Modeling with Node Degree Preservation Can Accurately Find Communities

Di Jin[1], Zheng Chen[2], Dongxiao He[1], Weixiong Zhang[2,3]

[1]School of Computer Science and Technology, Tianjin University, Tianjin 300072, China, [2]Department of Computer Science and Engineering, Washington University, St. Louis, MO 63130, USA, [3]Institute for Systems Biology, Jianghan University, Wuhan 430056, China
{jindi, hedongxiao}@tju.edu.cn, {zheng.chen, weixiong.zhang}@wustl.edu

Abstract

An important problem in analyzing complex networks is discovery of modular or community structures embedded in the networks. Although being promising for identifying network communities, the popular stochastic models often do not preserve node degrees, thus reducing their representation power and applicability to real-world networks. Here we address this critical problem. Instead of using a blockmodel, we adopted a random-graph null model to faithfully capture community structures by preserving in the model the expected node degrees. The new model, learned using nonnegative matrix factorization, is more accurate and robust in representing community structures than the existing methods. Our results from extensive experiments on synthetic benchmarks and real-world networks show the superior performance of the new method over the existing methods in detecting both disjoint and overlapping communities.

1. Introduction

Most networks, such as social and biological networks, are better organized and represented in communities or modules, where nodes within a community are relatively densely connected and nodes across communities are sparsely linked (Girvan and Newman 2002). Due to its utility in unraveling complex structures in networks from diverse fields, ranging from social sciences, engineering to biology and medicine, identification of communities in complex networks has attracted much attention in recent years (Fortunato 2010). Much effort has been devoted to developing methods for community detection. These methods can be grouped into that for finding disjoint communities (Girvan and Newman 2002; Blondel et al. 2008; Rosvall and Bergstrom 2008) and those for finding overlapping communities (Palla et al. 2005; Lancichinetti, Fortunato and Kertesz 2009; Ahn, Bagrow and Lehmann 2010).

Among the existing methods for community discovery are stochastic models of communities, which offer an effective technique for network analysis and has attracted much attention lately (Newman 2012). Several model-based methods have been proposed (Wang et al. 2011; Psorakis et al. 2011; Zhang and Yeung 2012; Ren et al. 2009; Shen, Cheng and Guo 2011; Karrer and Newman 2011; Ball, Karrer and Newman 2011), most of which are based on the popular stochastic blockmodel (Nowicki and Snijders 2001). In a simple form of the stochastic blockmodel, each of n nodes of a network is assigned to one of c communities, and two nodes are connected with a probability depending on the community memberships of the nodes. To be concrete, the stochastic blockmodel of a network of n nodes can be represented by a $c \times c$ probability matrix $\omega = (\omega_{g_i g_j})$, where g_i is the community that node i belongs to and $\omega_{g_i g_j}$ is the probability that an edge is introduced between nodes i and j.

Although simple, this blockmodel ignores variation of node degrees and thus does not preserve node degrees of a network to be modeled, rendering the model unsuitable for real networks (Karrer and Newman 2011). Node degrees of a real-world network are typically distinct from that of a random graph; node degrees of a real-world network often exhibit a power-law distribution (Barabasi and Albert 1999). Neglect or improper formulation of node degrees can result in inaccurate network structures and thereby affecting the quality of a community discovery algorithm.

The idea of preserving node degrees was proposed by (Karrer and Newman 2011). However, their model itself does not preserve node degrees; this is achieved by additional inference, which limits its generality. This critical issue has not been addressed adequately in the existing methods that adopt or extend the blockmodel, as discussed in Sec. 5. Here, we attempted to address this issue. In our study, we did not use the blockmodel. Rather, we introduced a new stochastic model atop a popular random-graph null model, i.e., null model of modularity (Newman and Girvan 2004). Using probabilistic community membership, our new model can not only describe those community structures that the original null model failed to capture, but also hold the property of retaining the expected node degrees in the model to be the same as in the original network. The community structure was learned by fitting the model to the given network. Our model preserves node degrees by itself without inference, and hence it doesn't have to rely

on specialized inference algorithms for this task. As a result, the new model and its companion network community discovery method are effective in discovering intrinsic and subtle community structures of many real-world networks.

This paper is organized as follows. We first show the importance of preserving node degrees by an example. We then introduce the new model and discuss the new MNDP method. We present experimental results thereafter. Finally, we discuss the related works and conclude this work.

2. A Motivating Example

We examine here the well-known "karate club" (Zachary 1977) to show the importance of preserving node degrees. This network has been considered as the *de facto* model for network analysis. It represents the relationships among the 34 members of a karate club. The club is known to have split into two groups or communities as a result of a dispute, and the members of each group are known.

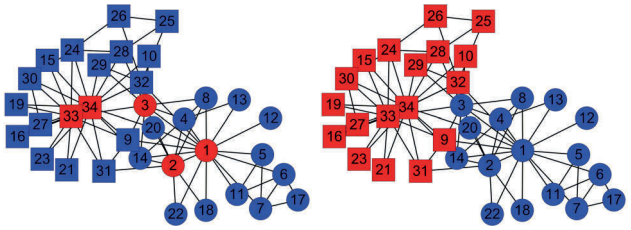

Figure 1: The Zachary's "karate club" network. The two node shapes represent the two actual communities reported: the club administrator's (square) and the instructor's (circle). **(a)** The two communities, in red and blue, reported by the basic blockmodel. **(b)** The two communities, also in red and blue, correctly identified by our model that is able to preserve node degrees.

