Identifying Interpretable Link Communities with User Interactions and Messages in Social Networks

Wei Li[†,‡,#], Meng Qin[†,‡,#], Kai Lei[†,‡,*]
[†]ICNLAB, School of Electronics and Computer Engineering (SECE), Peking University, Shenzhen, China
[‡]PCL Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen, China
1701213607@sz.pku.edu.cn, mengqin_az@foxmail.com, leik@pkusz.edu.cn
[#]Equal Contributions, [*]Corresponding Author

***Abstract*—Community detection is a fundamental step in social network analysis. In this study, we focus on the hybrid identification of overlapping communities with network topology (e.g., user interaction) and edge-induced content (e.g., user messages). Conventional hybrid methods tend to combine topology and node-induced content to explore disjoint communities, but fail to integrate edge-induced content and to derive overlapping communities. Moreover, although several semantic community detection approaches have been developed, which can generate the semantic description for each community besides the partition result, they can only incorporate the word-level content to derived coarse-grained descriptions. We propose a novel Semantic Link Community Detection (SLCD) model based on non-negative matrix factorization (NMF). It adopts the perspective of link communities to integrate topology and edge-induced content, where the partitioned link communities can be further transformed into an overlapping node-induced membership. Moreover, SLCD incorporates both the word-level and sentence-level content while considering the diversity of user messages. Hence, it can also derive strongly interpretable and multi-view community descriptions simultaneously when community partition is finished. SLCD's superior performance and powerful application ability over other state-of-the-art competitors are further demonstrated in our experiments on a series of real social networks.**

***Index Terms*—Social Networks Analysis, Community Detection, Semantic Description, Non-negative Matrix Factorization**

I. Introduction

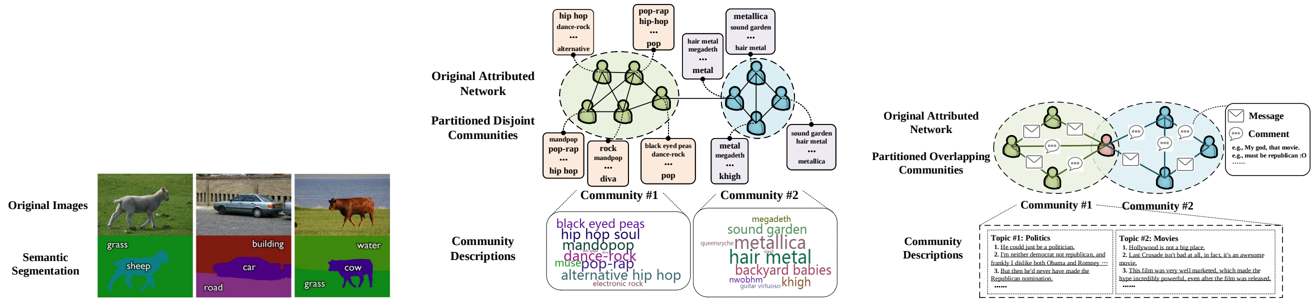
As a typical complex system, a specific social network can be generally formulated as a corresponding graph. In this graph, one can abstract each user as a node and users' interactions as the edges among nodes. Community is a significant substructure in most social networks, which usually corresponds to the real user groups with similar properties (e.g., interaction or semantic). According to [1], a typical community can be described as a group of nodes, within which the linage is dense (but between which is relatively sparse). As discussed in [2]–[4], it's also believed that the identification of communities can help to reveal the network's structure, function and semantic, while supporting the advanced applications of user profiling [5], recommender systems [6], etc. Hence, community detection can be considered as one of the fundamental steps in social network analysis.

Conventional community detection approaches (e.g., [1], [7]–[9]) tend to explore the latent characteristics of network topology (e.g., user interactions in social networks) to partition the community structures. Besides topology, content information (e.g., user's profile features) is another significant source available for most social network systems. It has also been verified that the network content can provide orthogonal and complementary information (about the network's significant properties) beyond the topological structures [3]. Hence, the integration of both topology and content can potentially lead to better community partition compared with those considering topology alone. Based on this motivation, several state-of-the-art methods (e.g., [10]–[12]) have tried to incorporate these two sources and have achieved significant improvement.

Inspired by the *semantic segmentation* task (see Fig. 1 (a)) from the field of computer vision (CV), the concept of *semantic community detection* (see Fig. 1 (b)) is proposed by previous work [2]–[4]. As illustrated in Fig. 1 (a), given an image, the *semantic segmentation task* is required to simultaneously partition the image into several different parts while annotating the semantic of each part (e.g., in terms of keywords). According to Fig. 1 (b), given an attributed network (e.g., with node-induced discrete attribute), a *semantic community detection* method can simultaneously derive the community membership while generating the semantic description (e.g., wordcloud) for each community. Note that such description is automatically generated without any additional manual inspection, which can reveal the network's deep semantic and knowledge in an unsupervised manner. In this case, the network content can not only improve the network's community partition, but also interpret the semantic of the partition result.

Despite the effectiveness of state-of-the-art hybrid methods with network topology and content (including existing *semantic community detection* approaches), there remain the following primary limitations.

First, most existing hybrid methods only incorporate the node-induced content (e.g., user's interest tags) to obtain the disjoint node communities (see Fig. 1 (b)). In this case, each node must belong to one unique community. However, they may fail to integrate the edge-induced content and to derive the overlapping communities. On the one hand, user message (e.g., email or comment) is common edge-induced content generated via user interactions in most social networks. On the other hand, overlapping communities, which allow each node to belong to multiple communities, are also ubiquitous in reality [8]. Moreover, the diversity of edge-induced message is also



(a) Semantic Segmentation (b) Existing Semantic Community Detection (c) Semantic Link Community Detection (SLCD)

Fig. 1. Illustrations of (a) *Semantic Segmentation*, (b) *Existing Semantic Community Detection* and (c) *Semantic Link Community Detection*. The concept of semantic community detection is primarily inspired from the semantic segmentation task in CV field. Existing semantic community detection methods tend to explore the word-level node-induced content and can only derive the disjoint community structure with coarse-grained descriptions. SLCD can obtain overlapping community structures while generating strongly interpretable and multi-view community descriptions.

a significant reason that leads to the overlapping membership in social networks. For instance, users may choose different communities with diverse interests or backgrounds, so they tend to send distinct messages to others in different communities. Existing methods may fail to capture such diversity as well as its induced overlapping community structures.

