Semantic community identification in large attribute networks

Xiao Wang[1,5], Di Jin[1], Xiaochun Cao[2,*], Liang Yang[2,3], Weixiong Zhang[4,5]
[1]School of Computer Science and Technology, Tianjin University, Tianjin 300072, China
[2]State Key Laboratory of Information Security, IIE, Chinese Academy of Sciences, Beijing 100093, China
[3]School of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China
[4]College of Math and Computer Science, Institute for Systems Biology, Jianghan University, Wuhan, Hubei 430056, China
[5]Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA
{wangxiao_cv, jindi}@tju.edu.cn, {caoxiaochun, yangliang}@iie.ac.cn, weixiong.zhang@wustl.edu
*Corresponding author.

Abstract

Identification of modular or community structures of a network is a key to understanding the semantics and functions of the network. While many network community detection methods have been developed, which primarily explore network topologies, they provide little semantic information of the communities discovered. Although structures and semantics are closely related, little effort has been made to discover and analyze these two essential network properties together. By integrating network topology and semantic information on nodes, e.g., node attributes, we study the problems of detection of communities and inference of their semantics simultaneously. We propose a novel nonnegative matrix factorization (NMF) model with two sets of parameters, the community membership matrix and community attribute matrix, and present efficient updating rules to evaluate the parameters with a convergence guarantee. The use of node attributes improves upon community detection and provides a semantic interpretation to the resultant network communities. Extensive experimental results on synthetic and real-world networks not only show the superior performance of the new method over the state-of-the-art approaches, but also demonstrate its ability to semantically annotate the communities.

Introduction

Complex systems can be represented in networks or graphs. One of the most prominent features of such networks is the community structure, where the nodes within a community are densely connected whereas nodes in different communities are sparsely connected (Girvan and Newman 2002). Community structures help reveal organizational structures and functional components of a complex system. Therefore, community detection is an essential step toward characterization of a complex system.

Network topology, an important network description, has been broadly exploited by the most existing methods for community detection. However, network topology reflects merely one aspect of a network and is often noisy. As a result, using network topology alone may not necessarily give rise to a satisfactory partition of a network. For instance, it is not uncommon that two nodes that belong to the same community are not directly connected, and a node connecting to multiple communities for distinct reasons is difficult to be assigned correctly to the right communities by only relying on network topology. Therefore, it is insufficient to accurately determine the community structure using network topology alone. In addition to network topology, semantic information, e.g., that of node attributes, is often available. For example, a node (i.e., a person) in a social network is often annotated by a personal profile with information such as education background, circle of friends and profession; a node (i.e., a paper) in a citation network is typically annotated with title, abstract and key words. Different from network topology, node semantics capture characteristics of individual nodes and provide a piece of valuable information orthogonal to information of network topology. Integration of network topological and semantic information holds a great potential for community identification.

Nevertheless, it is technically challenging to effectively combine these two pieces of valuable albeit orthogonal information. Particularly, two obstacles need to be addressed in order to properly integrate these two types of information. First, how to adequately characterize a community. The most existing methods for community detection mainly rely on network topologies. However, missing, meaningless or even erroneous edges are ubiquitous in real networks, which casts doubts on the accuracy and/or correctness of the network communities discovered based on network topology alone. While the nodes in a community are highly connected, they should also have similar characteristics, reflected by attributes. Thus, nodes attributes may carry essential information of communities that is complementary to the information of network topology. Therefore, even though two nodes are not directly connected, they may belong to the same community if they share the same characteristics, and the use of node attributes may enhance community discovery. Second, how to adequately interpret or semantically annotate communities. Functional analysis of network communities is typically and independent, post-processing task following community detection. The result from a community discovery often provides little information beyond network topology regarding why a group of nodes from a community, their semantic meaning, or potential functions. In order to semantically annotate a community, supplemental information, e.g., background information and/or domain knowledge, is usually required. Even though such domain information is available, how to fully utilize such information remains challenging, application specific and time consuming.

To address the above two problems, we propose and develop in this paper a method, named as Semantic Community Identification (SCI), to identify network communities with semantic annotation. The SCI method integrates network topological and node semantic information; it combines topology based community memberships and node-attribute based community attributes (or semantics) in the framework of nonnegative matrix factorization (NMF, (Seung and Lee 2001)). The key intuition behind SCI stems in two observations: two nodes are likely to be connected if their community memberships are similar, and two nodes likely belong to the same community if their attributes are consistent with the underlying community attributes to be learned. To make the novel SCI method effective, we introduce a sparsity penalty in order to select the most related attributes for each community and devise a multiplicative updating rule with a convergence guarantee. Extensive experiments on synthetic and real networks, in comparison with several state-of-the-art methods, are performed to assess the performance of SCI.