The basic blockmodel (Nowicki and Snijders 2001) found two communities (Fig. 1(a)), which are drastically different from the two actual groups of the club. The main problem of this model, as correctly pointed out by others (Karrer and Newman 2011), is that it fails to preserve node degrees so that it erroneously splits the club members into a group with (red) nodes of high degrees and another with (blue) nodes of low degrees.

On the contrary, our model correctly split the club members into two groups that perfectly match the actual network communities (Fig. 1(b)). The success of our model on this well-known network, as well as on other networks to be discussed in later sections, is mainly from faithfully modeling heterogeneous node degrees in the model (see Fig. 2 and Table 1). Besides, our result here is also better than that of the Karrer's degree-corrected blockmodel (Fig. 1(b) in (Karrer and Newman 2011)). The reason may be that: we adopt the random-graph null model while Karrer used the stochastic blockmodel; our model itself preserves node degrees while Karrer's model achieved this by specialized inference (see Sec. 5 for discussion).

3. The Method

We first describe the new model, discuss its properties, and then introduce an algorithm to learn the model parameters.

3.1 Stochastic Model

Consider an undirected graph of n nodes, $G = (V, E)$, represented by an adjacency matrix A. Assume that the nodes can be partitioned into c communities using a probabilistic group membership variable S, where S_{iz} is the probability that node i belongs to community z, subject to $\sum_z S_{iz} = 1$.

Given S_{iz}, network G can be viewed as an ensemble of c probabilistic communities, $\{C_1, C_2, \ldots, C_c\}$, where every node has a probabilistic membership in each community C_z. The degree of node i within community C_z can be defined as $d_{iz} = d_i S_{iz}$, where d_i is the degree of node i in G. Since C_z corresponds to only one community, it can be regarded as a random graph with no community structure. A popular random-graph null model, namely null model of modularity (Newman and Girvan 2004), is suitable for characterizing a probabilistic community. This null model describes random graphs with no communities where edges are rewired randomly among the nodes. We used the null model of modularity to describe each C_z with the given node degrees $\{d_{1z}, d_{2z},\ldots,d_{nz}\}$, so that we preserved node degrees in each of the probabilistic communities. In this null model, the expected number of links (or expected link weight) between nodes i and j in C_z can be written as

$$\hat{w}_{ij}^z = d_{iz} d_{jz} / \sum_k d_{kz} \ . \qquad (1)$$

Taking all the communities $\{C_z|z=1,2,..,c\}$ into consideration, the expected number of links between nodes i and j in the model can be written as

$$\hat{w}_{ij} = \sum_z \hat{w}_{ij}^z = \sum_z \left(d_{iz} d_{jz} / \sum_k d_{kz} \right). \qquad (2)$$

However, the community membership S_{iz} is unknown in advance, and it needs to be inferred by fitting the model parameters d_{iz} to the given network, which will be discussed in Sec. 3.3. When the parameters d_{iz} are available, the community membership S_{iz} can then be inferred by

$$S_{iz} = d_{iz} / \sum_r d_{ir} = d_{iz} / d_i \ . \qquad (3)$$

Both hard and overlapping communities can be derived from the probabilistic community membership, S_{iz}. Specifically, to derive a hard partition, node i can be assigned to community $r = arg\max_z \{S_{iz}, z=1,2,\ldots,c\}$. To construct a structure with overlapping communities, the strategy of (Zhang and Yeung 2012) can be adopted, i.e., the entries in vector $\{S_{i1}, S_{i2},\ldots,S_{ic}\}$ are first scaled to values in [0, 1], and those entries whose rescaled values exceed a threshold are set to 1, or 0 otherwise. Then we can use a suitable quality metrics of communities to determine the threshold.

3.2 Model Properties

The new model consists of a set of probabilistic communities, each of which is considered to be a random graph and represented by a null model of modularity. The new model

can in essence be regarded as a generalization to the null model of modularity. It not only incorporates the ability of describing communities which the original null model does not have, but also holds the property, as with the original model, that fixes the expected node degrees to be the same as that of the observed network, as captured by Proposition 1 below. Our model naturally describes the distributions of heterogeneous node degrees, making it more robust for complex network community structures than the blockmodels, which do not have this property.

Proposition 1. *In our model, the expected graph of G preserves the same node degrees as in the observed network G (see Appendix for proof).*

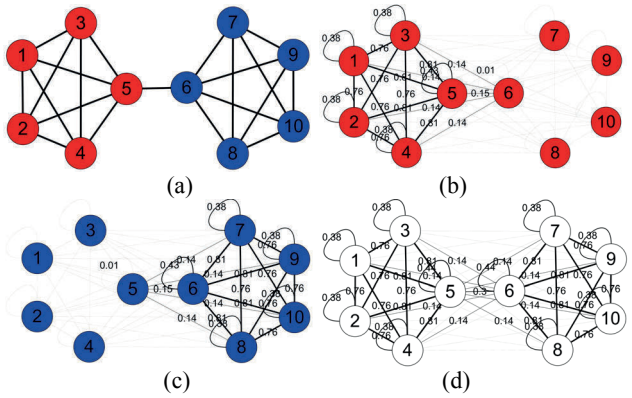

Figure 2: An example illustrating some properties of the model with network parameters d_{iz} listed in Table 1. **(a)** A given network G with two communities shown in red and blue. **(b)** and **(c)** The expected graphs of the red and blue communities, each of which is described by a null model of modularity. **(d)** The expected graph of G that combines the expected graphs of the two communities. The width of a link corresponds to its expected values; the links with values smaller than 1.0e-2 are omitted.