Second, existing *semantic community detection* methods tend to only use the word-level content (e.g., interest tags) to generate the interpretable descriptions (see Fig. 1 (b)). Such descriptions are coarse-grained, where (i) each community only has only one comprehensive description and (ii) each description is in the form of representative keywords. In fact, user messages in social networks usually contain complete sentences, which carry additional information about the organization of the word-level content. The ignored sentence-level content can be used to further enhance the interpretability of community descriptions. Moreover, user messages are also diverse, the unique description of each community may also fail to reflect such diversity.

To alleviate the aforementioned limitations, we proposed a novel Semantic Link Community Detection (SLCD) method based on the non-negative matrix factorization (NMF) [13] framework. Fig. 1 (c) demonstrates a brief sketch of SLCD. Compared with most existing hybrid methods combing network topology and content, SLCD's superiority is three-fold, which also forms our main contributions of this study.

- First, SLCD integrates the edge-induced content and derives the overlapping community membership based on the link communities. Particularly, we first partition edges (rather than nodes) into several subsets (i.e., link communities) with dense linkage and similar content. The link communities can be further transformed into a corresponding node-induced overlapping membership, which is treated as the partition result of SLCD.
- We consider both word-level as well as sentence-level content and use the *view of separated clustering membership* to integrate network topology and content. It enables SLCD to generate strongly interpretable and multi-view community descriptions. Concretely, each community can have multiple descriptions with each one reflecting one dominant topic. Moreover, each community description is in the form of representative sentences with much stronger interpretability.
- According to our experiments on a series of real social networks, SLCD consistently outperforms other competitors, while having better ability to reveal the network's deep semantic and knowledge.

In the rest of this paper, we first introduce the related work in Section II. Then, we give the formal problem definitions about (i) link communities, (ii) node overlapping communities and (iii) *separated clustering membership view* in Section III. The proposed SLCD model is elaborated in Section IV, where we in sequence derive the unified optimization objective and the solving strategy. Extensive experiments are further introduced in Section V, including the (i) performance evaluation of overlapping community detection and (ii) case study regarding semantic descriptions. Section VI concludes this paper indicates our future work.

II. Related Work

In the past few years, a series of approaches have been proposed for the community detection task. Several comprehensive reviews can be found in [14].

Conventional community identification methods tend to fully explore the latent characteristics of network topology. For instance, authors of [1] formulated the community detection task as a graph-cut problem based on the defined edge betweenness. [9] introduced an overlapping community detection method by ranking the node popularities, while [7] proposed a stochastic overlapping identification model based on the link clusters. In [8], both the partitions of nodes and edges were utilized to identify overlapping communities.

A series of state-of-the-art approaches have tried to combine both network topology and content for the sake of better partition performance. In [12], the integration of the complementary content information was formulated as a graph regularization on the conventional topology-based NMF model. [10] studied the overlapping communities in social circles by considering both circle structure and user profile similarity, while [11] explored the overlapping membership by combing edge structure and node attributes. Furthermore, authors of [2] developed an NMF-based *semantic community detection approach* to simultaneously derive the partition result and community descriptions.

Different from the aforementioned related work, our main focus is to (i) integrate the edge-induced content to explore overlapping community membership and (ii) generate strongly interpretable and multi-view community descriptions.

III. Problem Definition

We formulate the social network as an attributed graph (network). User interactions and messages are abstracted as the undirected unweighted topology and edge-induced content (with both the level of words and sentences considered).

Generally, a specific attributed network can be represented as a 6-tuple $G = (V, E, A, F_A, C, F_C)$, where $V = \{v_1, \cdots, v_M\}$ and $E = \{e_l(v_i, v_j) \,|v_i, v_j \in V\}$ are the set of nodes and edges; $A = \{a_1, \cdots, a_W\}$ is the set of keywords (terms); $C = \{c_1, \cdots, c_S\}$ is the set of sentences, with each sentence $c_s \subset A$ as a sequence of keywords; $F_A = \{F_A(e_l) \,|e_l(v_i, v_j) \in E\}$ and $F_C = \{F_C(e_l) \,|e_l(v_i, v_j) \in E\}$ are respectively the map from V to W and from V to C, with $F_A(e_l)$ and $F_C(e_l)$ as e_l's word set and sentence set.

Identification of Link Communities. Given the attributed network G, the goal of the hybrid link-based community detection is to partition the edge set E into K subsets (i.e., communities) $R = \{R_1, \cdots, R_K\}$ (with $R_r \subset E$) according to the topology E and content $\{F_A, F_C\}$, so that:

(i) within each community the linkage is dense and the content (i.e., keywords or sentences) has similar semantic;

(ii) between different communities the linkage is relatively loose and the content is distinct.

We assume that the link-based community structure is disjoint. Namely, $R_r \cap R_t = \emptyset$ holds for any $r \neq t$.

Transforming to Node-Induced Communities. The above link (edge-induced) community structure can be further transformed into a corresponding node-induced form. Concretely, if edge (v_i, v_j) belongs to the community R_r, its induced nodes v_i and v_j are simultaneously assigned with the label r. Such transformation is equivalent to partition the node set V into K subsets $R' = \{R'_1, \cdots, R'_K\}$ with dense linkage and similar semantic, in which $R'_r \subset V$ is the set of community R'_r's node members. Note that the transformed node-induced community structure should be overlapping. In other words, there exist $r \neq t$ that satisfies $R'_r \cap R'_t \neq \emptyset$.

How to extract the node-induced overlapping communities, which correspond to the groups and organizations in real social networks, is our main focus. Usually, better correspondence indicates better performance of community detection.

Correlation between Topology and Content. We adopt the *view of separated clustering membership* to explore the *intrinsic correlation* between the heterogeneous network topology and content. It can help to derive the multi-view community descriptions while enhancing the model's robustness.

Given an attributed network G, we can partition E into (i) K_1 topology clusters $R = \{R_1, \cdots, R_{K_1}\}$ and (ii) K_2 content clusters $L = \{L_1, \cdots, L_{K_2}\}$ only based on the (i) topology E and (ii) content $\{F_A, F_C\}$. Note that both R and L are in the edge-induced disjoint form. In this case, each edge e_l is simultaneously assigned with (i) a topology cluster label and (ii) a content cluster label. The *intrinsic correlation* can then be formulated as the correspondence between R and L. Usually, the larger proportion of the edges that share the similar clustering memberships described by R and L indicates better correspondence.

IV. Methodology

A. Modeling the Network Topology

For a given network G with M nodes and N edges, we assume the number of topology clusters extracted only based on G's topology is K_1. We utilize the *link structure matrix* $\mathbf{B} \in \Re^{N \times M}$ to describe the topology, in which $\mathbf{B}_{li} = 1$ if node v_i is the end node of edge e_l and $\mathbf{B}_{li} = 0$ otherwise. To further describe the clustering membership, we introduce the *link membership matrix* $\mathbf{X} \in \Re^{N \times K_1}$ and *node membership matrix* $\mathbf{Y} \in \Re^{M \times K_1}$. $\mathbf{X}_{lr}$ represents the propensity that edge e_l belongs to cluster R_r, while $\mathbf{Y}_{ir}$ is defined as the propensity that node v_i is the member of cluster R'_r.

