [10] and the *adjust rand index* [11] are other popular pair counting based metrics. If the number of groups of the two partitions is identical, the similarity can be calculated by the so called *overlap,* scaling from 0 to 1, which is the number of identical group labels between two partitions maximized over all possible permutations and properly normalized [3]. Within the framework of information theory, the problem of comparing partitions can be reformulated as a problem of message decoding with the assumption that if two partitions are similar, one needs very little information to infer one partition given the other. *Normalized mutual information* (NMI) proposed by Danon et al. [12], a popular and currently frequently adopted metric in the testing of community detection methods, defines the similarity of two partitions as the mutual information of the two partitions normalized by the entropies of the partitions. Different from overlap metric, NMI does not need the number of groups of the two partitions be identical. Similar to NMI, *variation of information* (VI) introduced by Meila [13] is another information-theoretic metric that has the properties of distance and is a local measure. However, the maximum of VI depends on the number of nodes of the networks and thus is not constant, making the similarity values of partitions of network with different number of nodes not comparable. A similar situation exists for other kind of metrics like modularity [14], conductance [15] to name a few, which are not similarity metrics of two partitions but that are often used as cost functions to find optimal partitions of a network. Although these metrics can be used to decide which of two or more partitions is the best, their

values depend strongly on the properties of the networks. As a result, metrics derived from cost functions cannot be used to evaluate the performance of the community detection methods in a systematic and consistent way. This differs from NMI used with available benchmarks, since NMI has bounded values [0, 1] on different networks and thus facilitates an easier explanation of the results. For its importance and the ascending popularity, we, in this paper, focus on the discussions of NMI as a mean for testing the performance of community detection methods. For more comprehensive introductions to various other metrics, the reader is referred to Refs. [1, 13].

Because of its useful properties NMI has quickly become a very popular performance measure soon after its introduction [1, 16, 17, 18]. However, it was recently observed that NMI may suffer from serious systematic errors in finite size networks [19]. Such errors are the result of using frequencies to approximate the probability of a randomly selected node to be in a community, when computing entropies. To overcome the finite size effect of NMI, Zhang proposed the *relative normalized mutual information* (rNMI) as an alternative similarity metric of two partitions of a network [19]. Although rNMI is designed to remove the bias in NMI, we found that it introduces another bias which we termed *reverse finite size effect* to differentiate it from the one of NMI. To overcome both the biases of NMI and rNMI, we propose a new similarity metric named *corrected mutual normalized information* (cNMI) and demonstrate its behaviors on the performance testing of popular community detection methods.

In the following, we first introduce the background concepts of contingency table and NMI in section II, and then demonstrate the bias of rNMI and explain why it occurs in section III. In section IV, we propose the new metric cNMI and compare it with other metrics on benchmark networks. We also derive an explicit expression of cNMI, and compare it with the empirical scheme of computing cNMI in section V. We give applications of cNMI in a more general context in section VI. Based on the above results and discussions, we then draw conclusions.

II. Contingency Table and Normalized Mutual Information

Given a network with N nodes $V=\{v_1, \ldots, v_N\}$, the relationship between two partitions X and Y of the network can be established through a contingency table. Suppose $\mathrm{X}=\{\mathrm{X}_1,\cdots,\mathrm{X}_{C_X}\}$ and $\mathrm{Y}=\{\mathrm{Y}_1,\cdots,\mathrm{Y}_{C_Y}\}$ represent respectively the partitions of nodes in V into C_X and C_Y communities, the elements in X and Y are then all not-empty subsets of the node set V, and are called communities or groups satisfying: i) $\mathrm{X}_r \cap \mathrm{X}_s = \emptyset$ for $1 \le r \ne s \le C_X$; ii) $\bigcup_{r=1}^{C_X} \mathrm{X}_r = V$; iii) $\mathrm{Y}_{r'} \cap \mathrm{Y}_{s'} = \emptyset$ for $1 \le r' \ne s' \le C_Y$; iv) $\bigcup_{r'=1}^{C_Y} \mathrm{Y}_{r'} = V$. If n_{rs} denotes the number of nodes in X_r of partition X that appear in Y_s of partition Y, the overlap between the two partitions X and Y can be written in the form of a contingency table as in Table 1, where $n_{r\cdot} = \sum_s n_{rs}$ and $n_{\cdot s} = \sum_r n_{rs}$ are respectively the number of nodes in X_r and Y_s.

Table 1. Contingency table of two partitions

		Partition Y				
Partition X	Community /group	Y_1	Y_2	…	Y_{C_Y}	$n_{r\cdot} = \sum_s n_{rs}$
	X_1	n_{11}	n_{12}	…	n_{1C_Y}	$n_{1\cdot}$
	X_2	n_{21}	n_{22}	…	n_{2C_Y}	$n_{2\cdot}$
	.	.	.	…	.	.
	.	.	.	…	.	.
	.	.	.	…	.	.
	X_{C_X}	n_{C_X1}	n_{C_X2}	…	$n_{C_XC_Y}$	$n_{C_X\cdot}$
	$n_{\cdot s} = \sum_s n_{rs}$	$n_{\cdot 1}$	$n_{\cdot 2}$	…	$n_{\cdot C_Y}$	$N = \sum_{rs} n_{rs}$

With the notations presented in Table 1, the normalized mutual information (NMI) is defined as [12]:

$$\mathrm{NMI}(X,Y) = \frac{-2\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y} n_{rs}\log\left(\frac{N\cdot n_{rs}}{n_{r\cdot}\cdot n_{\cdot s}}\right)}{\sum_{r=1}^{C_X} n_{r\cdot}\log\left(\frac{n_{r\cdot}}{N}\right)+\sum_{s=1}^{C_Y} n_{\cdot s}\log\left(\frac{n_{\cdot s}}{N}\right)}. \quad (1)$$

From the classical or frequentist interpretation of probability, the probabilities are viewed in terms of the frequencies of random, repeatable events. Consequently, the probability of a randomly selected node being in community/group X_r (or Y_s) is $P(r) = \frac{n_{r\cdot}}{n}$ (or $P(s) = \frac{n_{\cdot s}}{n}$), and similarly $P(r,s) = \frac{n_{rs}}{n}$ is the probability of a randomly selected node in group X_r being in Y_s. NMI can thus be reformulated as a function of probabilities:

$$\mathrm{NMI}(X,Y) = \frac{-2\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y}\frac{n_{rs}}{N}\log\left(\frac{\frac{n_{rs}}{N}}{\frac{n_{r\cdot}}{N}\cdot\frac{n_{\cdot s}}{N}}\right)}{\sum_{r=1}^{C_X}\frac{n_{r\cdot}}{N}\log\left(\frac{n_{r\cdot}}{N}\right)+\sum_{s=1}^{C_Y}\frac{n_{\cdot s}}{N}\log\left(\frac{n_{\cdot s}}{N}\right)} = \frac{2\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y} P(r,s)\log\left(\frac{P(r,s)}{P(r)\cdot P(s)}\right)}{-\sum_{r=1}^{C_X} P(r)\log P(r)-\sum_{s=1}^{C_Y} P(s)\log P(s)}. \quad (2)$$

Here $-\sum_{r=1}^{C_X} P(r)\log P(r) \triangleq H(X)$ is the entropy of the distribution of partition X, and $\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y} P(r,s)\log\left(\frac{P(r,s)}{P(r)\cdot P(s)}\right) = H(X)+H(Y)-H(X,Y) \triangleq I(X,Y)$ is the mutual information between partitions X and Y. When $P(r,s) = P(r)P(s)$ for each pair r and s, the two partitions X and Y are independent, and thus their NMI reaches its minimal value zero. Conversely, for two identical partitions, i.e. when for each community r in X there is only one matching community s in Y and vice versa (in this case, $P(r,s) = P(r) = P(s)$ or $n_{r\cdot} = n_{\cdot s} = n_{rs}$), NMI's value reaches its maximal value, one. It can be easily verified that NMI is a symmetric measure. As a result, NMI fulfills the desirable requirements for a useful similarity measure.