Table 1: The parameters d_{iz} used in Fig 2 and learned by NMF

d_{iz}	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$z=1$	3.999949	3.999974	3.999975	3.999972	4.253827
$z=2$	5.12E-05	2.63E-05	2.45E-05	2.76E-05	0.746173
d_{iz}	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$
$z=1$	0.748089	7.14E-04	7.39E-04	7.02E-04	6.89E-04
$z=2$	4.251911	3.999286	3.999261	3.999298	3.999311

Here we offer an example, shown in Fig. 2 and in Table 1, to illustrate our model of a small network. The red and blue communities are indexed by 'z=1' and 'z=2' in Table 1, respectively. Firstly, consider the within-community nodes, e.g., nodes 3 and 4. As d_{31} and d_{41} are both large, the expected number of red links between them is large ($\hat{w}^1_{3,4}$=0.76), so that the expected number of links between nodes 3 and 4 is large as well ($\hat{w}_{3,4}$ = 0.76). On the other hand, consider the nodes in different communities, such as nodes 3 and 7. As d_{71} is very small, although d_{31} is large, the expected number of red links between these two nodes is still small ($\hat{w}^1_{3,7}$ <1.0e-2); similarly, the expected number of blue links between them is also small ($\hat{w}^2_{3,7}$ < 1.0e-2). Thus, the expected number of links between nodes 3 and 7 is much smaller than that between 3 and 4 ($\hat{w}_{3,7} \ll \hat{w}_{3,4}$). This result correctly matches the intuition that intra-cluster connectivities are dense than inter-cluster connectivities.

More importantly, by preserving node degrees our model is consistent with the observation that nodes with high degrees are, with all other things being equal, more likely to be connected among themselves than those with low degrees. In Fig. 1(d), for instance, as node 5 has a larger degree than node 3, $\hat{w}_{5,4} = 0.81 > \hat{w}_{3,4} = 0.76$. Intuitively, two nodes of high degrees are more likely to be connected than two nodes of low degrees. Thus, by incorporating such observations, we can faithfully describe heterogeneous node degrees and accurately find communities.

3.3 Parameter Learning

The model is specified by a set of parameters d_{iz} of node degrees in the communities, for n nodes i =1,2,…,n and c communities z=1,2,…,c. These parameters have to be learned from the data of the given network G. The problem of fitting the model to the data of G can be cast as the following optimization problem,

$$\min_{d_{iz}\geq 0} \left\| A - \hat{A} \right\|_F^2 = \sum_{ij} \left(w_{ij} - \sum_z \left(d_{iz} d_{jz} / \sum_k d_{kz} \right) \right)^2, \quad s.t. \ \sum_z d_{iz} = d_i \qquad (4)$$

where $\|.\|_F$ denotes Frobenius norm. The best fit between the expected graph with adjacency $\hat{A} = (\hat{w}_{ij})_{n\times n}$ and a given network G with adjacency $A = (w_{ij})_{n\times n}$ can be achieved by optimizing (4). We give a nonnegative matrix factorization (NMF) method to solve the optimization in (4).

We first introduce an auxiliary matrix X, where X_{iz} is

$$X_{iz} = d_{iz} / \sqrt{\sum_j d_{jz}} \ . \qquad (5)$$

Function (4) can be written as a constrained NMF problem,

$$\min_{X\geq 0} \left\| A - XX^{\mathrm{T}} \right\|_F^2, \quad st. \ XX^{\mathrm{T}} 1_n = d \ , \qquad (6)$$

where $d = (d_1, d_2, \ldots, d_n)^{\mathrm{T}}$. It is nontrivial to directly optimize (6) with the hard constraints. We relax this optimization problem by introducing a penalty term that represents the hard constraints into the objective function, arriving at minimizing the following objective function,

$$O(X) = \frac{1}{2} \left\| A - XX^T \right\|_F^2 + \frac{\lambda}{2} \left\| XX^T 1_n - d \right\|_2^2 \qquad (7)$$

where λ is a hyperparameter that reflects the importance of the hard constraints. Violation to more hard constraints incurs a higher penalty to the objective function. In our experiments, we first get an initial value of X_0 by setting $\lambda = 0$. Then we restart the optimization with $X = X_0$ and let λ to a relatively large number, e.g., 1000, to minimize the chance of violating the degree constraints. The purpose of the initialization is to restrict the search for a model to start from some good approximations. Similar to other forms of NMF, the objective function in (7) is not convex w.r.t. X, so that it is computationally intractable to find global minima. Therefore, the gradient descent method is

adopted to search for local minima, which can be implemented in a multiplicative updating algorithm similar to the method for SNMF (Wang et al. 2011). In order to derive the updating rule, a Lagrange multiplier matrix Θ for the nonnegative constraints on X is introduced to (7), resulting in the following equivalent objective function,

$$\mathcal{L}(X) = Tr(XX^TXX^T)/4 - Tr(A^TXX^T)/2 + \lambda Tr(1_n^TXX^TXX^T1_n)/2 - \lambda Tr(1_n^TXX^Td) + Tr(\Theta^TX)$$

For any stationary state, we have

$$\partial\mathcal{L}/\partial X = XX^TX - AX + \lambda 1_n 1_n^T XX^TX + \lambda XX^T 1_n 1_n^T X - \lambda 1_n d^T X + \lambda d 1_n^T X + \Theta$$

Using complementary slackness condition $(\Theta)_{ik}(X)_{ik} = 0$, we have the following equation,

$$(XX^TX - AX + \lambda 1_n 1_n^T XX^TX + \lambda XX^T 1_n 1_n^T X - \lambda 1_n d^T X + \lambda d 1_n^T X)_{ik}(X)_{ik} = 0.$$

This leads to the following update rule for X:

$$X_{ik} = X_{ik}\left(\frac{\left((A + \lambda 1_n d^T + \lambda d 1_n^T)X\right)_{ik}}{\left(XX^TX + \lambda 1_n 1_n^T XX^TX + \lambda XX^T 1_n 1_n^T X\right)_{ik}}\right)^{\frac{1}{4}} \quad (8)$$

When the update rule converges, shown in Theorem 1 below, the solution satisfies the Karush-Kuhn-Tucker (KKT) conditions (Boyd and Vandenberghe 2004).