Related work

Several community detection methods, as reviewed in (Xie *et al.* 2013), have been developed to explore network topologies, including the well-known ones based on nonnegative matrix factorization (NMF) (Wang *et al.* 2011; Yang and Leskovec 2013) and stochastic blockmodel (SBM) (Karrer and Newman 2011). Among these methods are ones that combines network topologies and node attributes (content or features). In particular, a unified method was suggested to combine a conditional model for topology analysis and a discriminative model for making use of node attributes (Yang *et al.* 2009). However, this method focuses on community detection without inferring the most relevant attributes for each community. Edge content was also leveraged to improve community detection processes (Qi *et al.* 2012). However, this method is specifically designed for detecting communities of links, rather than communities of nodes. A heuristic linear combination between edges created by node attributes and the topological information of edges was proposed to create a new graph, which was used for the graph clustering (Ruan *et al.* 2013). However, this strategy did not use the semantic information of attributes when inferring topics of communities. A probabilistic model that can capture the relationship between community and attributes was developed (Yang *et al.* 2013), which simply added a sparsity term to the whole network rather than each community. Moreover, the updating rules for learning the parameters of the model are not guaranteed to converge. A heuristic algorithm to optimize the community score for recovering communities and minimize description complexity for inferring diverse community descriptions was proposed (Pool *et al.* 2014); this heuristic method reported too many relatively small communities, some of which have two or three nodes. A nonnegative matrix tri-factorization based clustering framework with graph regularization was proposed to combine social relations and user generated content in a social network (Pei *et al.* 2015). However, this method focused on utilizing additional content information to detect communities, and failed to study the relationship between communities and these content.

SCI: The network model

Consider an undirected network $G = (V, E)$ with n nodes V and e edges E, represented by a binary-valued adjacency matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$. Associated with each node i are its attributes $\mathbf{S}_i$, which may be semantic characteristics of the node. The attributes of a node are in the form of an m-dimensional binary-valued vector, and the attributes of all the nodes can be represented by a node attribute matrix $\mathbf{S} \in \mathbb{R}^{n\times m}$. The problem of community identification is to partition the network G into k communities as well as to infer the related attributes or semantics of each community.

Modeling network topologies. We define the propensity of node i belonging to community j as U_{ij}. The community membership of all the nodes in the network is then $\mathbf{U} = (U_{ij})$, where $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$. Consequently $U_{ir}U_{pr}$ presents the expected number of edges between nodes i and p in community r. Summing over all communities, the expected number of edges between i and p is $\sum_{r=1}^{k} U_{ir}U_{pr}$. This process of generating edges implies that if two nodes have similar community memberships, they have a high propensity to be linked. The expected number of edges between pairs of nodes should be as closely consistent as possible with the network topology denoted by $\mathbf{A}$, which gives rise to the following function in matrix formulation:

$$\min_{\mathbf{U}\geq 0} \|\mathbf{A} - \mathbf{U}\mathbf{U}^T\|_F^2. \tag{1}$$

Modeling node attributes. We define the propensity of community r to have attribute q as C_{qr}. So for all the communities, we have a community attribute matrix $\mathbf{C} = (C_{qr})$, for $q = 1, 2, ..., m$ and $r = 1, 2, ..., k$, where the r-th column, $\mathbf{C}_r$, is the attribute membership of community r. If the attributes of a node are highly similar to that of a community, the node may have a high propensity to be in the community. As a result, the nodes with similar attributes, described in $\mathbf{S}_i$, may form a community, which can be characterized by the common attributes of the nodes. Specifically, the propensity of node i belonging to community r can be formulated as $U_{ir} = \mathbf{S}_i\mathbf{C}_r$. Notice that if the attributes of node i and community r are completely inconsistent, node i must not belong to community r, i.e., $U_{ir} = 0$. As the community memberships of all the nodes $\mathbf{U}$ offer a guidance for combing the attributes of nodes and communities, we have the following optimization function:

$$\min_{\mathbf{C}\geq 0} \|\mathbf{U} - \mathbf{S}\mathbf{C}\|_F^2. \tag{2}$$

In order to select the most relevant attributes for each community, we add an l_1 norm sparsity to each column of matrix $\mathbf{C}$. In addition, to prevent the values of some columns of $\mathbf{C}$ too large, which means that each community has some meaningful attributes, we have the constraint on $\mathbf{C}$

$\sum_{j=1}^{k} \|\mathbf{C}(:,j)\|_1^2$, which gives rise to the following objective function together with (2):

$$\min_{\mathbf{C}\geq 0} \|\mathbf{U}-\mathbf{SC}\|_F^2 + \alpha \sum_{j=1}^{k} \|\mathbf{C}(:,j)\|_1^2, \tag{3}$$

where α is a nonnegative parameter to make a tradeoff between the first error term and the second sparsity term.

The unified model. By combining the objective functions of modeling the network topology specified in (1) and of modeling node attributes in (3), we have the following overall function:

$$\min_{\mathbf{U}\geq 0, \mathbf{C}\geq 0} L = \|\mathbf{U}-\mathbf{SC}\|_F^2 + \alpha \sum_{j=1}^{k} \|\mathbf{C}(:,j)\|_1^2 + \beta \|\mathbf{A}-\mathbf{UU}^T\|_F^2, \tag{4}$$

where β is a positive parameter for adjusting the contribution of network topologies.