III. The bias of NMI and its improved variant rNMI

Despite its advantages, NMI was recently shown to suffer from serious systematic errors in finite size networks [19]. Such errors are the result of using the

frequency $\frac{n_{r\cdot}}{N}$ to approximate the probability $\mathrm{P}(r)$ of a randomly selected node to be in community r, when computing entropies. Under the assumption of the Bernoulli distribution of community size $n_{r\cdot}$, NMI between two random partitions X and Y of a network with a finite number N of nodes can be approximated as [19]:

$$NMI_N^{random}(\mathrm{X},\mathrm{Y}) \cong \frac{1}{2N}\frac{C_X C_Y - C_X - C_Y + 1}{H(X)+H(Y)} \tag{3}$$

It can be inferred from equation (3) that two random partitions have a non-vanishing NMI. Owing to its origin, this phenomenon is termed *finite size effect*. Such an effect forces NMI to prefer a partition with larger number of communities, leading to a biased performance evaluation.

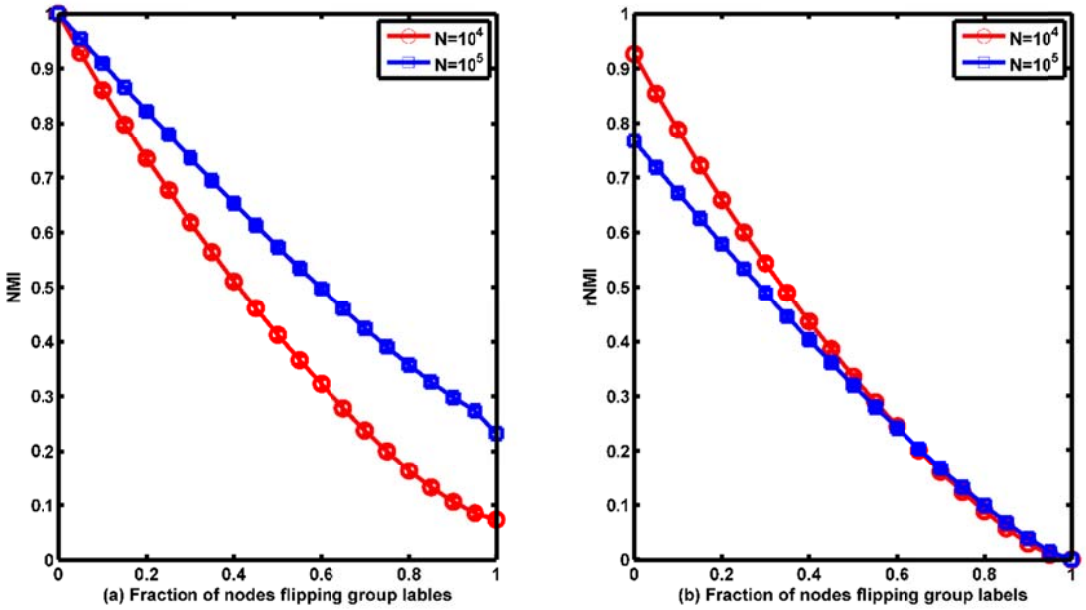

Figure 1. NMI-normalized mutual information (a) and rNMI-relative normalized mutual information (b) between two partitions with the computed partition obtained by flipping the group labels of a fraction of nodes in the known partition. The known partitions were generated with group sizes ranging from 20 to 400 in two different network sizes. Each point in the figure is averaged over 10 network instances and 10 repetitions.

By construction, the expected value of NMI between two partitions should decline with the dissimilarity of the two partitions, scaling from 1 to 0. To see more concretely how NMI is affected by the finite size effect, we generated known partitions with group sizes ranging from 20 to 400 and following power law distributions [5], and then obtained computed partitions from these known partitions by flipping the group labels of a fraction of nodes. When no group label is flipped, the NMI value between computed and known partitions is strictly 1, and declines, as expected, with the growing fraction of flipping nodes. However, the finite size effect prevents the NMI value from being zero when the computed partition becomes a random permutation of the known partition (see Figure 1(a), in this case the fraction equals 1). Since, under the modeling used here, a larger network comes with a larger number of communities, the NMI between a partition and its random permutation is

farther apart from 0 than in a smaller network, indicating that a network with a larger number of communities is more heavily affected by the finite size effect.

To overcome the finite size effect of NMI, Zhang proposed the ***relative normalized mutual information*** (rNMI), defined as [19]:

$$rNMI(X,Y) = NMI(X,Y) - < NMI(X,Z_Y) >, \quad (4)$$

where Z_Y is a random partition that has the same community-size distribution as partition Y, and $< NMI(X,Z_Y) >$ is the expectation of NMI between X and Z_Y. Practically, $< NMI(X,Z_Y) >$ is usually computed by averaging on a number of realizations of Z_Y. Due to the symmetry of NMI, rNMI is also symmetric. The finite size effect leads NMI to give higher value of a partition with larger number of communities. That is if partition Y has a larger number of communities, the value of $NMI(X,Y)$ will be larger, leading to a biased evaluation. Since the value of $NMI(X,Z_Y)$ between partition X and the random permutation Z_Y of partition Y will also be large, the bias will eventually be removed under the adjustment proposed for rNMI. As a consequence, rNMI has a vanishing value when two partitions X and Y are independent even if Y has a large number of communities, as it can be seen from the bottom right corner of Figure 1(b).

Although rNMI overcomes the finite size effect of NMI, we observed here that it suffers from another undesirable issue. In fact, rNMI is not a well normalized metric. In Figure 1(b) (see the upper left corner), it is shown that even if there is no group label flipping, rNMI computed for a partition against itself is smaller than 1, which is counterintuitive and practically undesirable.

We thus generated a series of synthetic partitions with equal size groups in network of different sizes, and with the number of groups ranging from 2 to N/5, to systematically reveal the defect of rNMI. As shown in Figure 2, rNMI values between two identical partitions decline with the increasing number of groups of partitions for different network sizes. This effect becomes more visible with smaller networks. For instance, when the number of groups is 10^2, rNMI between two identical partitions of a network with 10^3 nodes is the smallest and approaches zero. Furthermore, if the number of groups is large enough, the value between two identical partitions can be negligible.

To describe how this happens and simplify the analysis, we here assume that the distribution of the variable n_r. of group size follows a Bernoulli distribution, and that partitions X and Y are identical and both have C communities of equal size K and Z_Y is the random permutation of partition Y. From equation (3), the average NMI between X and Z_Y can be approximated by

$$\mathrm{NMI}_N(\mathrm{X},\mathrm{Z}) \cong \frac{1}{2CK}\frac{C^2-2C+1}{2logC} \quad (5)$$

As a result, the rNMI between X and Y is approximated to

$$\begin{aligned} rNMI(X,Y) &= rNMI(X,X) \\ &= NMI(X,X) - < NMI(X,Z_Y) > \\ &= 1 - \frac{1}{2CK}\frac{C^2-2C+1}{2logC} \end{aligned}$$

$$= 1 - \frac{1}{4K}\frac{C - 2 + \frac{1}{C}}{logC}$$

Since $C \gg logC$, it is obvious that $rNMI(X,Y)$ declines with increasing values of C. In contrast to the finite size effect of NMI that biases to a non-vanishing NMI value for two random partitions with a large number of groups, rNMI biases towards a value unable to maximize even for two identical partitions with a large number of groups, i.e. "1 - rNMI" is a non-vanishing value. When C is fixed, this bias is more evident in a smaller network since rNMI declines with the decreasing value of K. The two types of effect are both due to the finite size of a network, and we thus similarly termed such effect of rNMI as ***reverse finite size effect***.