From the perspective of generative model, $\sum_{r=1}^{K_1} \mathbf{X}_{lr}\mathbf{Y}_{ir}$ represents the expectation that v_j is the end vertex of e_i, which should be as close as possible to the real value $\mathbf{B}_{li}$. Hence, we can derive the following NMF-based optimization objective only related to network topology:

$$\mathop{\arg\min}_{\mathbf{X},\mathbf{Y}} \left\| \mathbf{B} - \mathbf{X}\mathbf{Y}^T \right\|_F^2 \text{ s.t. } \mathbf{X} \geq 0, \mathbf{Y} \geq 0. \qquad (1)$$

When the solution of this problem (notated as $\{\mathbf{X}^*, \mathbf{Y}^*\}$) is obtained, we can utilize $\mathbf{X}^*$ to extract the link communities, with $\mathbf{Y}^*$ as the auxiliary variable. Concretely, we assign the topology cluster label of edge e_l to be the column index with maximum propensity value in the l-th row of $\mathbf{X}^*$.

B. Modeling the Edge-Induced Content

We use the topic-enhanced multi-document summarization technique [15] to formulate network content. The basic model is derived via a probabilistic summarization model.

Given a network G, each edge e_l can be treated as a document with word-level and sentence-level content (i.e., $F_A(e_l)$ and $F_C(e_l)$). We adopt the unigram (Bag-of-Word) language model to describe the word-level content, where $p(a_w \,|e_l)$ is the probability that a_w is in e_l's attribute set (i.e., $a_w \in F_A(e_l)$). We further introduce the latent variables $c_s \in C$ and $L_t \in L$ to represent sentence and topic (content cluster), respectively. The observable distribution $p(a_w \,|e_l)$ can be decomposed into the following marginalized form:

$$p(a_w \,|e_l) = \sum_{L_t \in L} \sum_{c_s \in C} p(a_w \,|c_s) p(c_s \,|L_t) p(L_t \,|e_l), \qquad (2)$$

where $p(a_w \,|c_s)$ is also the observable distribution; $p(c_s \,|L_t)$ and $p(L_t \,|e_l)$ are the distributions needed to be inferred.

We further transform (2) into another equivalent NMF-based objective. For an attribute network with N edges and W words, we introduce the *link-word matrix* $\mathbf{C} \in \Re^{N \times W}$, where $\mathbf{C}_{lw} = 1$ when $a_w \in F_A(e_l)$ and $\mathbf{C}_{lw} = 0$, otherwise. Hence, we have $\mathbf{C}_{lw} = p(a_w \,|e_l\,)$.

Assume the number of attribute clusters is K_2. To describe the cluster membership with sentence-level content, we introduce the (i) *content membership matrix* $\mathbf{Z} \in \Re^{N \times K_2}$, (ii) *sentence description matrix* $\mathbf{U} \in \Re^{S \times K_2}$ and (iii) *content relation matrix* $\mathbf{V} \in \Re^{S \times W}$. Particularly, $\mathbf{Z}_{lt}$ represents the propensity that link e_l belongs to content cluster L_t. $\mathbf{U}_{st}$ is the propensity that L_t can be described by sentence C_s. Furthermore, $\mathbf{V}_{sw} = 1/|c_s|$ when $a_w \in c_s$ and $\mathbf{V}_{sw} = 0$, otherwise. In this case, we can use $\mathbf{Z}_{lt}$ and $\mathbf{U}_{st}$ to fit $p(L_t \,|e_l\,)$ and $p(c_s \,|L_t\,)$. We also have $\mathbf{V}_{sw} = p(a_w \,|c_s\,)$.

Based on the definitions of $\{\mathbf{C}, \mathbf{Z}, \mathbf{U}, \mathbf{V}\}$, we finally transform (2) into the following equivalent NMF problem:

$$\mathop{\arg\min}_{\mathbf{Z},\mathbf{U}} \left\|\mathbf{C} - \mathbf{Z}\mathbf{U}^T\mathbf{V}\right\|_F^2 \text{ s.t. } \mathbf{Z} \geq 0, \mathbf{U} \geq 0. \quad (3)$$

As discussed in Section III, the semantic in a certain topic (content cluster) should be relatively similar, but the semantic among different topics should be distinct. To enhance such property, we introduce the sparse constraint on each column of $\mathbf{U}$ by following [16]:

$$\mathop{\arg\min}_{\mathbf{Z},\mathbf{U}} \left\|\mathbf{C} - \mathbf{Z}\mathbf{U}^T\mathbf{V}\right\|_F^2 + \beta \sum_{t=1}^{K_2} \left\|\mathbf{U}_{:,t}\right\|_1^2 \text{ s.t. } \mathbf{Z} \geq 0, \mathbf{U} \geq 0, \quad (4)$$

with β as the parameter to control the second sparse term. Moreover, (4) can be further transformed into the following simplified form:

$$\mathop{\arg\min}_{\mathbf{Z},\mathbf{U}} \left\|\mathbf{C} - \mathbf{Z}\mathbf{U}^T\mathbf{V}\right\|_F^2 + \beta \left\|\mathbf{e}\mathbf{U}\right\|_F^2 \text{ s.t. } \mathbf{Z} \geq 0, \mathbf{U} \geq 0, \quad (5)$$

where $\mathbf{e} \in \Re^{1 \times K_1}$ is a K_1-dimensional vector with all the elements equal to one.

Similar to (1), when the solution of (5) (notated as $\{\mathbf{Z}^*, \mathbf{U}^*\}$) is obtained, one can utilize $\mathbf{Z}^*$ to extract the content cluster membership by exploring each row of $\mathbf{Z}^*$. $\mathbf{U}^*$ can be used to generate the summarization of each topic by selecting the top representative sentences from each column of $\mathbf{U}^*$.

C. The Unified Model

As discussed in III, we adopt the *view of separated clustering structure* to explore the intrinsic correlation between the heterogeneous network topology and content (with K_1 topology clusters and K_2 content clusters). Inspired by the Markov transition process, we introduce another attention-like *membership transition matrix* $\mathbf{R} \in \Re^{K_1 \times K_2}$. Concretely, $\mathbf{R}_{rt}$ is defined as the transition propensity from topology cluster R_r to content cluster L_t. Hence, there exist the relation between $\mathbf{X}$ and $\mathbf{Z}$ that $\mathbf{Z}_{lt} = \sum_{r=1}^{K_1} \mathbf{X}_{lr}\mathbf{R}_{rt}$.