Optimization

Since the objective function in (4) is not convex, it is impractical to obtain the optimal solution. Local minima of (4) can be achieved using Majorization-Minimization framework (Hunter and Lange 2004). Here we describe an algorithm that iteratively updates $\mathbf{U}$ with $\mathbf{C}$ fixed and then $\mathbf{C}$ with $\mathbf{U}$ fixed, which guarantees not to increase the objective function after each iteration. The specific formulas are shown as the following two subproblems.

U**-subproblem:** when update $\mathbf{U}$ with $\mathbf{C}$ fixed, we need to solve the following problem:

$$\min_{\mathbf{U}\geq 0} L(\mathbf{U}) = \|\mathbf{U}-\mathbf{SC}\|_F^2 + \beta \|\mathbf{A}-\mathbf{UU}^T\|_F^2. \tag{5}$$

To this end, we introduce a Lagrange multiplier matrix $\mathbf{\Theta} = (\Theta_{ij})$ for the nonnegative constraints on $\mathbf{U}$, resulting in the following equivalent objective function:

$$\begin{aligned} L(\mathbf{U}) = & \, tr(\mathbf{UU}^T - \mathbf{UC}^T\mathbf{S}^T - \mathbf{SCU}^T + \mathbf{SCC}^T\mathbf{S}^T) \\ & + \beta tr(\mathbf{AA} - \mathbf{AUU}^T - \mathbf{UU}^T\mathbf{A} + \mathbf{UU}^T\mathbf{UU}^T) \\ & + tr(\mathbf{\Theta U}^T). \end{aligned} \tag{6}$$

Set derivative of $L(\mathbf{U})$ with respect to $\mathbf{U}$ to 0, we have:

$$\mathbf{\Theta} = -2\mathbf{U} + 2\mathbf{SC} + 4\beta\mathbf{AU} - 4\beta\mathbf{UU}^T\mathbf{U}. \tag{7}$$

Following the Karush-Kuhn-Tucker (KKT) condition for the nonnegativity of $\mathbf{U}$, we have the following equation:

$$(-2\mathbf{U} + 2\mathbf{SC} + 4\beta\mathbf{AU} - 4\beta\mathbf{UU}^T\mathbf{U})_{ij} U_{ij} = \Theta_{ij} U_{ij} = 0. \tag{8}$$

This is the fixed point equation that the solution must satisfy at convergence. Given an initial value of $\mathbf{U}$, the successive update of $\mathbf{U}$ is:

$$U_{ij} \leftarrow U_{ij} \left(\frac{(\mathbf{SC} + 2\beta\mathbf{AU} - \mathbf{U})_{ij}}{2\beta(\mathbf{UU}^T\mathbf{U})_{ij}} \right)^{\frac{1}{4}}. \tag{9}$$

To guarantee the property that $\mathbf{U}$ is nonnegative, we set the diagonal elements in $\mathbf{A}$ to be larger than $\frac{1}{2\beta}$. The updating rule of $\mathbf{U}$ satisfies the following theorem, which guarantees the correctness of the rule.

Theorem 1. If the update rule of $\mathbf{U}$ converges, then the final solution satisfies the KKT optimality condition. *(Proof in Appendix A1).*

We now prove the convergence of the updating rule. Following (Seung and Lee 2001), we use an auxiliary function to achieve this goal.

Definition 1. (Seung and Lee 2001) A function $Q(\mathbf{U},\mathbf{U}')$ is an auxiliary function of function $L(\mathbf{U})$ if $Q(\mathbf{U},\mathbf{U}') \geq L(\mathbf{U})$, $Q(\mathbf{U},\mathbf{U}) = L(\mathbf{U})$ for any $\mathbf{U},\mathbf{U}'$.

The auxiliary function is useful because of the following lemma:

Lemma 1. (Seung and Lee 2001) If Q is an auxiliary function of L, then L is nonincreasing under the update rule $\mathbf{U}^{(t+1)} = arg\min_{\mathbf{U}} Q(\mathbf{U},\mathbf{U}^{(t)})$.

Now we have the specific form of the auxiliary function $Q(\mathbf{U},\mathbf{U}')$ for the objective function $L(\mathbf{U})$ in problem (5) based on the following lemma.

Lemma 2. The function

$$\begin{aligned} Q(\mathbf{U},\mathbf{U}') = & \, tr(\mathbf{SCC}^T\mathbf{S}^T + \beta\mathbf{AA}) + \beta tr(\mathbf{RU'}^T\mathbf{U'U'}^T) \\ & - tr(\mathbf{U'}^T\mathbf{A'Z}) - tr(\mathbf{Z}^T\mathbf{A'U'}) - tr(\mathbf{U'}^T\mathbf{A'U'}) \\ & - 2tr(\mathbf{C}^T\mathbf{S}^T\mathbf{Z}) - 2tr(\mathbf{C}^T\mathbf{S}^T\mathbf{U'}) \end{aligned} \tag{10}$$

is an auxiliary function for $L(\mathbf{U})$ in problem (5), where $R_{ij} = \frac{U_{ij}^4}{U_{ij}'^3}$, $Z_{ij} = U'_{ij} \ln \frac{U_{ij}}{U'_{ij}}$, $\mathbf{A}' = 2\beta\mathbf{A} - \mathbf{I}$, and $\mathbf{I}$ is an identity matrix. *(Proof in Appendix A2).*

Based on Lemmas 1 and 2, we can show the convergence of the updating rule.