Theorem 1. *Function O in (7) is non-increasing under the updating rule in (8). O is invariant under these updates if and only if X becomes stationary (see Appendix for proof).*

Finally, when the method converges, the parameters d_{iz} of the model can be computed using (5) as follows,

$$d_{iz} = X_{iz}\sqrt{\sum_j d_{jz}} = X_{iz}\sum_j X_{jz} \ . \quad (9)$$

Notice that, the time to calculate AX, $1_n(d^TX)$, $d(1_n^TX)$, $X(X^TX)$, $1_n(1_n^TXX^TX)$ and $X(X^T1_n)(1_n^TX)$ in (8) is $2mc$, $2nc$, $2nc$, $2nc^2$, $4nc$ and $4nc$, respectively, where n is the number of nodes, m is the number of edges and c is the number of clusters. So the time to evaluate (8) once is $O(mc+nc^2)$, and hence the time complexity of MNDP is $O(T(mc+nc^2))$, where T is the iteration number for convergence. As the number of communities is much smaller than network size (i.e., $c << m, n$), our method scales well to large networks.

4. Experiments

To test the performance of our MNDP, we evaluated it on synthetic and real networks. We also compared it with four related methods, i.e., Karrer's method (Karrer and Newman 2011), Ball's method (Ball, Karrer and Newman 2011), SNMF (Wang et al. 2011) and BNMTF (Zhang and Yeung 2012). Karrer's method considers node degrees, but it adopts likelihood probability as objective function which is optimized by local search. Ball's method was designed for link partitions (overlapping communities) and extended to node partitions with node degrees being taken into consideration. SNMF and BNMTF use squared loss objective functions and adopt NMF for optimization. Since the methods compared converge to local minima, we ran each method 20 times and reported the best result. We compared them in terms of hard partitions and overlapping communities. Besides, we compared them on a large weighted network to assess their applicability to weighted networks.

4.1 Synthetic Networks

Two types of synthetic benchmarks, one with disjoint communities (Lancichinetti, Fortunato and Radicchi 2008) and the other with overlapping communities (Lancichinetti and Fortunato 2009), were proposed. Here we considered them in our experiments. The node degrees and community size of a LFR benchmark graph follow power law distributions, a property that most real-world networks share. We did not compare the results of BNMTF here because it cannot finish in 100 hours on every trial attempted.

4.1.1 LFR Benchmark with Disjoint Communities

Proposed by (Lancichinetti, Fortunato and Radicchi 2008), this type of benchmark is designed to detect disjoint communities. We adopted the widely used Normalized Mutual Information (NMI) index as the quality metric in our study.

Following the parameter setting as the LFR benchmark used in (Lancichinetti, Fortunato and Radicchi 2008), we considered networks with 1000 nodes and the minimum community size c_{min} of 10 or 20. We varied the mixing parameter μ, which specifies the fraction of the links of a node connecting to nodes outside of the node's community, from 0 to 0.8 with an increment of 0.05. The remaining parameters were kept fixed: the average degree d was set to 20, the maximum degree d_{max} to $2.5\times d$, the maximum community size c_{max} to $5\times c_{min}$, the exponent of power-law distribution of node degrees τ_1 to -2 and community size τ_2 to -1. This design space led to two sets of benchmarks.

Fig. 3 shows the accuracy of each algorithm in NMI as a function of the mixing parameter μ. As shown, our MNDP outperformed Karrer's method, Ball's method and SNMF, especially when μ is in the range of 0.5 to 0.7.

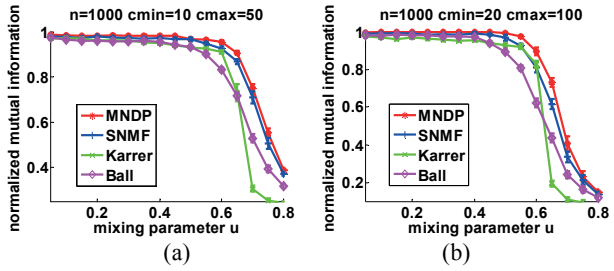

Figure 3: Comparison of different methods in terms of NMI accuracy on the LFR benchmark of disjoint communities. Error bars show the standard deviations estimated from 50 graphs. Shown are results on networks of **(a)** small communities (c_{min} = 10, c_{max} = 50) and **(b)** large communities (c_{min} = 20, c_{max} = 100).

4.1.2 LFR Benchmark with Overlapping Communities

This type of benchmark for overlapping communities was proposed by (Lancichinetti and Fortunato 2009). We used the generalized NMI (GNMI) (Lancichinetti, Fortunato and Kertesz 2009) as accuracy metric. Here we did not include Karrer's method, as it cannot find overlapping structures,

nor Ball's method as it finds highly overlapped structures, thus failing to detect clusters defined in this benchmark.

Adopting the same parameter setting as for the LFR benchmark in (Lancichinetti and Fortunato 2009), we used networks with 1000 nodes, the minimum community size c_{min} of 10 or 20, the mixing parameter μ of 0.1 or 0.3, and varying fraction of overlapping (o_n/n) from 0 to 0.5 with an increment of 0.05. The remaining parameters were fixed: the average node degree d was set to 20, the maximum degree d_{max} to 2.5×d, the maximum community size c_{max} to 5×c_{min}, the number of communities that an overlapping node belongs to (denoted as o_m) to 2, the exponents of the power-law distribution of node degrees τ_1 to -2 and community sizes τ_2 to -1, resulting in four sets of benchmarks.