As a result, we can formulate the unified objective of the proposed Semantic Link Community Detection (SLCD) model by combing (1) and (5):

$$\begin{array}{l} \mathop{\arg\min}\limits_{\mathbf{X},\mathbf{Y},\mathbf{U},\mathbf{R}} \left\|\mathbf{B} - \mathbf{X}\mathbf{Y}^T\right\|_F^2 + \alpha \left\|\mathbf{C} - \mathbf{X}\mathbf{R}\mathbf{U}^T\mathbf{V}\right\|_F^2 + \beta \left\|\mathbf{e}\mathbf{U}\right\|_F^2 \\ \text{s.t. } \mathbf{X} \geq 0, \mathbf{Y} \geq 0, \mathbf{R} \geq 0, \mathbf{U} \geq 0 \end{array}, \quad (6)$$

where we replace $\mathbf{Z}$ with $\mathbf{XR}$; α and β are the parameters to adjust the contributions of the first term (regarding content) and second term (regarding the sparse constraint).

Note that the effect of $\mathbf{R}$ is two-fold.

First, it incorporates the heterogeneous network topology and content into one unified objective, which can improve the community partition by using the complementary information provided by content.

Second, $\mathbf{R}$ also encodes the correspondence relation (i.e., intrinsic correlation) between topology clusters and content clusters. Concretely, if there is an explicit correspondence relation embedded in $\mathbf{R}$, each column (or row) of $\mathbf{R}$ has one or more dominant elements with significantly larger transition propensities than others. On the other hand, if there is an indistinct correspondence relation in $\mathbf{R}$, all the elements in a certain column (or row) of $\mathbf{R}$ may have close transition propensities. Such encoded correlation can be utilized to extract the multi-view fine-grained semantic description for each community. Namely, we can find one or more dominant topics (content clusters) for each topology cluster (by exploring each row of $\mathbf{R}$), where each topic has its own semantic description, reflecting one particular view of the network.

In summary, for the unified model (6), we treat the topology clusters as the final community structures, while utilizing the content clusters to extract the corresponding semantic descriptions. For the solution of (6) (notated as $\{\mathbf{X}^*, \mathbf{Y}^*, \mathbf{R}^*, \mathbf{U}^*\}$), we first use $\mathbf{X}^*$ to derive the link communities, which can be further transformed into the corresponding node-induced overlapping form. Concretely, as discussed in Section III, if edge $e_l(v_i, v_j)$ belongs to link community R_r, its induced nodes $\{v_i, v_j\}$ are assigned to be the member of node-induced community R_r'. Moreover, we utilize $\mathbf{R}^*$ to extract the dominant topics (content clusters) of each community (topology cluster) by exploring each row of $\mathbf{R}^*$. Note that each community may have multiple dominant topics. Finally, the semantic description of each topic can be generated by selecting the top sentences from each column of $\mathbf{U}^*$.

D. Model Optimization

We adopt the block coordinate descent method to solve the non-convex NMF problem (6), where we first properly initialize $\{\mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{U}\}$ and in turns take the following 4 steps to continuously update their values until converge.

1) $\mathbf{X}$-Step: In this step, we update $\mathbf{X}$ with $\{\mathbf{Y}, \mathbf{R}, \mathbf{U}\}$ fixed. To derive the updating rule, we first extract the terms only related to $\mathbf{X}$ in (6) and formulate the following objective:

$$\mathrm{O}_{\mathbf{X}}(\mathbf{X}) = \left\|\mathbf{B} - \mathbf{X}\mathbf{Y}^T\right\|_F^2 + \alpha \left\|\mathbf{C} - \mathbf{X}\mathbf{R}\mathbf{U}^T\mathbf{V}\right\|_F^2. \quad (7)$$

By applying $\left\|\mathbf{M}\right\|_F^2 = \mathrm{tr}(\mathbf{M}\mathbf{M}^T)$, we can get the partial derivative of $\mathrm{O}_{\mathbf{X}}(\mathbf{X})$ with respect to $\mathbf{X}$:

$$\frac{\partial \mathrm{O}_{\mathbf{X}}(\mathbf{X})}{\partial \mathbf{X}} = 2(\mathbf{X}\mathbf{Y}^T\mathbf{Y} - \mathbf{B}\mathbf{Y}) + 2\alpha(\mathbf{X}\mathbf{L}\mathbf{L}^T - \mathbf{C}\mathbf{L}^T), \quad (8)$$

where $\mathbf{L} = \mathbf{R}\mathbf{U}^T\mathbf{V}$.

According to the standard gradient descent process, we have the following addictive updating rule regarding $\mathbf{X}$:

$$\mathbf{X}_{lr} \leftarrow \mathbf{X}_{lr} - \eta_{lr}(\nabla_{+} - \nabla_{-})_{lr}, \tag{9}$$

with η_{lr} as the learning rate. In (9), we utilize the simplified notations ∇_{+} and ∇_{-} to represent the terms with positive coefficients and with negative coefficients, respectively. Namely, $\nabla_{+} = 2(\mathbf{X}\mathbf{Y}^T\mathbf{Y} + \alpha\mathbf{X}\mathbf{L}\mathbf{L}^T)$ and $\nabla_{-} = 2(\mathbf{B}\mathbf{Y} + \alpha\mathbf{C}\mathbf{L}^T)$.

If we properly set $\eta_{ir} = \mathbf{X}_{ir}/(\nabla_{+})_{ir}$, we can transform (9) into the following multiplicative form:

$$\mathbf{X}_{lr} \leftarrow \mathbf{X}_{lr}\frac{(\nabla_{-})_{lr}}{(\nabla_{+})_{lr}} = \mathbf{X}_{lr}\frac{(\mathbf{B}\mathbf{Y} + \alpha\mathbf{C}\mathbf{L}^T)_{lr}}{(\mathbf{X}(\mathbf{Y}^T\mathbf{Y} + \alpha\mathbf{L}\mathbf{L}^T))_{lr}}. \tag{10}$$

The above solving strategy is effective for most NMF-based models. It has been proved by [17] that if the variables (e.g., $\{\mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{U}\}$ in (6)) are initialized as non-negative values, the non-negative constraints of NMF (e.g., $\mathbf{X} \geq 0$) can be ensured during the iterative updating process. Moreover, the convergence of the solving strategy can also be guaranteed since it's based on the gradient descent process. Note that the multiplicative updating rule (e.g., (10)) has significant advantages beyond the gradient descent method with the addictive rule (e.g., (9)). The multiplicative rule eliminates the manual settings of the learning rate in addictive rule (e.g., η_{ir} in (9)). It can also be considered as an adaptive adjustment of the learning rate, which helps the updating process to converge much faster than the standard gradient descent. Hence, in the rest of this paper, we adopt the same strategy to derive the updating rules of other variables (i.e., $\{\mathbf{Y}, \mathbf{R}, \mathbf{U}\}$).