Theorem 2. The problem (5) is nonincreasing under the iterative updating rule (9). *(Proof in Appendix A3).*

C**-subproblem:** when update $\mathbf{C}$ with $\mathbf{U}$ fixed, we need to solve the following problem:

$$\min_{\mathbf{C}\geq 0} L(\mathbf{C}) = \|\mathbf{U}-\mathbf{SC}\|_F^2 + \alpha \sum_{j=1}^{k} \|\mathbf{C}(:,j)\|_1^2. \tag{11}$$

This is equivalent to the following optimization problem (Kim and Park 2008):

$$\min_{\mathbf{C}\geq 0} L(\mathbf{C}) = \left\| \begin{pmatrix} \mathbf{S} \\ \sqrt{\alpha}\mathbf{e}_{1\times m} \end{pmatrix} \mathbf{C} - \begin{pmatrix} \mathbf{U} \\ \mathbf{0}_{1\times k} \end{pmatrix} \right\|_F^2, \tag{12}$$

where $\mathbf{e}_{1\times m}$ is a row vector with all components equal to one and $\mathbf{0}_{1\times k}$ is a zero vector. So we have the following updating rule for (12):

$$\mathbf{C}_{ij} \leftarrow \mathbf{C}_{ij} \frac{(\mathbf{S'}^T\mathbf{U'})_{ij}}{(\mathbf{S'}^T\mathbf{S'C})_{ij}}, \tag{13}$$

where $\mathbf{S}' = \begin{pmatrix} \mathbf{S} \\ \sqrt{\alpha}\mathbf{e}_{1\times m} \end{pmatrix}$ and $\mathbf{U}' = \begin{pmatrix} \mathbf{U} \\ \mathbf{0}_{1\times k} \end{pmatrix}$. The convergence of (13) can be shown as in (Seung and Lee 2001).

At convergence, as $\mathbf{U}$ expresses the soft membership distribution over communities, we can either use $\mathbf{U}$ directly or $\mathbf{U} = \mathbf{SC}$ to get the final disjoint or overlapping communities. Each column of $\mathbf{C}$ indicates the relationship between a community and the attributes, where a larger value represents the more relevant the corresponding attribute to the community.

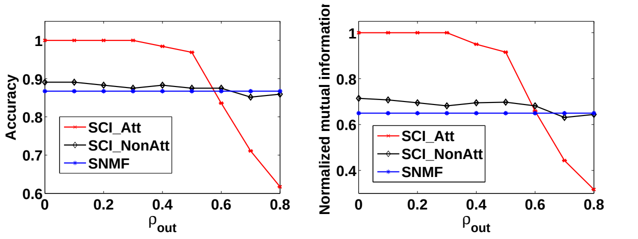

Figure 1: Performance comparison of SCI using $\mathbf{U}$ directly (SCI_NonAtt), SCI using $\mathbf{U} = \mathbf{SC}$ (SCI_Att) and SNMF.

Experimental evaluations

Synthetic network

We first evaluated SCI on a synthetic network constructed using the widely adopted Newman's model (Girvan and Newman 2002). The network consists of 128 nodes divided into 4 disjoint communities. Each node has on average z_{in} edges connecting it to members of the same community and z_{out} edges to members of other communities, with $z_{in} + z_{out} = 16$. Here we set z_{in} and z_{out} to 8, which becomes a challenging problem for most methods as there was no obvious community structure (Yang *et al.* 2014). Then we generated a $4h_{in}$-dimensional binary attributes for each node as follows. For each node within the i-th community, we used a binomial distribution with mean ρ_{in} to generate a h_{in}-dimensional binary vector as its $((i-1) \times h_{in} + 1)$-th to $(i \times h_{in})$-th attributes, and generated the rest attributes using a binomial distribution with mean ρ_{out}. In total, we have $(4h_{in})$-dimensional attributes for each node. Note that $\rho_{in} > \rho_{out}$, meaning that these generated h_{in}-dimensional attributes are associated with this community with high probability, while the rest are irrelevant (or noisy) attributes.

As mentioned earlier, the new method infers two parameters $\mathbf{U}$ and $\mathbf{C}$, so we may use the inferred $\mathbf{U}$ directly or the attributes $\mathbf{C}$ to derive a new $\mathbf{U}$ as in $\mathbf{U} = \mathbf{SC}$ to recover community structures. For convenience, we named these two schemes as SCI_NonAtt and SCI_Att, respectively. Our experiments were first designed to study the difference between these two schemes. We set $h_{in} = 50$, $\rho_{in} = 0.8$ and varied ρ_{out} from 0 to 0.8 with an increment of 0.1. We adopted SNMF (Wang *et al.* 2011) using network topologies alone as the baseline method for comparison. We used accuracy (AC) (Liu *et al.* 2012) and normalized mutual information (NMI) (Liu *et al.* 2012) as the quality metrics for performance assessment. As shown in Figure 1, both SCI_Att and SCI_NonAtt outperform SNMF, except when ρ_{out} almost reaches 0.8. The result shows that the quality of identified communities improves with the information of node attributes. Furthermore, SCI_Att usually significantly outperforms SCI_NonAtt before $\rho_{out} = 0.5$. As ρ_{out} increases beyond 0.5, the performance of SCI_Att deteriorates. This is in part because when ρ_{out} is greater than 0.5, node attributes provide less discriminative information on network community, meaning that there are less specific attributes associated with communities. This also implies that node attributes may not always be valuable to community detection, but rather may distort the results if they have low quality. However, in general, the node attributes have the underlying discriminative power that can be beneficial for distinguishing communities. So instead of using $\mathbf{U}$ directly, we specified $\mathbf{U} = \mathbf{SC}$ as the final community membership for all the following experiments.