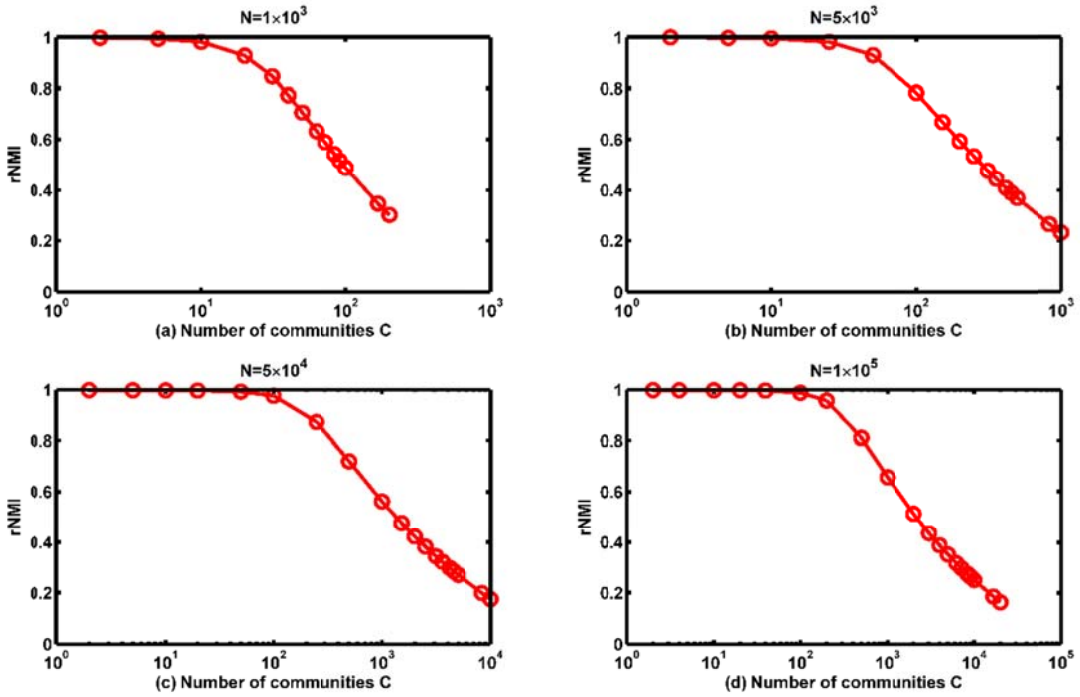

Figure 2. Relative normalized mutual information (rNMI) between identical partitions versus the number of communities in the corresponding partitions. The partitions were generated with different number of communities in different network sizes. (a) The partitions of networks with $N = 1x10^3$ nodes; (b) The partitions of networks with $N = 5x10^3$ nodes; (c) The partitions of networks with $N = 5x10^4$ nodes; (d) The partitions of networks with $N = 1x10^5$ nodes. Each point in the figure is averaged over 10 network instances.

IV. The proposed corrected normalized mutual information

To remove the new bias introduced by rNMI, a simple and naive solution to overcome this latter issue is to properly normalize rNMI:

$$rNMI_{norm}(X,Y) = \frac{rNMI(X,Y)}{rNMI(X,X)}, \quad (6)$$

a solution also adopted in a recent work and there called *relative rNMI* [20]. The denominator of the above equation has vanishing probability to be zero if the size of a network and the number of groups is large enough, making the normalization possible. Consequently, for two identical partitions of a network, the value of $rNMI_{norm}$

theoretically equals 1. Since rNMI gives vanishing value of two random partitions, $rNMI_{norm}$ will thus maintain this property. However, $rNMI_{norm}$ is no longer symmetric with such normalization. As discussed previously, rNMI is symmetric, but the denominator of $rNMI_{norm}(\mathrm{Y},\mathrm{X})$ will be different from that of $rNMI_{norm}(\mathrm{X},\mathrm{Y})$ unless partition Y is identical to partition X. The asymmetry of $rNMI_{norm}$ rules out the possibility to have it as a desirable metric for similarity. In addition, the maximal value of $rNMI_{norm}$ is not strictly equal to 1 but mildly larger, which makes the explanation of the value of $rNMI_{norm}$ between two identical partitions less rigorous.

We thus call for a new metric that should not be affected by any of the finite size effects, and that is symmetric. In this paper, we introduce the ***corrected normalized mutual information*** (cNMI) and define it as:

$$cNMI(X,Y) = \frac{rNMI(X,Y)+rNMI(Y,X)}{rNMI(X,X)+rNMI(Y,Y)}. \quad (7)$$

Let Z_X and Z_Y be, respectively, the random permutations that have the same group-size distribution as partition X and Y, exploiting the symmetry of rNMI, cNMI can be simplified as

$$cNMI(X,Y) = \frac{2NMI(X,Y)-<NMI(X,Z_Y)>-<NMI(Y,Z_X)>}{2-<NMI(X,Z_X)>-<NMI(Y,Z_Y)>}, \quad (8)$$

taking advantage of the fact that NMI between two identical partitions is 1.

It can be verified that cNMI has the desirable properties of a similarity metric of partitions. First, cNMI removes the finite size effect so that the cNMI value between two random partitions can reach zero, even if the partitions have a large number of groups. Suppose that partition Y is a random partition with regard to partition X, leading to the fact that $NMI(X,Y) \cong < NMI(X,Z_Y) >$ and $(Y,X) \cong < NMI(Y,Z_X) >$, then

$$cNMI(X,Y) = \frac{2NMI(X,Y)-< NMI(X,Z_Y) > -< NMI(Y,Z_X) >}{2-< NMI(X,Z_X) > -< NMI(Y,Z_Y) >}$$
$$\cong \frac{< NMI(X,Z_Y) > +< NMI(Y,Z_X) > -< NMI(X,Z_Y) > -< NMI(Y,Z_X) >}{2-< NMI(X,Z_X) > -< NMI(Y,Z_Y) >}$$
$$= 0$$

Second, the value of cNMI between two identical partitions is strictly equal to 1, and thus is not influenced by the reverse finite size effect.

$$cNMI(X,X) = \frac{2NMI(X,X)-< NMI(X,Z_X) > -< NMI(X,Z_X) >}{2-< NMI(X,Z_X) > -< NMI(X,Z_X) >}$$
$$= \frac{2-<NMI(X,Z_X)>-<NMI(X,Z_X)>}{2-<NMI(X,Z_X)>-<NMI(X,Z_X)>} = 1.$$

Finally, cNMI is symmetric due to the symmetry of NMI.

$$cNMI(X,Y) = \frac{2NMI(X,Y)-< NMI(X,Z_Y) > -< NMI(Y,Z_X) >}{2-< NMI(X,Z_X) > -< NMI(Y,Z_Y) >}$$
$$= \frac{2NMI(Y,X)-< NMI(Y,Z_X) > -< NMI(X,Z_Y) >}{2-< NMI(Y,Z_Y) > -< NMI(X,Z_X) >}$$
$$= cNMI(Y,X)$$

As shown in Figure 3, cNMI is well normalized and scales from assigning to value 0 the similarity of two random partitions to assigning to value 1 the similarity of two

identical partitions.

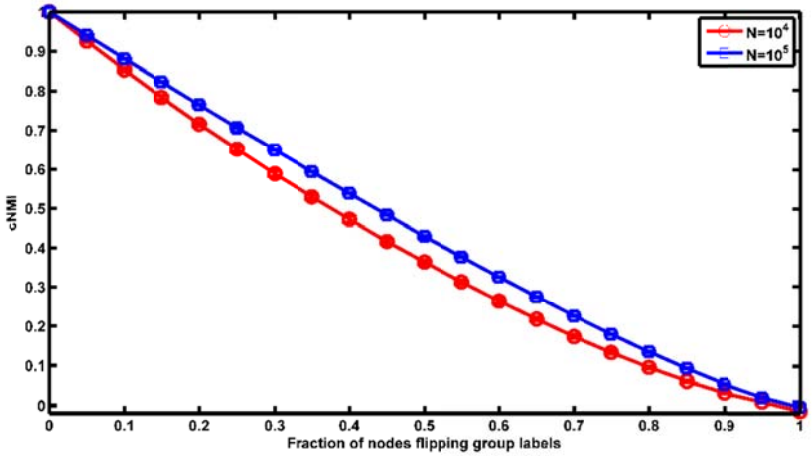

Figure 3. Corrected normalized mutual information (cNMI) between two partitions with the computed partition obtained by flipping the group labels of a fraction of nodes in the known partition. The known partitions were generated with group sizes ranging from 20 to 400 in two different network sizes. Each point in the figure is averaged over 10 network instances and 10 repetitions.