2) $\mathbf{Y}$-Step: In the $\mathbf{Y}$-step, we update $\mathbf{Y}$ with $\{\mathbf{X}, \mathbf{R}, \mathbf{U}\}$ fixed. The derivation of the updating rule is similar to that of (10). Due to the space limit, we omit the derivation details and directly list it as follow:

$$\mathbf{Y}_{ir} \leftarrow \mathbf{Y}_{ir}(\mathbf{B}^T\mathbf{X})_{ir}/(\mathbf{Y}\mathbf{X}^T\mathbf{X})_{ir}. \tag{11}$$

3) $\mathbf{R}$-Step: In the $\mathbf{R}$-Step, we fix $\{\mathbf{X}, \mathbf{Y}, \mathbf{U}\}$ and update $\mathbf{R}$. Similar to (10), we can obtain the following updating rule:

$$\mathbf{R}_{rt} \leftarrow \mathbf{R}_{rt}(\mathbf{X}^T\mathbf{C}\mathbf{T}^T)_{rt}/(\mathbf{X}^T\mathbf{X}\mathbf{R}\mathbf{T}\mathbf{T}^T)_{rt}, \tag{12}$$

with $\mathbf{T} = \mathbf{U}^T\mathbf{V}$.

4) $\mathbf{U}$-Step: The $\mathbf{U}$-step is applied to update $\mathbf{U}$ via the following rule (with $\{\mathbf{X}, \mathbf{Y}, \mathbf{R}\}$ fixed):

$$\mathbf{U}_{st} \leftarrow \mathbf{U}_{st}(\alpha\mathbf{V}\mathbf{C}^T\mathbf{P})_{st}/(\alpha\mathbf{V}\mathbf{V}^T\mathbf{U}\mathbf{P}^T\mathbf{P} + \beta\mathbf{E}\mathbf{U})_{st}, \tag{13}$$

where $\mathbf{P} = \mathbf{X}\mathbf{R}$ and $\mathbf{E} = \mathbf{e}^T\mathbf{e}$.

We utilized the NNDSVD strategy [18] to initialize $\{\mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{U}\}$. This strategy is originally designed for the standard NMF problem that $\min \left\|\mathbf{M} - \mathbf{W}\mathbf{H}^T\right\|_F^2$. For the convenience of discussion, we use the notation $\{\mathbf{W}^{(0)}, \mathbf{H}^{(0)}\} = \mathrm{N}(\left\|\mathbf{M} - \mathbf{W}\mathbf{H}^T\right\|_F^2)$ to represent the initialization of $\{\mathbf{W}, \mathbf{H}\}$ in the objective $\left\|\mathbf{M} - \mathbf{W}\mathbf{H}^T\right\|_F^2$. For the detailed algorithm of NNDSVD, please refer to [18].

In SLCD, we first initialize $\{\mathbf{X}, \mathbf{Y}\}$ by setting that

$$\{\mathbf{X}^{(0)}, \mathbf{Y}^{(0)}\} = \mathrm{N}(\left\|\mathbf{B} - \mathbf{X}\mathbf{Y}^T\right\|_F^2). \tag{14}$$

To initialize $\mathbf{U}$, we introduce an auxiliary *link-sentence matrix* $\mathbf{S} \in \Re^{N \times S}$, where $\mathbf{S}_{ls} = 1$ if sentence c_s is in edge e_l's sentence set $F_C(e_l)$ and $\mathbf{S}_{ls} = 0$, otherwise. Then, $\mathbf{U}$'s initialized value is set via

$$\{\mathbf{Z}^{(0)}, \mathbf{U}^{(0)}\} = \mathrm{N}(\left\|\mathbf{S} - \mathbf{Z}\mathbf{U}^T\right\|_F^2), \tag{15}$$

where $\mathbf{Z}$ is the auxiliary variable for the further initialization of $\mathbf{R}$. Finally, we adopt the following strategy to initialize $\mathbf{R}$:

$$\{:, \mathbf{R}^{(0)}\} = \mathrm{N}(\left\|\mathbf{Z}^{(0)} - \mathbf{X}'\mathbf{R}\right\|_F^2). \tag{16}$$

In conclusion, we summarized the aforementioned process in Algorithm 1. By reasonably assuming $\max\{K_1, K_2\} \ll \min\{M, N, W, S\}$, the time complexities to update $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{R}$ and $\mathbf{U}$ once are $O(NMK_1 + WSK_2)$, $O(NMK_1)$, $O(WSK_2 + NWK_2)$ and $O(NWK_2 + WSK_2)$, respectively. Hence, the complexity of the SLCD algorithm is no more than $O(T(NMK_1 + WSK_2 + NWK_2))$ for T iterations to converge. Please note that such complexity is only the upper bound of SLCD. It can be further reduced by considering the sparsity of network topology and content. Therefore, all the variables $\{\mathbf{B}, \mathbf{C}, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{U}\}$ are treated as sparse matrices in our implementation, which has much lower complexity.

To determine whether the iterative updating process has converged, we adopt a criterion based on the objective function's value (i.e., (6)'s value) . In each iteration, we record (6)'s value and calculate its relative error with respect to the previous iteration. When such relative error is smaller than a pre-set threshold (e.g., 10^{-6} in our experiments), we determine the algorithm has converged and stop the updating.

Algorithm 1: Semantic Link Community Detection

```
Input: B, C, S, V, $\{K_1, K_2\}$, $\{\alpha, \beta\}$
Output: $\mathbf{X}^*$, $\mathbf{Y}^*$, $\mathbf{R}^*$, $\mathbf{U}^*$
initialize $\{\mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{U}\}$ via (14), (15) and (16)
while not converge do
    update $\mathbf{X}$ via (10) with $\{\mathbf{Y}, \mathbf{U}, \mathbf{R}\}$ fixed //$\mathbf{X}$-Step
    update $\mathbf{Y}$ via (11) with $\{\mathbf{X}, \mathbf{U}, \mathbf{R}\}$ fixed //$\mathbf{Y}$-Step
    update $\mathbf{R}$ via (12) with $\{\mathbf{X}, \mathbf{Y}, \mathbf{U}\}$ fixed //$\mathbf{R}$-Step
    update $\mathbf{U}$ via (13) with $\{\mathbf{X}, \mathbf{Y}, \mathbf{R}\}$ fixed //$\mathbf{U}$-Step
extract the (disjoint) link community membership via $\mathbf{X}^*$
transform the link membership into the overlapping node
    membership
explore the correlation between topology and content clusters via $\mathbf{R}^*$
extract each content cluster's description via $\mathbf{U}^*$
```

V. Experimental Evaluation

A. Datasets

In our experiments, *Enron*[1] (*EN*) [19] and *Reddit*[2] [20] were used as the testing datasets.