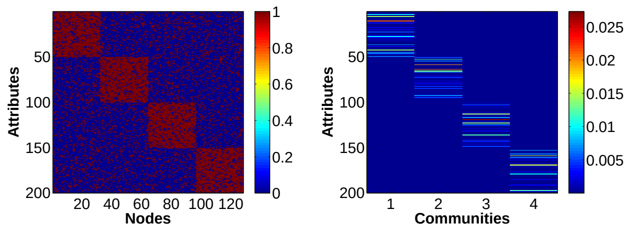

Figure 2: **Left:** the node attributes matrix. **Right:** the inferred community attributes matrix.

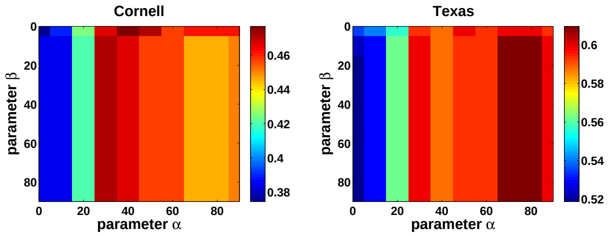

Figure 3: The effect of parameter α and β. Different color means different accuracy and color close to red indicates high accuracy.

Further, we studied the community attributes $\mathbf{C}$ inferred by SCI. We fixed $\rho_{in} = 0.8$, $\rho_{out} = 0.2$, and $h_{in} = 50$. The generated node attribute matrix is shown in the left figure of Figure 2. As shown, the nodes of each the communities have 50-dimensional relevant attributes, and the rest attributes are irrelevant. We noticed that the attributes of each community are very different, as shown in the right figure of Figure 2, meaning that unique attributes have been recovered for each community. Besides, the community attributes are consistent with the relevant attributes of the nodes in the community. In short, the new method is able to identify network modular structures as well as infer community attributes which provide semantic information of the communities.

Real networks

We considered three real networks with node attributes and ground-truth community labels. The CiteSeer network[1] (6 communities) consists of 3312 scientific publications with 4732 edges, and the Cora[1] network (7 communities) consists of 2708 scientific publications with 5429 edges. The publications in Citeseer and Cora are associated with 3703- and 1433-dimensional binary-valued word attributes, respectively, indicating whether a corresponding word is in a publication. The WebKB network[1] consists of 4 subnetworks gathered from 4 universities (Cornell, Texas, Washington and Wisconsin). Each subnetwork is divided into 5 communi-

[1] http://linqs.cs.umd.edu/projects/projects/lbc/

Table 1: Performance comparison of disjoint communities (bold numbers represent the best results).

Metrics	Methods	Cornell	Texas	Washington	Wisconsin	Cora	Citeseer
AC	PCL-DC	0.3487	0.3690	0.4087	0.3547	**0.5543**	**0.6525**
	SNMF	0.3179	0.3583	0.2783	0.3283	0.4173	0.2539
	SBM	0.3436	0.3743	0.2826	0.2981	0.3833	0.2844
	CAN	0.4154	0.4706	0.5087	0.4717	0.3021	0.2129
	SMR	0.3179	0.5401	0.4565	0.4226	0.3002	0.2111
	SCI	**0.4769**	**0.6096**	**0.5435**	**0.5245**	0.4169	0.3442
NMI	PCL-DC	0.0813	0.0686	0.1031	0.0719	**0.3830**	**0.3816**
	SNMF	0.0332	0.0476	0.0211	0.0803	0.1994	0.0403
	SBM	0.0543	0.0839	0.0211	0.0428	0.2047	0.0512
	CAN	0.0614	0.0908	0.1175	0.0702	0.0132	0.0079
	SMR	0.0845	0.1150	0.0381	0.0777	0.0078	0.0032
	SCI	**0.1520**	**0.2197**	**0.2096**	**0.1852**	0.1780	0.0922

ties. There are 877 webpages with 1608 edges. Each webpage is annotated by 1703-dimensional binary-valued word attributes.

