Enterprise Community Detection

Jiawei Zhang[†], Philip S. Yu[†,¶], Yuanhua Lv[‡]
[†]University of Illinois at Chicago, Chicago, IL, USA
[¶]Institute for Data Science, Tsinghua University, Beijing, China
[‡]Microsoft, Sunnyvale, CA, USA
jzhan9@uic.edu, psyu@cs.uic.edu, yuanhual@microsoft.com

***Abstract*—Employees in companies can be divided into different social communities, and those who frequently socialize with each other are treated as close friends and will be grouped in the same community. In the enterprise context, a large amount of information about the employees is available in both (1) offline company internal sources and (2) online enterprise social networks (ESNs). What's more, each of the information sources can also contain multiple categories of employees' socialization activity information at the same time. In this paper, we propose to detect the social communities of the employees in companies based on these different information sources simultaneously, and the problem is formally called the "Enterprise Community Detection" (ECD) problem. To address the problem, a novel community detection framework named "HeterogeneoUs Multi-sOurce ClusteRing" (HUMOR) is introduced in this paper. Based on the various *enterprise social intimacy* measures introduced in this paper, HUMOR detects a set of *micro community structures* of the employees based on these different categories of information available in the online and offline sources respectively.**

***Keywords*-Community Detection; Enterprise Social Network; Data Mining**

I. Introduction

People in social organizations (e.g., schools, companies, and even countries) can generally be divided into different social communities, where individuals who frequently socialize with each other will be in the same community, while those who rarely interact with each other will be in different communities instead. Meanwhile, detecting the social communities of people within social organizations is formally called the *community detection* problem.

Problem Studied: In this paper, we propose to study the social community structures of employees in companies based on both the company internal professional information, like *enterprise organizational chart* [3], [4], [6], job title and employee workplace, and the personal information available in the online "*enterprise social networks*" (ESNs) [3], [4], [6]. Formally, the problem is named as the ECD (Enterprise Community Detection) problem. The ECD aims at fusing multiple information sources for the community detection problem, which is an important problem for big data researches.

To illustrate the ECD problem more clearly, we also provide an example in Figure 1. Based on the information available in both online ESNs and offline organizational chart, we can compute the enterprise social intimacy scores among the employees (denoted as the width of the green links among the employees). These 7 employees are divided into 3 communities, and those with large intimacy scores are grouped into the same communities respectively.

To address the ECD problem, a new enterprise social community detection framework HUMOR (HeterogeneoUs Multi-sOurce ClusteRing) is introduced in this paper. A new concept named *enterprise social intimacy* is formally defined to measure the closeness among employees in this paper. HUMOR calculates the intimacy scores among employees based on the heterogeneous information available in online ESNs and offline company internal sources respectively. Based on each category of the information, HUMOR can detect unique community structures involving the employees, which are called the *micro enterprise communities* in this paper. Framework HUMOR obtains the globally consistent enterprise community structure by fusing these detected *micro enterprise communities* with two hierarchical phases: (1) *intra-fusion* of the *micro enterprise communities* detected in online ESNs and company internal sources respectively, and (2) *inter-fusion* of the enterprise community structures between online ESNs and company internal sources.

II. Problem Formulation

In this section, we will first define several important concepts used in this paper, based on which, we will introduce the formulation of the ECD problem next.

A. Terminology Definition

Enterprise social networks studied in this paper are a new type of social networks launched in the firewalls of companies especially, which can be formally defined as follows.

Definition 1 (Enterprise Social Network (ESN)): An *enterprise social network* can be represented as a heterogeneous information network $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of nodes, and $\mathcal{E}$ represents the complex links in the network.

In this paper, we will take Yammer as an example of online ESNs. Users in Yammer can perform various kinds of social activities, e.g., (1) follow other users, (2) create and join social groups of their interests, (3) write posts, and (4) comment on/reply/like posts written by others. Besides the online ESNs, a large amount of information (e.g., the organizational chart, and various other attribute information) about the employees is available inside the company, which can be formally represented as an attribute augmented organizational chart.

Definition 2 (Attribute Augmented Organizational Chart): An *attribute augmented organizational chart* can be represented

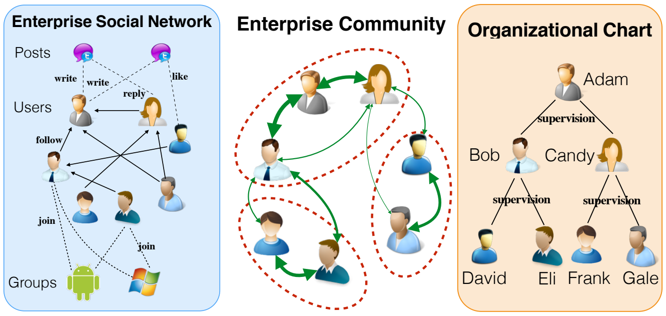

Fig. 1. An example of enterprise community detection (red circles: detected community structures; green links: enterprise intimacy scores calculated among employees, and wider links denote larger intimacy scores).

as a rooted tree $T = (\mathcal{N}, \mathcal{L}, root, \mathcal{A})$, where $\mathcal{N}$ denotes the set of employees in the company, $\mathcal{L}$ represents the management links from managers to subordinates, and $root \in \mathcal{N}$ indicates the CEO of the company. Set $\mathcal{A}$ is the set of attributes attached to employees in $\mathcal{N}$. For instance, in this paper, each employee in the company is associated with both the *job title* and *workplace* attributes. Therefore, set $\mathcal{A}$ can be represented as $\mathcal{A} = \mathcal{A}_t \cup \mathcal{A}_l$, where job title attribute set $\mathcal{A}_t = \{A_t(u_1), A_t(u_2), \cdots, A_t(u_{|\mathcal{N}|})\}$ involves the job title attribute of all the employees and employee workplace attribute set $\mathcal{A}_l = \{A_l(u_1), A_l(u_2), \cdots, A_l(u_{|\mathcal{N}|})\}$ contains the workplace attribute of the employees.