Application of such findings is shown here on networks generated using stochastic block model (SBM), a popular and important benchmark model for testing community detection methods [2, 3]. To generate a commonly studied network with C communities of equal size, SBM assigns the probability p_{in} to connect a pair of nodes in the same community and p_{out} if they are in different communities. $\varepsilon = p_{out}/p_{in}$ thus reflects the strength of the community structure in a network. If $\varepsilon = 1$, the partition of an SBM network is said to be in the *undetectable phase*, and no algorithm can group nodes better than chance [3, 21]. For such cases, rNMI can give zero value even when algorithms wrongly detect communities while NMI fails due to the finite size effect. We thus used cNMI to reevaluate the performance of two efficient and popular community detection methods: InfoMap and Louvain method. InfoMap aims at finding a partition that produces an optimal compression of information diffusion on a network [22]. Louvain method iteratively conducts a two-step procedure of nodes moving and aggregation to search an optimal partition that maximizes the modularity [23]. The benchmark networks generated have C=100 communities. As shown in Figure 4, in contrast to NMI, cNMI also reports zero value as rNMI does for InfoMap and Louvain method in detecting communities in these networks, meaning that there is no community structure in these networks.

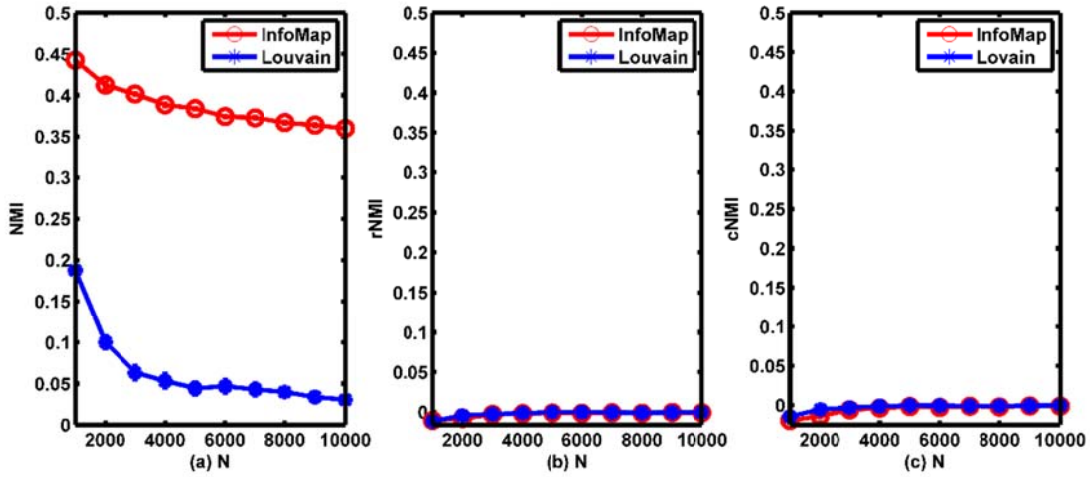

Figure 4. The values of different metrics for evaluating the performance of InfoMap and Louvain method on benchmark networks generated by stochastic block model with $\varepsilon = 1$ and the number of communities C=100. (a) Normalized mutual information (NMI); (b) Relative normalized mutual information (rNMI); (c) Corrected normalized mutual information (cNMI). Each point in the figure is averaged over 10 instances of the model in each network size N.

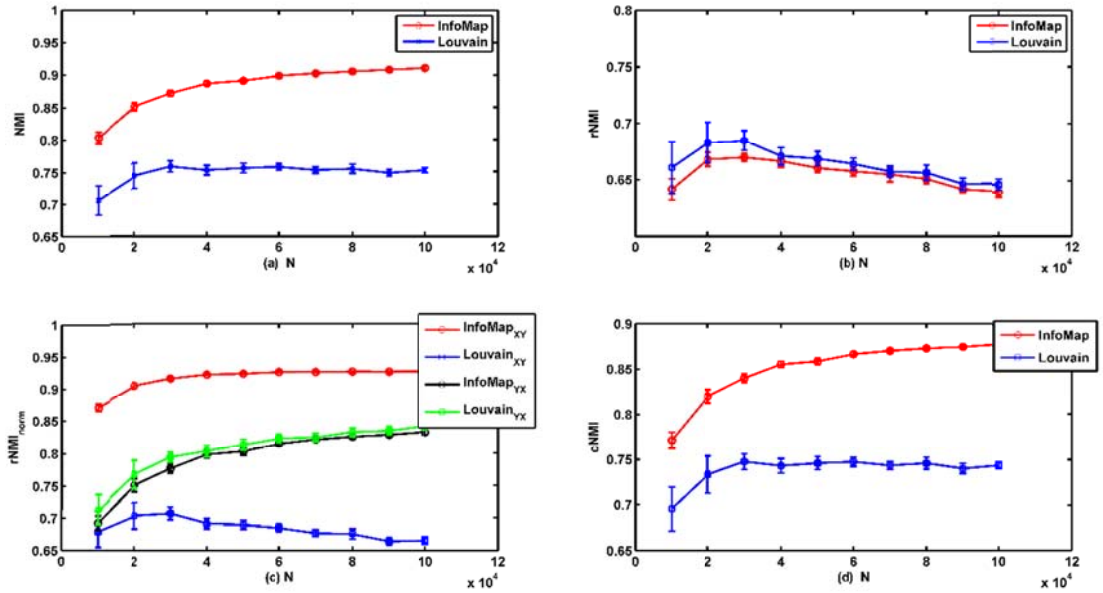

Figure 5. Performance evaluations of InfoMap and Louvain method on benchmark networks generated with different sizes by LFR model. (a) Performance evaluation under normalized mutual information (NMI); (b) Performance evaluation under relative normalized mutual information (rNMI); (c) Performance evaluation under relative normalized mutual information with naive normalization ($rNMI_{norm}$); $InfoMap_{XY}$ represents $rNMI_{norm}$ (X, Y) and $InfoMap_{YX}$ is $rNMI_{norm}$ (Y, X) , where Y is the partition obtained by InfoMap and X is the ground truth partition. Symbols for Louvain method are similar. (d) Performance evaluation under corrected normalized mutual information (cNMI). Each point in the figure is averaged over 10 instances of the model in each network size N.

Recently, Zhang [19] reported that by rNMI, the performance of Louvain method was better than InfoMap, against the prior results obtained with NMI (InfoMap being more accurate than Louvain method). To contribute to this analysis and reveal further the differences between cNMI and other NMI-type metrics, we generated a series of

synthetic networks by LFR model [5], with different network sizes. The average degrees of the networks are all set to 8, and the maximal degrees to 50, with mixing parameter μ=0.45. The community sizes range from 20 to 400 to obtain more heterogeneous communities and other parameters are using default values. Figure 5 shows the testing results of InfoMap and Louvain method under different metrics.