Fig. 4 shows the results comparing MNDP and SNMF in terms of the GNMI index. As shown, MNDP outperforms SNMF on all four sets of benchmarks. Particularly, when overlapping (o_n/n) is larger, MNDP performs better.

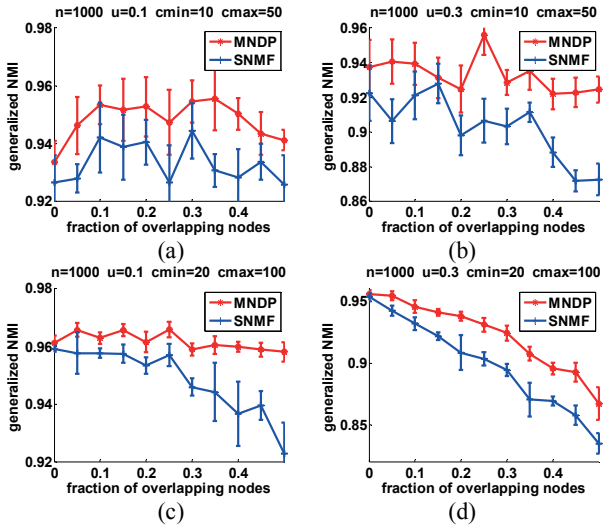

Figure 4: The GNMI accuracy of each algorithm as a function of the fraction of overlapping nodes. Shown are results on networks of **(a)** small mixing parameter & small communities (μ = 0.1, c_{min} = 10, c_{max} = 50), **(b)** big mixing parameter & small communities (μ = 0.3, c_{min} = 10, c_{max} = 50), **(c)** small mixing parameter & big communities (μ = 0.1, c_{min} = 20, c_{max} = 100) and **(d)** big mixing parameter & big communities (μ = 0.3, c_{min} = 20, c_{max} = 100).

4. 2 Real-World Networks

Many real-world networks have distinct topological properties from synthetic networks so that different network analysis methods may render performance on such networks different from that on synthetic networks. We compared the algorithms on two types of real-world networks, i.e., networks with known community structures and networks with actual community structures unknown. We included networks with disjoint and overlapping communities, as well as a large weighted network in comparison.

4.2.1 Networks with Known Community Structures

The real networks analyzed (Newman 2013; Xie, Kelley and Szymanski 2013) and the methods compared are listed in Table 2. For comparison on networks with known community structures, we adopted the NMI index as the quality metric. As shown in the table, the new method MNDP has the best performance on five of the seven networks, and is also competitive with the other methods on the left two networks. On average, our method is 6.7543%, 3.9414%, 1.2814% and 2.5186% more accurate than Karrer's method, Ball's method, SNMF and BNMTF, respectively.

Table 2: Accuracies of 5 methods compared, in NMI index, on 7 networks. 'Friendship6' and 'friendship7' have the same network, but are described by different "true" community structures.

Datasets	n	m	c	NMI index (%)				
				MNDP	Karrer	Ball	SNMF	BNMTF
Zachary's karate club	34	78	2	**100**	83.72	**100**	**100**	**100**
Dolphin social network	62	160	2	**88.88**	**88.88**	81.41	81.41	81.41
High school friendship6	69	220	6	**79.30**	77.02	78.08	78.64	71.22
High school friendship7	69	220	7	84.26	**85.10**	83.93	82.11	84.30
Political books	105	441	3	53.01	54.20	53.39	**56.48**	51.18
American college football	115	613	12	**92.42**	87.06	91.34	90.38	**92.42**
Political blogs	1,490	16,717	2	**71.07**	45.68	53.20	70.95	70.78

4.2.2 Networks with no Known Community Structures

These five methods were further compared on real-world networks whose community structures are unknown (Newman 2013; Nelson, McEvoy and Schreiber 2013). The results are in Table 3. As no "true" community structure is known, Louvain method (Blondel et al. 2008), which is regarded as one of the best algorithms for community detection by (Fortunato 2010), was applied to estimate the numbers of communities that were used by the 5 methods. Because some of the algorithms compared are able to find both disjoint and overlapping community structures, we used two widely used quality metrics: the modularity Q (Newman and Girvan 2004) for evaluating hard partitions and the generalized map equation L for testing overlapping communities (Esquivel and Rosvall 2011). Because Louvain method was designed to optimize Q, it is not surprising that it is able to derive higher Q-values. Besides, as Louvain method is unable to find overlapping communities, we did not include it in comparison here.

Table 3: Some real-world networks without ground-truths.

Datasets	n	m	c (Louvain)
Les Miserables	77	254	6
Word adjacencies	112	425	7
Jazz musicians collaborations	198	2,742	4
C. Elegans neural	297	2,148	5
E. coli metabolic	453	2,025	10
E-mail network URV	1,133	5,451	11
Network Science collaborations	1,589	2,742	277
Power grid	4,941	6,594	39
Word association (unweighted)	5,018	55,234	12
Word association (weighted)	5,018	55,234	30

Table 4: The comparison of community detection algorithms on several real networks. The greater a Q-value, the better; and the smaller a L-value, the better. '–' denotes run time >100 hours. Karrer's method cannot find overlapping communities. Karrer's method and Ball's method cannot deal with weighted networks.