We compared SCI against three topology based methods: SNMF (Wang *et al.* 2011), SBM (Karrer and Newman 2011), BIGCLAM (Yang *et al.* 2013); two node attributes based methods: CAN (Nie *et al.* 2014) and SMR (Hu *et al.* 2014); three methods that combine network topologies and node attributes: PCL_DC (Yang *et al.* 2009), CESNA (Yang *et al.* 2013), DCM (Pool *et al.* 2014). The methods compared may provide disjoint or overlapping communities, so we chose different evaluation metrics. For disjoint communities, we adopted accuracy (AC) (Liu *et al.* 2012) and normalized mutual information (NMI) (Liu *et al.* 2012). For overlapping communities, generalized normalized mutual information (GNMI) (Lancichinetti *et al.* 2009) was used. In addition, we compare a set of detected communities M with the ground-truth communities M^* as in (Yang *et al.* 2013): $\frac{1}{2|M^*|}\sum_{M_i^*\in M^*}\max_{M_j\in M}\delta(M_j^*,M_j)+\frac{1}{2|M|}\sum_{M_j\in M}\max_{M_j^*\in M^*}\delta(M_j^*,M_j)$, where $\delta(M_i^*,M_j)$ is a similarity measure (F-score and Jaccard similarity) between communities M_i^* and M_j.

We verified the effectiveness of SCI on both disjoint and overlapping community results, shown in Tables 1 and 2, respectively. As shown in Table 1, SCI outperforms the other methods on four of the six network instances (Cornell, Texas, Washington and Wisconsin). As shown in Table 2, when measured with F-score and Jaccard metrics, SCI achieves the best performances on all the tested networks; it outperforms the other methods on four of the six networks in terms of GNMI, further demonstrating the effectiveness of SCI. Besides, we noticed that the accuracies to be vastly different across different networks (even using the same set of methods), and this may reflect diverse characteristics of networks analyzed.

We tested the effect of parameters α and β of the new method on the real networks, i.e., α and β are the parameters for adjusting the contributions of sparsity term and network topologies, respectively. We varied each parameter from 1 to 100 with an increment of 10. Because the results of different networks have similar tends, here we just showed two networks (Cornell and Texas) in Figure 3. Notice that SCI is relatively stable with varying parameter β, whereas it is significantly affected by α, suggesting the importance of the sparsity term. Therefore we suggest to set β to either 1 or a value between 10 and 100 and fine tune $\alpha \in \{1, 10, 20, ..., 100\}$ so as to achieve a high performance.

Since SCI converges to the local optimum, we tested its robustness on Cornell, Texas, Washington and Wisconsin datasets. We repeated SCI with ten different initializations. The mean values of loss functions are $81.9509 \pm 0.2844, 87.8448 \pm 0.1728, 102.4153 \pm 0.2852, 105.5379 \pm 0.4075$, respectively. The variances are all less than 0.4%, which shows the stability of SCI. The main computation of SCI is for the updating rules in (9) and (13). The complexity is $O(T(mnk + n^2k))$ for T iterations to converge. We also reported the running time of SCI on Cornell, Texas, Washington, Wisconsin, Cora and Citeseer here. On a PC with "RAM: 8G; CPU: Intel I7; Platform: Matlab", they are $0.4509s, 0.1917s, 0.3234s, 0.4571s, 88.6821s$ and $69.8382s$, respectively.

Analysis of detected communities

We closely examined some of the communities detected by SCI. Here we used LASTFM dataset[2] from an online music system *Last.fm*, whose 1892 users are connected in a social network generated from *Last.fm* "friend" relations. Each user has 11946-dimensional attributes, including a list of most listened music artists, and tag assignments. Because the network does not have ground-truth labels, we did not quantitatively evaluate it in the previous section. We used Louvain method (Blondel *et al.* 2008) to set the number of communities to 38. Four example community attributes are shown as word clouds in Figure 4. The size of a word is proportional to its community attribute value, i.e., more relevant an attribute, larger it is in the figure.

For each community, we selected the top ten attributes. We observed these four communities have their unique attributes. In particular, the community in Figure 4 (a) shows that this is a group of fans of "heavy metal" bands or music. For example, "metallica," "queensryche," "backyard babies," "sound garden" and "skid row" are all heavy mental bands. Besides, the music genre of "slash" and "nik-

[2] http://ir.ii.uam.es/hetrec2011/datasets.html

Table 2: Performance comparison of overlapping communities (bold numbers represent the best results).

Metrics	Methods	Cornell	Texas	Washington	Wisconsin	Cora	Citeseer
GNMI	BIGCLAM	0.0051	0.0034	0.0028	0	0.0244	5.551e-17
	CESNA	0.0704	0.0008	**0.1151**	**0.1573**	0.0179	0
	DCM	1.110e-16	0.0090	0.0062	1.110e-16	2.220e-16	0
	SCI	**0.0901**	**0.0955**	0.0859	0.0879	**0.1039**	**0.0506**
F-score	BIGCLAM	0.2267	0.2097	0.2002	0.2399	0.2927	0.1386
	CESNA	0.3368	0.2352	0.3527	0.4393	0.3160	0.1360
	DCM	0.1438	0.0908	0.1127	0.1052	0.0345	0.0245
	SCI	**0.4766**	**0.4740**	**0.4718**	**0.5063**	**0.3835**	**0.3651**
Jaccard	BIGCLAM	0.1294	0.1190	0.1120	0.1380	0.1797	0.0829
	CESNA	0.2120	0.1406	0.2551	0.3164	0.1940	0.0794
	DCM	0.0795	0.0484	0.0607	0.0563	0.0177	0.0125
	SCI	**0.3225**	**0.3413**	**0.3303**	**0.3642**	**0.2519**	**0.2275**