From Figure 5, all the metrics except rNMI demonstrate that InfoMap performs much better than Louvain method on these LFR networks, while under rNMI we may conclude that Louvain method works better than InfoMap. In addition, although $rNMI_{norm}$ can give evaluation of InfoMap and Louvain method similar to NMI and cNMI (see the red and blue curves in Figure 5(c)), we found that $rNMI_{norm}$ would lead to the opposite conclusion (see the black and green curves in Figure 5(c)). Since $rNMI_{norm}$ is not a symmetric measure, when reversing the order of computed and known partitions in equation (6) for computing $rNMI_{norm}$, the results are different. If partitions of InfoMap and Louvain method are used as the first input partition, the conclusion is that InfoMap performs much better than Louvain method, while with the second input partitions the conclusion is reversed. As a result, $rNMI_{norm}$ shows important limitations as a metric for performance testing of community detection.

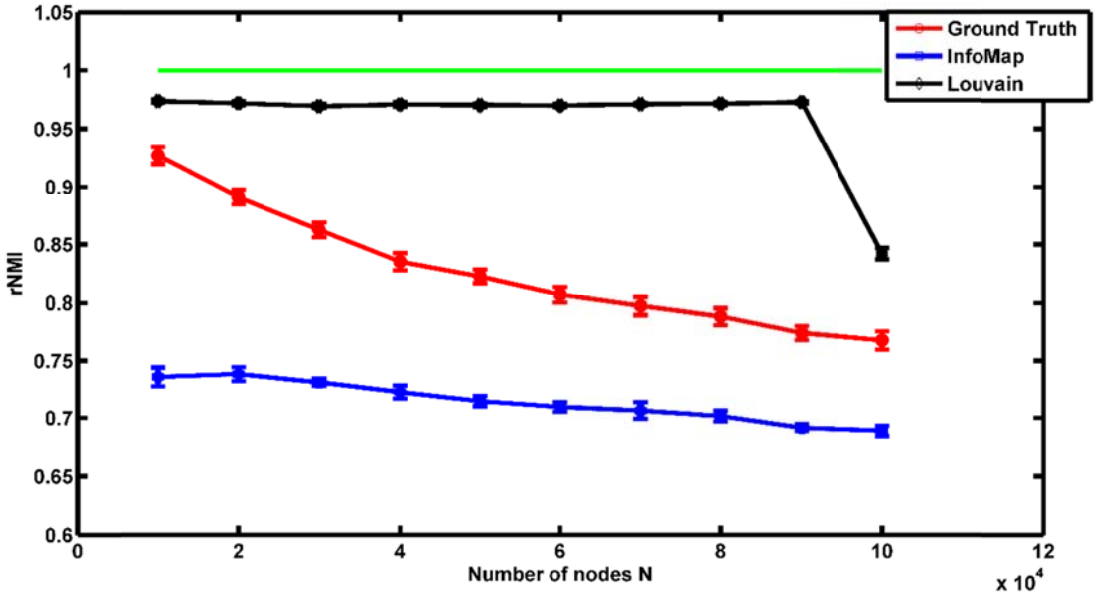

Figure 6. The values of rNMI between identical partitions from different methods. The green straight line is for the expected values of similarity between two identical partitions. Each point in the figure is averaged on 10 network instances.

To disambiguate the situations we computed rNMI values of the identical partitions obtained by InfoMap, Louvain method and *ground truth*. The ground truth partitions are the real community structure of the generated networks. The results are shown in Figure 6. As it can be seen from Figure 6, the values of rNMI on identical partitions obtained by Louvain method are very large and stable across different network sizes, while those for partitions obtained by InfoMap and ground truth decline as the network size grows, with the ones for InfoMap being the smallest. When the network size grows, the number of communities in the ground truth

partition grows as well. Due to the reverse finite size effect, rNMI biases to the partition with smaller number of communities. Since the partitions obtained by Louvain method usually contain smaller number of communities, the values of rNMI are thus much larger than that of partitions obtained from ground truth and InfoMap, leading to the erroneous conclusion that Louvain method works better than InfoMap [19]. In contrast, cNMI not only removes the finite size effect of NMI but also the reverse finite size effect of rNMI, which, based on the self-similarity experiment, leads to the evidence that InfoMap works better than Louvain method.

V. The closed form expression of cNMI

In the previous sections, both rNMI and cNMI adopt a correction to remove the bias from NMI by using the partitions sampled from a null model that has the same number of communities and nodes as the computed partition. As a consequence, the expectation of NMI between the ground-truth partition and a random partition is computed by averaging on realizations of the null model. In fact, there are closed form expressions for both rNMI and cNMI. In this section, we discuss such possible closed form expressions of cNMI.

If one assumes that the distribution of the variable of the number of nodes n_r. in any community r follows the Bernoulli distribution, the closed form expression of the expectation of NMI can be obtained from equation (3), as in [19]. This is of interest, as an explicit closed form expression for the expectation of NMI may make the implementation more efficient.

One possible type of closed form expression of cNMI based on the approximation of equation (3), denoted as $cNMI_B$, is defined as:

$$cNMI_B(X,Y) = \frac{2NMI(X,Y) - \frac{1}{N}\frac{C_X C_Y - C_X - C_Y + 1}{H(X)+H(Y)}}{2 - \frac{1}{2N}\frac{C_X^2 - 2C_X + 1}{2H(X)} - \frac{1}{2N}\frac{C_Y^2 - 2C_Y + 1}{2H(Y)}} \quad (9)$$

However, as pointed out by Zhang [19] (see Figure 2 therein), adopting the Bernoulli distribution makes the value of the expectation of NMI less accurate than the simulation value. As it can be seen from Figure 7(a), $cNMI_B$ gives poorer evaluation on results obtained from Louvain method when the number of nodes is larger. In fact, we find that equation (3) is an even worse estimation of the expectation of NMI. When the number of communities in a network is large enough, say C_X and C_Y are larger than $\sqrt{2N}$, equation (3) is dominated by the term $C_X C_Y$, making the estimated expectation of NMI much larger than 1. This unrealistic estimation makes the value of rNMI negative and that of $cNMI_B$ larger than 1, which is why on results from Louvain method $cNMI_B$ is increasing with the increasing number of nodes (also increasing the number of communities, see Figure 7(a)) and is even much larger than 1 on the results from InfoMap (data not shown).

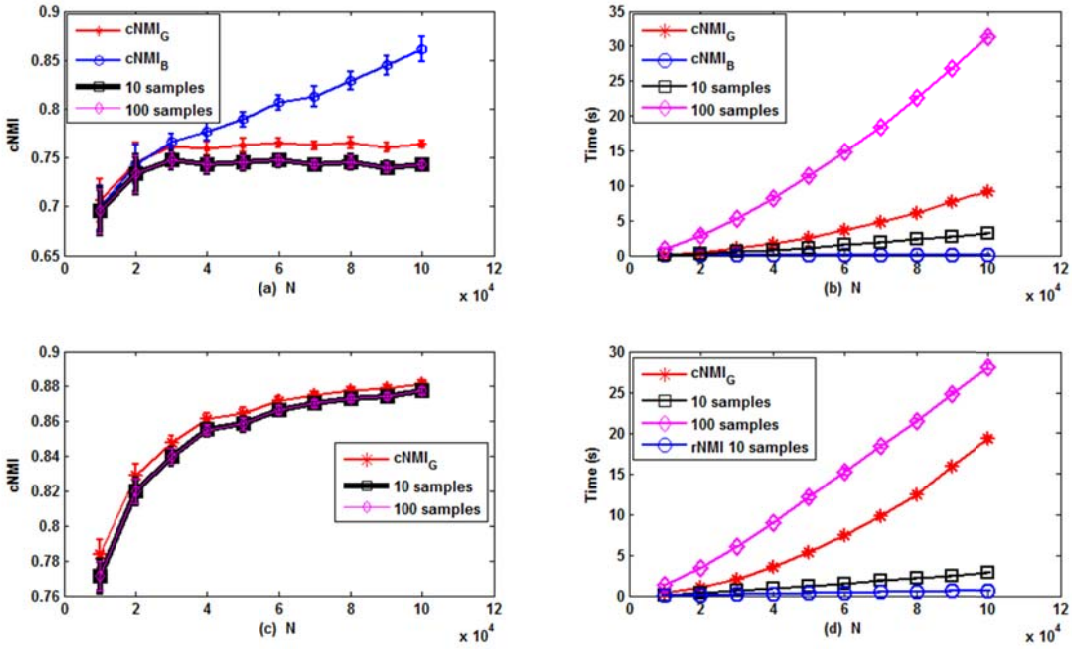

Figure 7. cNMI and the computational time on community partition results, computed by different strategies. (a) cNMI values of community partition results obtained by Louvain method. (b) The computational time of computing cNMI by different strategies on results obtained by Louvain method. (c) cNMI values of community partition results obtained by InfoMap. (d) The computational time of computing cNMI by different strategies on results obtained by InfoMap, together with the computational time of rNMI for comparison. Each point in the figure is averaged on 10 different networks. The computational time was obtained by the implementation of C++ under Linux on DELL PC with Dural 3.4GHz CPU cores and 8GB memory.