Datasets	Modularity Q (disjoint)					Map equation L (overlaps)			
(abbr)	MNDP	Karrer	Ball	SNMF	BNMTF	MNDP	Ball	SNMF	BNMTF
Les Mis	0.5434	0.4575	0.5211	0.5453	**0.5487**	**4.7283**	4.8922	4.7640	4.7713
Adjnoun	**0.2712**	-0.1041	0.2594	0.2672	0.2634	6.6474	6.7941	6.6869	**6.6397**
Jazz	**0.4377**	0.3696	0.4352	0.4348	0.4347	**6.8327**	6.9121	6.8465	6.8506
Neural	**0.3811**	0.2617	0.3638	0.3701	0.3689	**7.6108**	7.7163	7.6276	7.6893
Metabolic	0.3796	0.2656	0.3689	**0.3879**	0.3834	7.6252	7.7358	**7.5434**	7.6129
Email	**0.5154**	0.5126	0.4814	0.5007	0.4685	**8.6075**	9.0586	8.6202	8.7130
Netscience	**0.8336**	0.6402	0.7296	0.7986	0.7592	5.2338	**3.7445**	5.2922	6.7284
Power	**0.8683**	0.1796	0.6641	0.8649	–	**8.1487**	8.8804	8.1512	–
Word_u	0.3613	**0.4595**	0.3625	0.3546	–	**11.7068**	12.2743	11.7225	–
Word_w	**0.5190**	*NaN*	*NaN*	0.5030	–	**10.4104**	*NaN*	10.4499	–

The results are in Table 4. As shown, MNDP has the best performance on seven of the ten networks in terms of modularity Q. On average, our method is 0.1722, 0.0451, 0.0083 and 0.0193 better than Karrer's method, Ball's method, SNMF and BNMTF, respectively. As Q-values are normally in the range of 0.3 to 0.8 (Newman and Girvan 2004), MNDP obviously outperforms the existing methods. In terms of map equation L, MNDP has the best performance on seven of the ten networks. On average, our method is 0.0963, 0.0153 and 0.2456 bits better than Ball's method, SNMF and BNMTF, respectively. It is known that the improvement of L is typically very small near optimal solutions (see Table II of (Kim and Jeong 2011)), thus our improvement here is also nontrivial. Furthermore, we used the unweighted and weighted versions of the 'word association' network in the comparison. The performance of each method on the weighted network is much better than that on unweighted network. This suggests that edge weights contain additional information of the networks, and thus it is important for an algorithm to be able to incorporate weight information in the weighted networks. Importantly, MNDP outperformed the other two NMF-based methods on both unweighted and weighted versions of this network.

5. Related Works

Several model-based methods have been developed, most of which are built atop the stochastic blockmodel or its variations and employ different optimization methods for model learning. For instance, some works extend the basic blockmodel by introducing soft community membership and use nonnegative matrix factorization (NMF) or its variations as the optimization methods to learn model parameters. In particular, (Wang et al. 2011) use a squared loss and a symmetric nonnegative matrix factorization (SNMF) to minimize the loss function. (Psorakis et al. 2011) adopt generalized KL-divergence as the loss function, and develop a Bayesian nonnegative matrix factorization (BNMF) for optimization. (Zhang and Yeung 2012) remove the constraint that the sum of probabilities for each node belonging to different communities equals to 1 to better model overlapping structures. Further, they use both squared loss and KL-divergence as the loss functions, and devise a bounded nonnegative matrix tri-factorization (BNMTF) for optimization. While these methods are all based on NMF with soft membership, they do not preserve node degrees, which may distort the community detection results.

There are also other works that adopt similar models (Ren et al. 2009; Shen, Cheng and Guo 2011; Karrer and Newman 2011; Ball, Karrer and Newman 2011). However, rather than using loss functions, they adopt likelihood probabilities as the objectives, and take different algorithmic approaches, e.g., Expectation-Maximization algorithm, to learn the models. Of particular interest is the Karrer's method (Karrer and Newman 2011), which attempts to preserve node degrees. However, it does not preserve the node degrees in its model, but achieves this objective by a special inference algorithm, which may limit its generality. Specifically, the expected degree of node i in Karrer's model is $\sum_j \theta_i \theta_j \omega_{g_i g_j}$, which is not necessary equal to node i's degree d_i. For example, if we use squared loss instead of Poisson distribution to fit Karrer's model to the given network, its property of preserving node degree will be lost. In contrast, the model in our MNDP preserves node degrees by itself with no inference, which makes it a generalized model for this problem. Specifically, the expected degree of node i in our model is $\sum_j \left(d_{iz} d_{jz} / \sum_k d_{kz} \right)$, which is equal to node i's degree d_i without any inference (proof in Proposition 1). Furthermore, Karrer's model is an extension of blockmodel to correct node degrees by using a block matrix ω (discussed in Sec. 1) to specify relationships among communities. It has two objective functions to fit: communities and disassortative structures. In contrast, our model is a generalization to the null model of modularity to incorporate community structures, which does not use any types of block matrices. It has one objective (i.e. communities) to fit, which may improve the ability to detect communities. Moreover, Karrer's model focuses on hard clustering that can only detect disjoint communities, while our model uses soft community membership and thus is able to find both disjoint and overlapping communities. Finally, Karrer's model cannot handle weighted networks as it uses Poisson distribution to derive the objective function, while our model can handle weighted networks.