ki sixx" also includes heavy mental. Particularly, the tags "heavy mental" and "glam punk" appear here. The topic of the community in Figure 4 (b) should be related to singer "rihanna" or popular music, because the word "rihanna" is the largest and she is one of the best-selling artists of all time and featured on the worldwide hits. Her song "We Found Love" was ranked by Billboard as the 24th biggest US Billboard Hot 100 hit of all time. "raining men" is one of her songs, and "rated r" is her fourth studio album. "xtina" is another popular singer "Christina Aguilera". For the community in Figure 4 (c), it is mainly related to the rock band "duran duran" and the rock music. Moreover, "new romantic," synth-rock" and "new wave" are all their genres. Also, according to Wikipedia[3], "supergroup" is usually used in the context of rock and pop music and "duran duran" is one of them. For the community in Figure 4 (d), its topic is mainly about social, livelihood, or political issues. In particular, "deutsche welle" is a German international broadcaster which broadcasts news and information towards audiences outside of Germany. Different from the previous music communities, it talks about "female empowerment" and other topics like life, "sickness" and "the cure". In summary, these four communities carry their distinct attributes; by leveraging these attributes, we are able to explain and understand these communities.

Concluding remarks

We developed a novel semantic community identification method, SCI, to detect network community structures and infer their semantics simultaneously. A salient property of SCI is its ability to semantically or functionally annotate each of the communities identified. The key idea underlying SCI is to adequately integrate information of network topologies and information of node attributes under the framework of nonnegative matrix factorization (NMF). We formulated SCI as an optimization problem in NMF and designed efficient updating rules with a convergence guarantee. The extensive experimental results demonstrated the superior performance of SCI over several state-of-the-art approaches in accurately identifying network community structures. More importantly, it can effectively infer community semantics or attributes so as to explain and understand community structures.

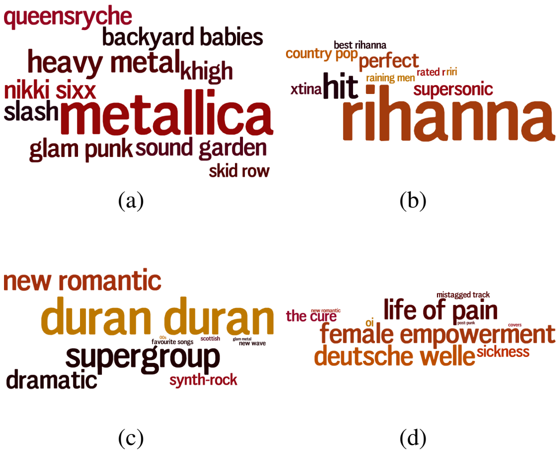

Figure 4: Word clouds for different communities. Top ten attributes of four communities are shown here. The size of a word is proportional to its community attribute value.

Appendix

A1. Proof of Theorem 1

At convergence, $\mathbf{U}^{(\infty)} = \mathbf{U}^{(t+1)} = \mathbf{U}^{(t)} = \mathbf{U}$, where t denotes the t-th iteration, i.e.,

$$U_{ij} = U_{ij}\left(\frac{(\mathbf{SC} + 2\beta\mathbf{AU} - \mathbf{U})_{ij}}{2\beta(\mathbf{UU}^T\mathbf{U})_{ij}}\right)^{\frac{1}{4}}, \tag{14}$$

which is equivalent to

$$(-2\mathbf{U} + 2\mathbf{SC} + 4\beta\mathbf{AU} - 4\beta\mathbf{UU}^T\mathbf{U})_{ij}U_{ij}^4 = 0, \tag{15}$$

which is equivalent to (8). □

A2. Proof of Lemma 2.

$$\begin{aligned} L(\mathbf{U}) = & \, tr(\mathbf{UU}^T - \mathbf{UC}^T\mathbf{S}^T - \mathbf{SCU}^T + \mathbf{SCC}^T\mathbf{S}^T) \\ & + \beta tr(\mathbf{AA} - 2\mathbf{AUU}^T + \mathbf{UU}^T\mathbf{UU}^T). \end{aligned} \tag{16}$$

[3] https://en.wikipedia.org/wiki/Supergroup_(music)

By Lemmas 6 and 7 of (Wang *et al.* 2011), we have

$$tr(\mathbf{U}\mathbf{U}^T\mathbf{U}\mathbf{U}^T) \leq tr(\mathbf{P}\mathbf{U'}^T\mathbf{U'}) \leq tr(\mathbf{R}\mathbf{U'}^T\mathbf{U'}\mathbf{U'}^T), \tag{17}$$

where $P_{ij} = \frac{(\mathbf{U}^T\mathbf{U})^2_{ij}}{(\mathbf{U'}^T\mathbf{U'})_{ij}}$, $R_{ij} = \frac{U^4_{ij}}{U'^3_{ij}}$.