Differently from setting the randomness on the number of nodes in a community, Hubert and Arabie modeled the randomness on the variable n_{rs} of the number of nodes that were common to community r and s [11]. They assumed that the variable n_{rs} follows the generalized hypergeometric distribution that the number of communities and nodes therein are the same of the computed partition, i.e. the contingency table is constructed from the same distribution. We here employ this generalized hypergeometric distribution to derive an explicit expression for the expectation of NMI.

As previously discussed, to compute the expectation $< \mathrm{NMI}(\mathrm{X}, \mathrm{Z}_Y) >$, the partition X is fixed while the partition Z_Y is generated from the generalized hypergeometric model with $n_{\cdot l}(1 \le l \le C_Y)$ fixed. Consequently, this procedure introduces the randomness in the contingency table between X and Z_Y, making n_{rs} a random variable. Therefore, the number of different ways to create contingency tables that makes n_{rs} nodes common to community X_r and Z_s reads:

$$\emptyset(\mathrm{n}_{rs}) = \binom{\mathrm{n}_{r\cdot}}{\mathrm{n}_{rs}} \binom{N-n_{r\cdot}}{n_{\cdot s}-n_{rs}} \cdot \prod_{s'=1, s' \neq s}^{C_Y} \binom{N-n_{\cdot s}-\sum_{l=1}^{s'-1} n_{\cdot l}}{n_{\cdot s'}}. \quad (10)$$

On the other hand, the total number of different ways to create contingency tables between partitions X and Z_Y without constraint on any entry of the table is:

$$\emptyset(\cdot) = \prod_{s=1}^{C_Y} \binom{N-\sum_{l=1}^{s-1} n_{\cdot l}}{n_{\cdot s}}. \quad (11)$$

As a result, the distribution of random variable n_{rs} can be obtained as follows:

$$\mathrm{P}(\mathrm{n}_{rs}) = \frac{\emptyset(\mathrm{n}_{rs})}{\emptyset(\cdot)} = \binom{\mathrm{n}_{r\cdot}}{\mathrm{n}_{rs}}\binom{N-n_{r\cdot}}{n_{\cdot s}-n_{rs}}/\binom{N}{n_{\cdot s}}. \quad (12)$$

To make sense the equation for $\mathrm{P}(\mathrm{n}_{rs})$, conditions that $\mathrm{n}_{r\cdot} \geq \mathrm{n}_{rs}$ and $N - n_{r\cdot} \geq n_{\cdot s} - n_{rs}$, as well as $\mathrm{n}_{rs} \leq n_{\cdot s}$, should be satisfied for $1 \leq r \leq C_X, 1 \leq s \leq C_Y$, leading to the constraint that $\max(0, n_{r\cdot} + n_{\cdot s} - N) \leq \mathrm{n}_{rs} \leq \min(n_{r\cdot}, n_{\cdot s})$. With this distribution, the expectation of NMI can then be computed:

$$< \mathrm{NMI}(\mathrm{X}, \mathrm{Z}_Y) > = < \frac{-2\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y}\frac{n_{rs}}{N}\log\left(\frac{\frac{n_{rs}}{N}}{\frac{n_{r\cdot}}{N}\cdot\frac{n_{\cdot s}}{N}}\right)}{\sum_{r=1}^{C_X}\frac{n_{r\cdot}}{N}\log\left(\frac{n_{r\cdot}}{N}\right)+\sum_{s=1}^{C_Y}\frac{n_{\cdot s}}{N}\log\left(\frac{n_{\cdot s}}{N}\right)} >$$

$$= \frac{-2}{H(X)+H(\mathrm{Z}_Y)} < \sum_{r=1}^{C_X}\sum_{s=1}^{C_Y}\frac{n_{rs}}{N}\log\left(\frac{\frac{n_{rs}}{N}}{\frac{n_{r\cdot}}{N}\cdot\frac{n_{\cdot s}}{N}}\right) >$$

$$= \frac{-2}{H(X)+H(Y)}\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y}[< \frac{n_{rs}}{N}\log\left(\frac{n_{rs}}{N}\right) > - \log(\frac{n_{r\cdot}}{N}\cdot\frac{n_{\cdot s}}{N}) < \frac{n_{rs}}{N} >] \quad (13)$$

The last equation holds under the assumption that all entries n_{rs} of the contingency table are *independent and identically distributed* (iid) variables. The expectation terms in the above equation are given by:

$$< \frac{n_{rs}}{N}\log\left(\frac{n_{rs}}{N}\right) > = \sum_{n_{rs}=\max(0,\, n_{r\cdot}+n_{\cdot s}-N)}^{\min(n_{r\cdot},\, n_{\cdot s})}\frac{n_{rs}}{N}\log\left(\frac{n_{rs}}{N}\right)\cdot P(n_{rs}),$$

$$< \frac{n_{rs}}{N} > = \frac{1}{N} < n_{rs} > = \frac{1}{N}\frac{n_{r\cdot}n_{\cdot s}}{N}.$$

Consequently, the closed form of the expectation of NMI under the generalized hypergeometric model can be obtained:

$< \mathrm{NMI}(\mathrm{X}, \mathrm{Z}_Y) > =$

$$\frac{-2}{H(X)+H(Y)}\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y}\sum_{n_{rs}=\max(0,\, n_{r\cdot}+n_{\cdot s}-N)}^{\min(n_{r\cdot},\, n_{\cdot s})}\frac{n_{rs}}{N}\log\left(\frac{n_{rs}}{N}\right)\cdot\binom{\mathrm{n}_{r\cdot}}{\mathrm{n}_{rs}}\binom{N-n_{r\cdot}}{n_{\cdot s}-n_{rs}}/\binom{N}{n_{\cdot s}} -$$

$$\frac{-2}{H(X)+H(Y)}\sum_{r=1}^{C_X}\sum_{s=1}^{C_Y}\frac{1}{N}\frac{n_{r\cdot}n_{\cdot s}}{N} \quad (14)$$

To simplify the denotations, let $N_X = (n_{r\cdot}, \cdots, n_{C_X\cdot})$ and $N_Y = (n_{r\cdot}, \cdots, n_{C_Y\cdot})$ and denote $< \mathrm{NMI}(\mathrm{X}, \mathrm{Z}_Y) > \triangleq < \mathrm{NMI}(\mathrm{X}, \mathrm{Z}_Y; N_X, N_Y) >$ to make explicit the dependence of the expectation of NMI on the community sizes, the closed form expression of cNMI between partitions X and Y under the generalized hypergeometric model (denoted as $cNMI_G(X, Y)$) reads:

$$cNMI_G(X, Y) = \frac{2NMI(X,Y) - 2<NMI(X, Z_Y; N_X, N_Y)>}{2 - <NMI(X, Z_X; N_X, N_X)> - <NMI(Y, Z_Y; N_Y, N_Y)>}. \quad (15)$$

Here we make use of the fact that $< NMI(X, Z_Y; N_X, N_Y) >$ is equal to $< NMI(Y, Z_X; N_Y, N_X) >$ according to equation (12).