Enron is a labeled subnetwork extracted from the email system of Enron corporation. In this dataset, each email contains the addresses of the related sender and the receiver(s) (i.e., an email can be sent to multiple users) as well as the text content. We abstracted each user as a node and recorded a certain link

[1] http://bailando.sims.berkeley.edu/enron_email.html

[2] https://sites.google.com/site/sunnycdwhl/home

(v_i, v_j) if user v_i sent an email to v_j. The email text was considered as the edge-induced content. For each email, we extracted all the sentences and words from the text. We further removed the stop words and conducted the stemming process to construct the word dictionary. Moreover, all the emails were manually labeled by the dataset's providers with 13 classes (e.g., company image, legal advice, etc.) according to their content. We adopted the edge-induced labels as the ground-truth and transformed them into the node-induced form, which described an overlapping community structure.

Reddit is a dataset extracted from 3 sub-forums (i.e., Movies, Politics and Science) in the social news aggregation and discussion website Reddit[3]. It contains the interactive records of user comment from 2012-8-25 to 2012-8-31. During pre-processing, we treated each user as a node. If user v_i commented on user v_j, we extracted a corresponding link (v_i, v_j). The comment text was processed similarly with *Enron*. The subforums in which a user posted comments were considered as the node-induced ground-truth, which was found to be overlapping. Namely, there exist users that have posted content in multiple sub-forums. Furthermore, to have more datasets to be evaluated, we extracted 3 timeslices (2018-8-25, 2018-8-26 and 2018-8-27) of *Reddit*, forming 3 subsets notated as *Reddit25* (*R25*), *Reddit26* (*R26*) and *Reddit27* (*R27*).

The statistic details of the 4 datasets after pre-processing are presented in Table I, with M, N, S, W and C as the number of nodes, edges, sentences, words (terms) and communities.

TABLE I
STATISTIC DETAILS OF THE TESTING DATASETS

		M	*N*	*S*	*W*	*C*
Reddit27	**R27**	2,143	2,290	5,794	6,635	3
Reddit26	**R26**	1,590	1,714	3,952	5,055	3
Reddit25	**R25**	1,314	1,339	3,273	4,616	3
Enron	**EN**	974	1,557	31,563	15,382	13

B. Performance Evaluation

In the experiment, we compared the performance of SLCD with 12 baselines, which can be divided into 3 types.

- First, *LMBP* [7], *BigCLAM* [21] and *CNLP* [8] are state-of-the-art overlapping community detection methods only consider network topology.
- Second, *SMR* [22] and *LDA* [23] are adopted as two typical content-based approaches, in which we treat each edge as a corresponding document with text. Particularly, these content-based methods partition the links according to their content, forming a disjoint edge-induced clustering structure. Similar to the link communities, the edge-induced memberships are further transformed into the overlapping node-induced forms.
- Third, *Circles* [10], *EIMF-LAP* [17], *EIMF-LP* [17], *CESNA* [11] and *DNEM* [24] are all the hybrid overlapping community detection methods integrating network topology and attribute.

Moreover, the basic model (1) and (5) are also selected as the topology-based and content-based baseline (notated as *NMF-T* and *NMF-C*). They can further verify SLCD's performance improvement by incorporating the two sources.

As the datasets all provide the ground-truth, we follow [21], [24] to adopt the following generalized metric to evaluate the accuracy of the extracted overlapping communities:

$$\tfrac{1}{2|R|}\sum_{R_r \in R} \max_{G_t \in G} f(R_r, G_t) + \tfrac{1}{2|G|}\sum_{G_t \in G} \max_{R_r \in R} f(R_r, G_t), \quad (17)$$

in which $f(R_r, G_t)$ can be defined as some typical similarity metric between the node member set R_r and G_t. We adopted F-Score and Jaccard similarity as the candidate metrics. Particularly, (i) $R = \{R_1, \cdots, R_C\}$ and (ii) $G = \{G_1, \cdots, G_C\}$ are the overlapping membership of the (i) method to be evaluated and (ii) ground-truth. For all the methods (including SLCD) we set the number of communities C according to the dataset's ground-truth and adjusted their related hyper-parameters to report the best performance metric. With regard to SLCD, we set $K_1 = K_2 = C$ for the convenience of evaluation.

The evaluation results in terms of generalized F-Score and Jaccard similarity are illustrated in Table II, with the best quality metric in **bold** and the second-best underlined.

TABLE II
PERFORMANCE EVALUATION IN TERMS OF GENERALIZED F-SCORE (%) AND JACCARD SIMILARITY (%)

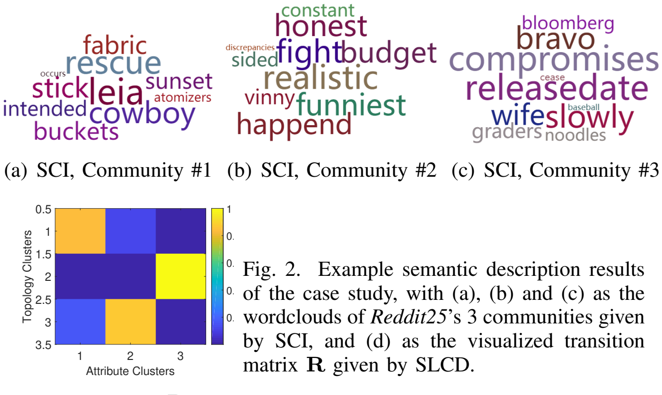
Methods	**F-Score**				**Jaccard**			
	R27	***R26***	***R25***	***EN***	***R27***	***R26***	***R25***	***EN***
NMF-T	55.27	43.00	52.36	48.41	39.26	29.11	37.01	33.42
LMBP	51.94	48.66	52.60	43.69	35.44	32.52	37.62	30.21
BigCLAM	17.81	24.29	20.36	18.90	10.58	16.33	12.63	10.92
CNLP	60.23	54.98	52.57	48.58	44.31	39.71	36.85	33.67
NMF-C	37.16	37.53	38.43	31.82	28.34	30.64	29.50	19.01
SMR	43.03	47.40	42.42	44.14	29.96	35.99	30.21	31.06
LDA	45.23	40.20	35.54	36.88	30.28	26.50	22.49	23.24
Circles	52.55	51.08	50.23	45.22	38.20	37.28	36.09	32.11
EIMF-Lap	44.84	42.27	56.44	43.83	29.42	30.83	42.06	28.73
EIMF-LP	55.37	52.33	54.70	31.28	39.93	39.56	38.97	21.70
CESNA	27.81	33.96	34.88	30.15	17.15	21.53	25.93	20.21
DNEM	55.48	53.47	54.19	50.29	40.48	42.02	40.74	**37.89**
SLCD	**62.23**	**58.27**	**60.83**	**51.32**	**47.28**	**42.30**	**45.60**	36.36