By Lemma 4 of (Wang *et al.* 2011), we have

$$\begin{aligned} &- tr[(2\beta\mathbf{A} - \mathbf{I})\mathbf{U}\mathbf{U}^T] = -tr(\mathbf{A'}\mathbf{U}\mathbf{U}^T) \\ &\leq -tr(\mathbf{U'}^T\mathbf{A'}\mathbf{Z}) - tr(\mathbf{Z}^T\mathbf{A'}\mathbf{U}) - tr(\mathbf{U'}^T\mathbf{A'}\mathbf{U'}). \end{aligned} \tag{18}$$

By Lemma 2 of (Wang *et al.* 2011), we have

$$- tr(\mathbf{U}\mathbf{C}^T\mathbf{S}^T) \leq -tr(\mathbf{C}^T\mathbf{S}^T\mathbf{Z}) - tr(\mathbf{C}^T\mathbf{S}^T\mathbf{U'}). \tag{19}$$

For both (18) and (19), $Z_{ij} = U'_{ij} \ln \frac{U_{ij}}{U'_{ij}}$. By combining (17), (18) and (19), we have the final auxiliary function in Lemma 2. □

A3. Proof of Theorem 2.

Lemma 2 provides a specific form $Q(\mathbf{U}, \mathbf{U'})$ of the auxiliary function for $L(\mathbf{U})$ in problem (5). We can have the solution for $\min_{\mathbf{U}} Q(\mathbf{U}, \mathbf{U'})$ by the following KKT condition

$$\begin{aligned} \frac{\partial Q(\mathbf{U}, \mathbf{U'})}{\partial U_{ij}} &= 4\beta(\mathbf{U'}\mathbf{U'}^T\mathbf{U'})_{ij}\frac{U^3_{ij}}{U'^3_{ij}} \\ &- \frac{U'_{ij}}{U_{ij}}(2(\mathbf{A'}\mathbf{U'})_{ij} + 2(\mathbf{SC})_{ij}) = 0, \end{aligned} \tag{20}$$

which gives rise to the updating rule in (9). Following Lemma 1, under this updating rule the objective function $L(\mathbf{U})$ of (5) will be nonincreasing. □

Acknowledgments. This work was supported by National High-tech R&D Program of China (2014BAK11B03), National Basic Research Program of China (2013CB329305), National Natural Science Foundation of China (No. 61422213, 61503281, 61502334, 61303110), "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDA06010701), Foundation for the Young Scholars by Tianjin University of Commerce (150113), National Training Programs of Innovation and Entrepreneurship for Undergraduates (201410069040), the Talent Development Program of Wuhan, the municipal government of Wuhan, Hubei, China (2014070504020241), and an internal research grant of Jianghan University, Wuhan, China, the United States National Institutes of Health (R01GM100364) and United States National Science Foundation (DBI-0743797).

References

V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

M. Girvan and M.E.J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

Han Hu, Zhouchen Lin, Jianjiang Feng, and Jie Zhou. Smooth representation clustering. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3834–3841. IEEE, 2014.

David R Hunter and Kenneth Lange. A tutorial on mm algorithms. *The American Statistician*, 58(1):30–37, 2004.

B. Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

Jingu Kim and Haesun Park. Sparse nonnegative matrix factorization for clustering. 2008.

Andrea Lancichinetti, Santo Fortunato, and János Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015, 2009.

Haifeng Liu, Zhaohui Wu, Xuelong Li, Deng Cai, and Thomas S Huang. Constrained nonnegative matrix factorization for image representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(7):1299–1311, 2012.

Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 977–986. ACM, 2014.

Yulong Pei, Nilanjan Chakraborty, and Katia Sycara. Nonnegative matrix tri-factorization with graph regularization for community detection in social networks. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2083–2089. AAAI Press, 2015.

Simon Pool, Francesco Bonchi, and Matthijs van Leeuwen. Description-driven community detection. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(2):28, 2014.

Guo-Jun Qi, Charu C Aggarwal, and Thomas Huang. Community detection with edge content in social media networks. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pages 534–545. IEEE, 2012.

Yiye Ruan, David Fuhry, and Srinivasan Parthasarathy. Efficient community detection in large networks using content and links. In *Proceedings of the 22nd international conference on world wide web*, pages 1089–1098. International World Wide Web Conferences Steering Committee, 2013.

D Seung and L Lee. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13:556–562, 2001.

F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493–521, 2011.

Jierui Xie, Stephen Kelley, and Boleslaw K Szymanski. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (CSUR)*, 45(4):43, 2013.

Jaewon Yang and Jure Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM, 2013.

Tianbao Yang, Rong Jin, Yun Chi, and Shenghuo Zhu. Combining link and content for community detection: a discriminative approach. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 927–936. ACM, 2009.

Jaewon Yang, Julian McAuley, and Jure Leskovec. Community detection in networks with node attributes. In *Data Mining (ICDM), 2013 IEEE 13th international conference on*, pages 1151–1156. IEEE, 2013.

Liang Yang, Xiaochun Cao, Di Jin, Xiao Wang, and Dai Meng. A unified semi-supervised community detection framework using latent space graph regularization. 2014.