Figure 7 shows cNMI values computed by different strategies on community partition results obtained by Louvain method and InfoMap. Although the computational time for $cNMI_B$ is constant (Figure 7(b)), $cNMI_B$ cannot correctly evaluate partition results, as discussed previously (Figure 7(a)) and we thus exclude

the curve of $cNMI_B$ in Figure 7(c) for results by InfoMap for its inaccurate values larger than 1. As an alternative, cNMI computed by averaging on different samples from null model gives an accurate empirical estimation and is easy to compute. In practice, such estimation makes the estimated cNMI converge fast to its true value. As it can be seen from the Figure 7, cNMI computed by averaging on 10 samples has almost the same curve as by averaging on 100 samples, but the former obviously takes far less computational time. Figure 7 (a) and (c) show that the closed form expression $cNMI_G$ also gives good evaluations on the community partition results, with the computation time lying between averaging on 10 and 100 samples. $cNMI_G$ gives mildly higher cNMI scores across networks of different scales than that of the empirical calculations by averaging on samples, however such differences are negligible, as shown in Figure 8. The small overestimation is the result of the violation of the iid assumption in the calculation of the expectation in equation (13), since n_{rs} is not strictly independent under the constraints that $n_{.s}$ is fixed for $1 \leq s \leq C_Y$. The effect of such violation is alleviated for growing values C_Y of communities, which is exemplified by the fact that $cNMI_G$ gives more accurate values by InfoMap on larger networks when compared to those by Louvain method.

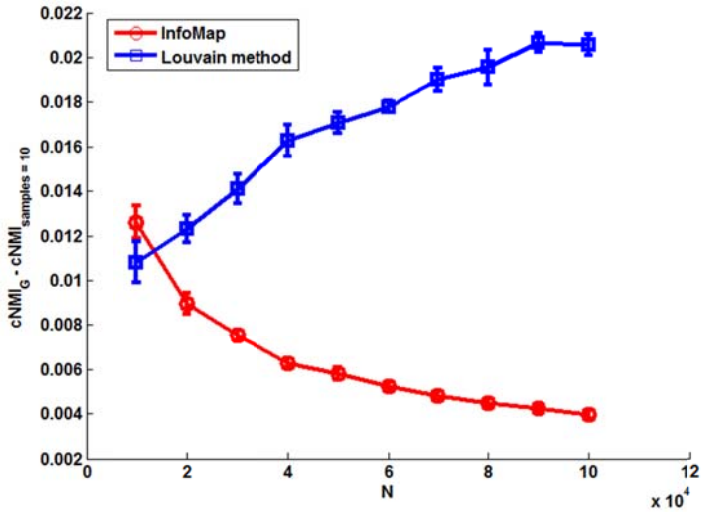

Figure 8. The difference between cNMI in closed form $cNMI_G$ and cNMI computed by averaging on 10 samples of the null model $cNMI_{sample=10}$, of the results obtained by InfoMap and Louvain method. Each point in the figure is averaged on 10 different networks.

In addition to Figure 7 (b) and (d) showing that the computation of cNMI is efficient enough, complexity analysis can also be exploited. To compute $\mathrm{NMI}(X, Y)$, one needs first construct a contingency table between partitions X and Y and then compute NMI according to equation (1), taking respectively $\mathrm{O}(\mathrm{N})$ and $\mathrm{O}(\mathrm{C}_X\mathrm{C}_Y + \mathrm{C}_X + \mathrm{C}_Y)$. Thus, it takes $\mathrm{O}(\mathrm{N} + \mathrm{C}_X\mathrm{C}_Y)$ time to compute NMI. Computing the expectation $\mathrm{NMI}(X, Z_Y)$ by averaging on samples from null model requires first to permute partition Y to obtain Z_Y, which takes $\mathrm{O}(\mathrm{N})$. As a consequence, the time complexity of computing $\mathrm{rNMI}(X, Y)$ is $\mathrm{O}(\mathrm{I}(\mathrm{N} + \mathrm{C}_X\mathrm{C}_Y))$ and $\mathrm{cNMI}(X, Y)$ is $\mathrm{O}(\mathrm{I}(\mathrm{N} + \mathrm{C}_X^2 + \mathrm{C}_Y^2 + 2\mathrm{C}_X\mathrm{C}_Y))$, where I is the number of samples. On the other hand, it takes at most $\mathrm{O}(\mathrm{n}_{r.} + \mathrm{n}_{.s})$ time to compute the term

$\sum_{n_{rs}=\max(0,\, n_{r\cdot}+n_{\cdot s}-N)}^{\min(n_{r\cdot},\, n_{\cdot s})} \frac{n_{rs}}{N} \log\left(\frac{n_{rs}}{N}\right) \binom{n_{r\cdot}}{n_{rs}} \binom{N-n_{r\cdot}}{n_{\cdot s}-n_{rs}} / \binom{N}{n_{\cdot s}}$, and thus $O((C_X + C_Y)N)$ to compute the expectation $< \mathrm{NMI}(X, Z_Y) >$ and $cNMI_G$.

VI. Applications beyond community detection

Community detection in networks consists of clustering nodes into groups, which in essence is a special case of the general clustering problem [1, 24]. As an important and improved measure for performance evaluation, cNMI is by no means only limited to community detection field, but can be applied to a wider context for evaluating general clustering results.

In this section, we apply cNMI to evaluate clustering results on real-world datasets as COIL-20 [25] and COIL-100 [26]. COIL-20 contains 20 objects, each of which has 72 images that are represented as 1024-dimensional vectors, while COIL-100 contains similarly represented objects but for 100 objects and thus 7200 instances. Figure 9 shows the results on these two datasets by a recently designed clustering method SCC (sparse concept coding) [27] and the classical K-means method [28], evaluated by rNMI and cNMI.

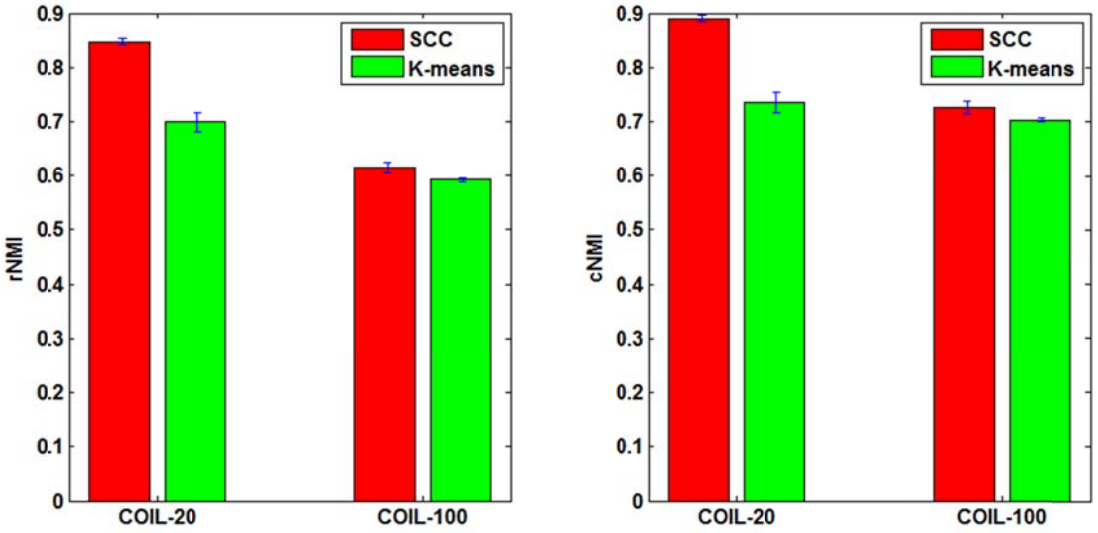

Figure 9. The values of rNMI (a) and cNMI (b) of the clustering results obtained by SCC and K-means on datasets COIL-20 and COIL-100. Each bar in the figure is obtained by averaging on 20 repetitions of algorithms.