For generalized F-Score, the proposed SLCD method achieves the best performance on all the testing datasets (with the average improvement of **7.75%** compared with the second-best baseline). With regard to generalized Jaccard similarity, SLCD performs the best on 3 of the 4 datasets (with the average improvement of **8.63%** compared with the second-best competitor) and performs the second-best on the rest *Enron* dataset. In summary, the proposed SLCD method has the best performance over other state-of-the-art approaches on most of the testing datasets. Especially, SLCD achieves significantly better performance over *NMF-T* and *NMF-C* for both the metrics, indicating that the incorporation of the heterogeneous topology and content can further improve the network's community partition.

C. Case Study

To further verify SLCD's capacity to generate the semantic description, we adopt the result on *Reddit25* (with $\alpha = 1$,

[3] www.reddit.com



(a) SCI, Community #1 (b) SCI, Community #2 (c) SCI, Community #3

(d) SLCD, $\mathbf{R}$

Fig. 2. Example semantic description results of the case study, with (a), (b) and (c) as the wordclouds of *Reddit25*'s 3 communities given by SCI, and (d) as the visualized transition matrix $\mathbf{R}$ given by SLCD.

TABLE III
EXAMPLE SEMANTIC DESCRIPTION OF COMMUNITY #1'S FIRST DOMINANT TOPIC (#1) AND SECOND DOMINANT TOPIC (#2)

Topics	Top Sentences
#1	must be republican :O
	Do you actually seem either seducing each other?
	He could just be a politician.
	I'm neither democrat not republican, and frankly I dislike both Obama and Romney, but Obama has had plenty of time in office.
	But then he'd never have made the Republican nomination.
#2	Hollywood is not a big place
	Very few actors play anyone but a version of themselves with minor variations.
	Last Crusade isn't bad at all, in fact, it's an awesome movie.
	This film was very well marketed, which made the hype incredibly powerful, even after the film was released.
	Shoot, the beginning of the film is an almost shot for shot recreation of Raiders.

$\beta = 0.1$ and $K_1 = K_2 = C = 3$) as an example. For comparison, we utilize SCI [2] as another baseline. SCI can be considered as a representative conventional method that can derive the semantic descriptions based on the node-induced word-level content. Note that the content information of *Reddit25* is originally edge-induced, which cannot be directly used by SCI. Hence, we transformed the content into the corresponding node-induced form for SCI. Concretely, the keywords of a certain edge (v_i, v_j) are set to be shared by the induced nodes v_i and v_j.

The difference between the descriptions of SCI and SLCD is two-fold. First, SCI can only generate one comprehensive description for each community. However, SLCD is capable to derive the fine-grained descriptions, allowing each community to have more than one aspects (topics). Second, SCI can only give the semantic descriptions by selecting the representative keywords, while SLCD's descriptions are in the form of representative sentences (i.e., extractive summarization [15]).

Since the *Reddit* dataset (including *Reddit25*) is collected from 3 sub-forums, which are 'Movies', 'Politics' and 'Science'. These 3 sub-forums can be treated as 3 distinct topics to verify whether the descriptions given by a certain method are semantically related.

The semantic descriptions (with the top-10 keywords) of the 3 communities given by SCI are illustrated in Fig. 2 (a), (b) and (c). Moreover, we visualize the attention-like transition matrix $\mathbf{R}$ of SLCD in Fig. 2 (d). As demonstrated in Fig. 2 (d), topic (content cluster) #1 and #2 are respectively the first and second dominant topics of community #1. We further generate the summarization of such 2 topics by selecting the top sentences from $\mathbf{U}$, with the results presented in Table III.

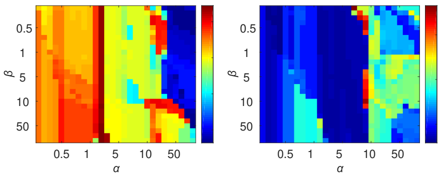
According to Fig. 2 (a)-(c), the wordclouds of SCI are coarse-grained and ambiguous. For instance, (a) can be considered as the mixture description of 'Movie' and 'Science'. Particularly, 'leia' may refer to the character Princess Leia Organa in the movie Star Wars. The keywords 'cowboy', 'rescue' and 'buckets' may also refer to the names of 3 different movies, but such reference relationships are relatively weak. The wordcloud doesn't provide the additional information to further check the reference. Moreover, 'fabric', 'sunset' and 'atomizers' may be the related keywords regarding 'Science', but they are still ambiguous. Similar problems can also be found in Fig 2 (b) and (c).

The descriptions presented in Table III are much clearer. For topic #1, there is a discussion about the Republican Party and Democratic Party, which is highly related to 'Politics'. The keywords 'republican', 'politician', 'democrat', 'Obama' and 'Romney' further enhance the primary topic regarding 'Politics'. Furthermore, with regard to topic #2, the representative sentences are all about 'Movie'. There are a series of keywords such as 'Hollywood', 'actors', 'movie' and 'film', which highlight the primary topic. Compared with the representative keywords, the summarization given by SLCD is much easier to understand, because the representative sentences can carry additional information about the relations among the words. Furthermore, topic #1 and #2 have distinct semantic, reflecting different views of community #1.

In conclusion, SLCD has the powerful capacity to derive strongly explainable and multi-view community descriptions (besides the overlapping community structures), which is superior to reveal the network's deep knowledge and semantic compared with conventional methods.

D. Analysis of Parameters and Convergence

In the experiments, we also tested the effects of $\{\alpha, \beta\}$ on all the datasets. To determine their proper value ranges, we in sequence set $\alpha, \beta \in \{0.1, 0.2, \cdots, 0.9\}$, $\{1, 2, \cdots, 9\}$ and $\{10, 20, \cdots, 100\}$. Due to the space limit, we illustrate the results on *Reddit27* and *Reddit26* (in terms of generalized F-Score) in Fig. 3 as the examples.



(a) *Reddit27* (b) *Reddit26*

Fig. 3. Parameter adjustment results in terms of generalized F-Score, with (1) and (2) corresponding to *Reddit27* and *Reddit26*.