The work proposed by (Ball, Karrer and Newman), which can be considered as a relaxation to Karrer's degree-corrected blockmodel, is also related. Nevertheless, Ball's model was initially designed for link communities and then extended to node communities. In contrast, our model focused on node communities. Furthermore, Ball's model is parameterized by parameters θ_{iz}'s, where θ_{iz} denotes that node i to have links in community z; and then it takes $\theta_{iz}\theta_{jz}$ as the expected number of links in community z connecting nodes i and j. In comparison, our model is parameterized by parameters d_{iz}'s, where d_{iz} is the expected node degree of i in community z; and takes $d_{iz}d_{jz}/\sum_k d_{kz}$ as the expected number of links between nodes i and j in community z, following the null model of modularity. In short, these two models use different ideas to describe communities. Specifically, if we map $d_{iz} = \theta_{iz} \sum_j \theta_{jz}$, Ball's model as-

signs a node to a community for which $d_{iz}/\sum_{j=1}^{n} d_{jz}$ is the largest, while our model assigns a node to a community for which $d_{iz}/\sum_{r=1}^{c} d_{ir}$ is the largest, resulting in two different types of community structures. More importantly, similar to Karrer's model, Ball's model does not preserve node degrees by itself. Specifically, the expected degree of node i in Ball's model is $\sum_j \theta_{iz}\theta_{jz}$, which may not be equal to node i's degree d_i. When using squared loss instead of Poisson distribution to fit the Ball's model to the given network, its objective function will be $\| A - XX^{\mathrm{T}} \|_F^2$, which is a basic NMF and will not preserve node degrees. Finally, Ball's model cannot deal with weighted networks.

We used the null model of modularity to describe probabilistic communities because of its simplicity and good performance. Some other random models with given degree sequence (e.g., that due to Havel-Hakimi (Havel 1955; Hakimi 1962)) may also be suitable for this task. We will include these models in our framework in the future. Our model offers no criterion for determining the number of communities, which is a critical parameter for analyzing community structures. This is a common drawback shared by almost all model methods (Newman 2012). The methods of statistical model selection (Brunet et al. 2004; Tan and Févotte 2012) may in principle be used to address this issue; nevertheless it is too computationally demanding to be useful for any but some small graphs (Ball, Karrer and Newman 2011). More research is needed to address this issue, a direction we plan to pursue in future research.

6. Conclusion

We studied the importance of preserving node degrees for identifying community structures in complex networks. We developed a novel model and an efficient algorithm for finding such structures. They can be used to identify both overlapping and disjoint communities and are applicable to weighted networks. The new model generalizes the null model of modularity (Newman and Girvan 2004) and utilizes probabilistic community memberships to characterize community structures. In our method, we introduced a squared loss function based on this extended model, and optimized the model parameters using a NMF approach.

Our extensive experimental comparison of the new method and four existing competing methods on synthetic benchmarks and real-world networks demonstrated the superior performance of the new MNDP method in detecting both disjoint and overlapping communities.

Acknowledgments

The work was supported by National Basic Research Program (973 Program) of China (2013CB329301, 2012CB316301), Natural Science Foundation of China (61303110, 61271128, 61133011, 31300999), Major Project of National Social Science Fund (14ZDB153), the municipal government of Wuhan, China (2014070504020241 and the Talent Development Program), an internal grant of Jianghan University, Wuhan, China, and United States National Institutes of Health (R01GM100364).

Appendices

A1. Proof of Proposition 1

Let d_i be the degree of node i in G. Given an arbitrary set of variables d_{iz}, subject to $\sum_z d_{iz} = d_i$, which correspond to our model parameters, then using (2) the degree of node i in the expected graph of G can be inferred as

$$\hat{d}_i = \sum_j \hat{w}_{ij} = \sum_j \sum_z \left(d_{iz} d_{jz} / \sum_k d_{kz} \right)$$
$$= \sum_z \left(d_{iz} \sum_j d_{jz} \right) / \sum_k d_{kz} = \sum_z d_{iz} = d_i ,$$

which equals to node i's degree in the observed network G.

A2. Proof of Theorem 1

We adopt the auxiliary function approach used in Expectation-Maximization and NMF. The basic idea is to construct an auxiliary function $C(X, \tilde{X})$ such that:

$$\mathcal{O}(X) = C(X, X) \le C(X, \tilde{X}) \le C(\tilde{X}, \tilde{X}) = \mathcal{O}(\tilde{X})$$

If we can minimize $C(X, \tilde{X})$ w.r.t to X, then we are guaranteed to drive $\mathcal{O}(X)$ down. Note that,

$$\mathcal{O}(X) = Tr(XX^TXX^T)/4 - Tr(A^TXX^T)/2$$
$$+\lambda Tr(1_n^TXX^TXX^T1_n)/2 - \lambda Tr(1_n^TXX^Td)$$
$$\le Tr(P\tilde{X}\tilde{X}^T)/4 + \lambda Tr(P1_n1_n^T\tilde{X}\tilde{X}^T)/2 - Tr(A^TXX^T)/2$$
$$-\lambda Tr(d1_n^TXX^T) \text{ (by Lemma 6 of (Wang et al. 2011))}$$
$$\le Tr(R^T\tilde{X}\tilde{X}^T\tilde{X})/4 + \lambda Tr(R^T1_n1_n^T\tilde{X}\tilde{X}^T\tilde{X})/4$$
$$+ \lambda Tr(\tilde{X}^T1_n1_n^T\tilde{X}\tilde{X}^TR)/4 - Tr(Z^TA\tilde{X}) - \lambda Tr(Z^Td1_n^T\tilde{X})$$
$$- \lambda Tr(\tilde{X}^Td1_n^TZ) - Tr(\tilde{X}^TA\tilde{X})/2 - \lambda Tr(\tilde{X}^Td1_n^T\tilde{X})$$
(by Lemma 7 and 3 of (Wang et al. 2011))
$$\equiv C(X, \tilde{X}),$$