SCC is a two-step procedure that first learns *K* bases and then derives sparse representations of objects under these bases for *K*-means clustering. We used default parameters' values to apply SCC on COIL-20 and compared its performance with that of *K*-means. Both rNMI and cNMI reveal that SCC performs better than *K*-means on COIL-20, as previously shown [27]. When applied to COIL-100, SCC is also shown to be a better method on this scaling up dataset with more diverse objects under both rNMI and cNMI. However, more subtle differences in this context can be learned from the figure if we are concerned with the performance stability of SCC and *K*-means methods. Under cNMI, the performance of *K*-means remains good when shifting from the application on a smaller dataset COIL-20 to the application on

another larger dataset COIL-100, showing good stability of the performance of *K*-means. In contrast, if one uses instead rNMI to evaluate the clustering results, the performance of *K*-means drops much faster (4% vs. 15%) compared to cNMI when the variety of objects enriches, which may mislead to the inaccurate conclusion that the stability of the performance of *K*-means on these datasets is not good. Similarly, cNMI also reveals that SCC has higher stability than when using rNMI, since under cNMI the performance drops only about 18% instead of 27% under rNMI. All these phenomena are due to the reverse finite size effect of rNMI that cannot give the score 1 when two partitions are identical. In fact, rNMI scores the ground truth partition against itself 0.9509 on COIL-20 and 0.8355 on COIL-100.

VII. Conclusions

In this paper, we theoretically and experimentally show a relevant problem of a recently proposed metric rNMI for performance testing of community detection methods. Although rNMI is originally designed to remove the finite size effect of NMI, it suffers from a different limitation, the reverse finite size effect that biases to a partition with small number of communities and small community size. If the number of communities is large enough or the size of the community is small enough, the value of rNMI, even between two identical partitions may shrink to a very small or even vanishing value, and not to the intuitive value of 1 that the similarity of two identical partitions should have. We thus proposed a new metric cNMI on the basis of the definition of rNMI. cNMI scales from giving value 0 to two random partitions to giving value 1 to two identical partitions, and thus is a well normalized metric. cNMI is also a symmetric metric. It not only removes the finite size effect of NMI but also the reverse finite size effect of rNMI. Compared to rNMI that in some cases may give a wrong estimate of performance of community detection methods, cNMI can more objectively test different community detection methods. We thus suggest to use cNMI for performance testing of community detection methods.

We further discuss different strategies to compute cNMI, that is, by averaging on instances sampled from the hypothesized null models and by closed form approximations. In the closed form approximations of cNMI, we find that the approximation by setting Bernoulli distribution of the number of nodes in a community is poor for both rNMI and cNMI. If the randomness of the number of common nodes between two communities is modeled to follow the generalized hypergeometric distribution, the resulting closed form expression of cNMI becomes a good approximation. Since the calculation of such closed form cNMI introduces the iid assumption on the random variable of the number of common nodes, its value is slightly, but negligibly, overestimated. Compared to the closed form expressions of cNMI, cNMI by averaging on limited samples can be computed faster and more accurately. It is thus preferable to use cNMI by averaging on samples for the performance testing of community detection methods and other clustering methods in the much wider clustering field as exemplified in the previous section.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grants No. 61202262), the Natural Science Foundation of Jiangsu Province (Grants No. BK2012328), and Specialized Research Fund for the Doctoral Program of Higher Education of China (Grants No. 20120092120034).

References

[1]. S. Fortunato, Community detection in graphs, (2010) Phys. Rep. 486 75.

[2]. K. Nowicki, and T. A. B. Snijders, Estimation and prediction for stochastic blockstructures, (2001) J. Am. Stat. Assoc 96, p.1077-1087.

[3]. A. Decelle, F. Krzakala, C. Moore, Cristopher and L. Zdeborova, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, (2011) Phys. Rev. E 84, 066106.

[4]. M. B.Hastings, Community detection as an inference problem, (2006) Phys. Rev. E 74, 035102.

[5]. A. Lancichinetti, S. Fortunato and F. Radicchi, Benchmark graphs for testing community detection algorithms, (2008) Phys. Rev. E 78, 046110.

[6]. W. W. Zachary, An information flow model for conflict and fission in small groups, (1977) Journal of Anthropological Research 33, p.452-473.

[7]. M. Girvan and M. E. J. Newman, Community structure in social and biological networks, (2002) Proc. Natl. Acad. Sci. USA 99, p.7821-7826.

[8]. L. A. Adamic and N. Glance, The political blogosphere and the 2004 US Election, (2005) in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem.

[9]. W.M. Rand, Objective criteria for the evaluation of clustering methods, (1971) J. Am. Assoc. 66 (336), p.846-850.

[10]. A. Ben-Hur, A. Elisseeff and I. Guyon, A stability based method for discovering structure in clustered data, (2002) in: Pacific Symposium on Biocomputing, p.6–17.

[11]. L. Hubert, P. Arabie, Comparing partitions, (1985) J. Classification 2, p.193–218.

[12]. L. Danon, A. Díaz-Guilera, J. Duch and A. Arenas, Comparing community structure identification, (2005) J. Stat. Mech. P09008.

[13]. M. Meila, Comparing clusterings—an information based distance, (2007) Journal of multivariate analysis 98, p.873-895.

[14]. M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, (2004) Phys. Rev. E 69, 026113.

[15]. J. Leskovec, K.J. Lang, A. Dasgupta and M.W. Mahoney, Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, (2009) Internet Mathematics. 6 (1), p.29-123.

[16]. G. Orman, V. Labatut and H. Cherifi, Comparative evaluation of community detection algorithms: a topological approach, (2012) J. Stat. Mech. P08001.

[17]. H Papadakis, C Panagiotakis and P Fragopoulou, Distributed detection of communities in complex networks using synthetic coordinates, (2014) J. Stat. Mech. P03013.

[18]. Diego R Amancio, Osvaldo N Oliveira Jr and L da F Costa, Robustness of community structure to node removal, (2015) J. Stat. Mech. P03003.

[19]. P. Zhang, Evaluating accuracy of community detection using the relative normalized mutual information, (2015) J. Stat. Mech. P11006.

[20]. J. Zhang, T. Chen, and J. Hu, On the relationship between Gaussian stochastic blockmodels and label propagation algorithms, (2015) J. Stat. Mech. P03009.

[21]. P. Zhang, and C. Moore, Scalable detection of statistically significant communities and hierarchies, using message passing for modularity, (2014) Proc. Natl. Acad. Sci. USA 111, p.18144-18149.

[22]. M. Rosvall and C. T. Bergstrom, Maps of random walks on complex networks reveal community structure, (2008) Proc. Natl. Acad. Sci. USA 105, p.1118-1123.

[23]. V. D. Blondel,J-L Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks, (2008) J. Stat. Mech. P10008.

[24]. S. Schaeffer, Graph clustering, (2007) Computer Science Review, 1 (1), p.27-64.

[25]. S. A. Nene, S. K. Nayar, H. Murase, Columbia Object Image Library (COIL-20),Technical Report CUCS-005-96, February 1996. URL http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php.

[26]. S. A. Nene, S. K. Nayar, H. Murase, Columbia Object Image Library (COIL-100),Technical Report CUCS-006-96, February 1996. URL http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php.

[27]. Deng Cai, Hujun Bao and Xiaofei He, Sparse Concept Coding for Visual Analysis, (2011) Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Washington, DC, USA, p.2905-2910.

[28]. J.B. MacQueen, Some methods for classification and analysis of multivariate observations, in: L.M.L. Cam, J. Neyman (Eds.),(1967) Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, USA, p.281-297.