According to Fig. 3, the result of SLCD is more sensitive to α rather than β. Furthermore, it's hard to find a fixed parameter setting that can always result in the best performance on arbitrary datasets. The results on other datasets also present similar tendencies. One significant reason is that the hyper-parameters (i.e., $\{\alpha, \beta\}$ in SLCD) can reflect the prior knowledge and bias of the dataset. Usually, different datasets may have difference priors and biases. However, we can still give the recommended search space with $\alpha \in \{1, 2, \cdots, 10\}$ and $\beta \in \{0.1, 0.2, \cdots, 1\}$. In fact, it's better to determine the parameter settings in a semi-supervised way, where we can utilize only a small proportion of the supervised label information to explore the dataset's prior knowledge. We intend to study this semi-supervised parameter setting strategy in our future work.

Besides, we also analyzed the convergence of SLCD on all the testing datasets, where we recorded the objective function (6)'s value in each iteration. Since all the results have similar tendencies, we demonstrate the convergence curve on *Reddit27* in the first 20 iterations in Fig. 4 as an example. As shown in Fig. 4, the (6)'s value decreases fast in the first 2 iterations and tends to converge in the rest iterations. It further verifies the convergence of SLCD. According to our records, the relative error of (6) can reach the precision of 10^{-4} and 10^{-6} within the first 10 and 20 iterations in most cases.

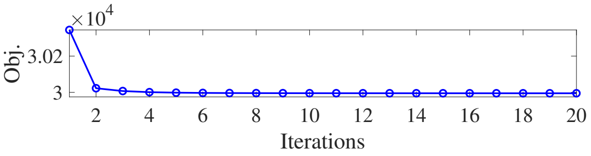


Fig. 4. Convergence curve of the objective function (6)'s value in the first 20 iterations on *Reddit27*.

VI. Conclusion

In this paper, we focused on the identification of overlapping communities with edge-induced content and introduced a novel SLCD method. It integrated the edge-induced content by partitioning link communities, which can be further transformed into the corresponding node-induced overlapping community membership. Moreover, we utilized the *view of separated clustering membership* to explore the diversity of content (with word level and sentence level), enabling SLCD to generate strongly interpretable and multi-view community descriptions. In our future work, we intend to explore an advanced semi-supervised parameter setting strategy, which can reduce the parameter search space with only a small proportion of the supervised label information. Furthermore, how to further speed up the computation via parallel and distributed computing techniques is also our next focus.

Acknowledgment

This work has been financially supported by the Shenzhen Key Laboratory Project (ZDSYS201802051831427) and the PCL Future Regional Network Facilities for Large-scale Experiments and Applications.

References

[1] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[2] X. Wang, D. Jin, X. Cao, L. Yang, and W. Zhang, "Semantic community identification in large attribute networks," in *AAAI*, 2016.

[3] D. He, Z. Feng, D. Jin, X. Wang, and W. Zhang, "Joint identification of network communities and semantics via integrative modeling of network topologies and node contents," in *AAAI*, 2017.

[4] M. Qin, D. Jin, K. Lei, B. Gabrys, and K. Musial-Gabrys, "Adaptive community detection incorporating topology and content in social networks," *Knowledge-Based Systems*, vol. 161, pp. 342–356, 2018.

[5] K. Ikeda, G. Hattori, C. Ono, H. Asoh, and T. Higashino, "Twitter user profiling based on text and community mining for market analysis," *Knowledge-Based Systems*, vol. 51, pp. 35–47, 2013.

[6] C. Yan, Y. Huang, Y. Wan, and G. Liu, "Community-based matrix factorization model for recommendation," in *International Conference on Cloud Computing and Security*. Springer, 2018, pp. 464–475.

[7] D. He, D. Liu, D. Jin, and W. Zhang, "A stochastic model for detecting heterogeneous link communities in complex networks," in *AAAI*, 2015.

[8] D. Jin, B. Gabrys, and J. Dang, "Combined node and link partitions method for finding overlapping communities in complex networks," *Scientific reports*, vol. 5, p. 8600, 2015.

[9] D. Jin, H. Wang, J. Dang, D. He, and W. Zhang, "Detect overlapping communities via ranking node popularities," in *AAAI*, 2016.

[10] J. Leskovec and J. J. Mcauley, "Learning to discover social circles in ego networks," in *Advances in neural information processing systems*, 2012, pp. 539–547.

[11] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *ICDM*. IEEE, 2013, pp. 1151–1156.

[12] J. Cao, H. Wang, D. Jin, and J. Dang, "Combination of links and node contents for community discovery using a graph regularization approach," *Future Generation Computer Systems*, vol. 91, pp. 361–370, 2019.

[13] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, p. 788, 1999.

[14] E. ADNANI, "A comprehensive literature review on community detection: Approaches and applications," *Procedia Computer Science*, vol. 151, pp. 295–302, 2019.

[15] D. Wang, S. Zhu, T. Li, Y. Chi, and Y. Gong, "Integrating clustering and multi-document summarization to improve document understanding," in *CIKM*. ACM, 2008, pp. 1435–1436.

[16] J. Kim and H. Park, "Sparse nonnegative matrix factorization for clustering," Georgia Institute of Technology, Tech. Rep., 2008.

[17] G.-J. Qi, C. C. Aggarwal, and T. Huang, "Community detection with edge content in social media networks," in *ICDE*. IEEE, 2012, pp. 534–545.

[18] C. Boutsidis and E. Gallopoulos, "Svd based initialization: A head start for nonnegative matrix factorization," *Pattern Recognition*, vol. 41, no. 4, pp. 1350–1362, 2008.

[19] Z. Zhao, S. Feng, Q. Wang, J. Z. Huang, G. J. Williams, and J. Fan, "Topic oriented community detection through social objects and link analysis in social networks," *Knowledge-Based Systems*, vol. 26, pp. 164–173, 2012.

[20] C.-D. Wang, J.-H. Lai, and S. Y. Philip, "Neiwalk: Community discovery in dynamic content-based networks," *TKDE*, vol. 26, no. 7, pp. 1734–1748, 2013.

[21] J. Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in *WSDM*. ACM, 2013, pp. 587–596.

[22] H. Hu, Z. Lin, J. Feng, and J. Zhou, "Smooth representation clustering," in *CVPR*, 2014, pp. 3834–3841.

[23] X.-H. Phan and C.-T. Nguyen, "Gibbslda++: Ac/c++ implementation of latent dirichlet allocation (lda)," *Tech. rep.*, 2007.

[24] D. Jin, X. Wang, R. He, D. He, J. Dang, and W. Zhang, "Robust detection of link communities in large social networks by exploiting link semantics," in *AAAI*, 2018.