where $P_{kl} = [XX^T]_{kl}^2/[\tilde{X}\tilde{X}^T]_{kl}$, $R_{ik} = [X]_{ik}^4/[\tilde{X}]_{ik}^3$, and $Z_{ij} = \tilde{X}_{ij}\ln(X_{ij}/\tilde{X}_{ij})$. The equality holds when $X = \tilde{X}$. Then $C(X, \tilde{X})$ satisfied the conditions of being an auxiliary function for $\mathcal{O}(X)$. So we can define the updating rules as:

$$X^{(t+1)} = \min_X C(X, X^{(t)})$$

$$\frac{\partial C(X, \tilde{X})}{\partial X_{ik}} = \frac{X_{ik}^3}{\tilde{X}_{ik}^3}[\tilde{X}\tilde{X}^T\tilde{X} + \lambda 1_n1_n^T\tilde{X}\tilde{X}^T\tilde{X} + \lambda\tilde{X}\tilde{X}^T1_n1_n^T\tilde{X}]_{ik}$$
$$-\frac{\tilde{X}_{ik}}{X_{ik}}[A\tilde{X} + \lambda d1_n^T\tilde{X} + \lambda 1_n d^T\tilde{X}]_{ik} = 0$$

Thus we get the update rule for X as in (8).

References

Ahn, Y. Y.; Bagrow, J. P.; and Lehmann, S. 2010. Link communities reveal multiscale complexity in networks. *Nature* 466: 761-764.

Ball, B.; Karrer, B.; and Newman, M. E. J. 2011. Efficient and principled method for detecting communities in networks. *Phys. Rev.* E 84: 036103.

Barabasi, A.; and Albert, R. 1999. Emergence of scaling in random networks. *Science* 286: 509-512.

Blondel, V. D.; Guillaume, J. L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008: P10008.

Boyd, S., and Vandenberghe, L. eds. 2004. *Convex optimization*. Cambridge, UK: Cambridge University Press.

Brunet, J-P.; Tamayo, P.; Golun, T. R.; and Mesirov, J. P. 2004. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci.* 101: 4164-4169.

Danon, L.; Duch, J.; Diaz-Guilera, A.; and Arenas, A. 2005. Comparing community structure identification. *J. Stat. Mech.* 2005: P09008.

Esquivel, A. V.; and Rosvall, M. 2011. Compression of flow can reveal overlapping-module organization in networks. *Physical Review* X 1: 021025.

Fortunato, S. 2010. Community detection in graphs. *Phys. Rep.* 486: 75-174.

Girvan, M.; and Newman, M. E. J. 2002. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* 99: 7821-7826.

Havel, V. 1955. A remark on the existence of finite graphs. *Časopis pro pěstování matematiky (in Czech)* 80: 477-480.

Hakimi, S. L. 1962. On realizability of a set of integers as degrees of the vertices of a linear graph. *Journal of the Society for Industrial and Applied Mathematics* 10: 496-506.

Karrer, B.; and Newman, M. E. J. 2011. Stochastic blockmodels and community structure in networks. *Phys. Rev.* E 83: 016107.

Kim, Y.; and Jeong, H. 2011. Map equation for link communities. *Phys Rev* E 84: 026110.

Lancichinetti, A.; Fortunato, S.; and Kertesz, J. 2009. Detecting the overlapping and hierarchical community structure in complex networks. *New J. Phys.* 11: 033015.

Lancichinetti, A.; Fortunato, S.; and Radicchi, F. 2008. Benchmark graphs for testing community detection algorithms, *Phys. Rev.* E 78: 046110.

Lancichinetti, A.; and Fortunato, S. 2009. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev.* E 80: 016118.

Nelson, D. L.; McEvoy, C. L.; and Schreiber, T. A. 2013. The University of South Florida, word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation/

Newman, M. E. J.; and Girvan, M. 2004. Finding and evaluating community structure in networks. *Phys. Rev.* E 69: 026113.

Newman, M. E. J. 2012. Communities, modules and large-scale structure in networks. *Nature Physics* 8: 25-31.

Newman, M. E. J. 2013. Real-world network data in Newman's hompage. http://www-personal.umich.edu/~mejn/netdata/

Nowicki, K.; and Snijders, T. A. B. 2001. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96: 1077-1087.

Palla, G.; Derenyi, I.; Farkas, I.; and Vicsek, T. 2005. Uncovering the overlapping community structures of complex networks in nature and society. *Nature* 435: 814-818.

Psorakis, I.; Roberts, S.; Ebden, M.; and Sheldon, B. 2011. Overlapping community detection using Bayesian non-negative matrix factorization. *Phys. Rev.* E 83: 066114.

Ren, W.; Yan, G.; Liao, X.; and Xiao, L. 2009. Simple probabilistic algorithm for detecting community structure. *Phys. Rev.* E 79: 036111.

Rosvall, M.; and Bergstrom, C. T. 2008. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* 105: 1118-1123.

Shen, H.; Cheng, X.; and Guo, J. 2011. Exploring the structural regularities in networks. *Phys. Rev.* E 84: 056111.

Tan, V. Y. F.; and Févotte, C. 2012. Automatic relevance determination in nonnegative matrix factorization with the β-divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35: 1592-1605.

Wang, F.; Li, T.; Wang, X.; Zhu, S.; and Ding, C. H. Q. 2011. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery* 22: 493-521.

Xie, J.; Kelley, S.; and Szymanski, B. K. 2013. Overlapping community detection in networks: the state of the art and comparative study. *ACM Comput Surv* 45: Article No. 43.

Zachary, W. W. 1977. An information flow model for conflict and fission in small groups. *J Anthropol Res* 33: 452-473.

Zhang, Y.; and Yeung, D. 2012. Overlapping community detection via bounded nonnegative matrix tri-factorization. *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 606-614. New York, NY: ACM Press.