B. Problem Definition

Based on the concepts defined above, we can define the ECD problem formally in this section.

Problem Definition (The ECD Problem): Given the online ESN G and the offline attribute augmented organizational chat T, the ECD problem aims at inferring the community structure of employees in the company. More specifically, we aim at partition the employee set $\mathcal{N}$ in the company into K disjoint social communities $\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$. In this paper, we don't consider that case that each employee is involved in multiple communities. Therefore, for these detected community structures we have $C_i \cap C_j = \emptyset, \forall i, j \in \{1, 2, \cdots, K\}, i \neq j$ and $\bigcup_i C_i = \mathcal{N}$. Generally, the employees grouped in each community (e.g., C_i) tend to interact with each other more frequently and have larger closeness scores with each other compared with those in other communities (i.e., $\mathcal{C} \setminus C_i$).

III. Proposed Methods

A. ESNs based Community Detection

In this part, we will introduce the *enterprise intimacy* concept based on ESNs. With the various categories of social information in online ESNs, we can calculate different types of enterprise intimacy scores among employees and detect the corresponding micro enterprise communities respectively.

1) Enterprise Social Intimacy in Online ESN: Generally, social intimacy is a measure of how people closely interact with each other. At an individual level, specifically, the enterprise social intimacy involves the quality and number of connections one has with other employees in the company.

Definition 3 (ESNs based Enterprise Social Intimacy): Based on the social interactions information available in the online *enterprise social network* $G = (\mathcal{V}, \mathcal{E})$, the *ESNs based enterprise social intimacy* between employees u and v denotes how close u and v are in the online ESNs G, which can be represented as $EI^g(u, v)$ in this paper.

In this section, we will mis-use "employees" and "users" to represent the individuals involved in online ESNs. As introduced in the previous section, employees in online ESNs can perform various types of social activities, based on which, different enterprise social intimacy measures can be calculated, which will not be introduced here due the limited space. Full information about the enterprise social intimacy measures is available in [5].

2) ESNs based Enterprise Community Detection and Intra-Fusion: **Micro Community Structure Detection in ESNs**

Based on each enterprise social intimacy measure introduced in the previous section, a concrete enterprise social intimacy matrix can be constructed. For instance, according to the social connection based enterprise social intimacy, we can define the enterprise social intimacy matrix $\mathbf{A}_s^g \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{U}|}$, where entry $A_s^g(i,j) = EI_s^g(u_i, u_j)$, $u_i, u_j \in \mathcal{U}$ denotes the social connection information based enterprise social intimacy between employees u_i and u_j. In a similar way, we can represent the enterprise social intimacy scores among employees based on the group participation and user generated content information as matrices $\mathbf{A}_g^g, \mathbf{A}_p^g \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{U}|}$ respectively.

Based on each of the enterprise social intimacy matrix (e.g., $\mathbf{A}_s^g$), various methods can be applied to detect the community structure of the employees. In this paper, we propose to use the non-negative matrix factorization (NMF) method to partition the intimacy matrix $\mathbf{A}_s^g$ due to its outstanding performance and wide applications in clustering problems [2]. The NMF method aims at minimizing the following objective equation to infer the community structure hidden factor matrix $\mathbf{U}$:

$$J_C(\mathbf{A}_s^g) = \left\| \mathbf{A}_s^g - \mathbf{U}_s \mathbf{U}_s^\top \right\|_F^2.$$

Here, the entry $U_s(i,j)$ in the hidden factor matrix represents the confidence of employee $u_i \in \mathcal{U}$ belonging to the j_{th} community $C_j \in \mathcal{C}$. Formally, the enterprise social community structure described by matrix $\mathbf{U}_s$ is called the *micro enterprise community structure* in this paper.

Similarly, we can define the objective equations $J_C(\mathbf{A}_g^g)$ and $J_C(\mathbf{A}_p^g)$ of the enterprise social intimacy matrix based on the group membership and user generated content information. The corresponding hidden factor matrices $\mathbf{U}_g$ and $\mathbf{U}_p$ can be obtained by minimizing the following two objective equations

$$J_C(\mathbf{A}_g^g) = \left\| \mathbf{A}_g^g - \mathbf{U}_g \mathbf{U}_g^\top \right\|_F^2, \;\; J_C(\mathbf{A}_p^g) \;\; = \left\| \mathbf{A}_p^g - \mathbf{U}_p \mathbf{U}_p^\top \right\|_F^2.$$

Intra-Fusion of Micro Community Structures in ESNs

The detected hidden factor matrices $\mathbf{U}_s, \mathbf{U}_g, \mathbf{U}_p$ characterize the community structures of the employee users in ESNs from different aspects, which can be different from each other. Meanwhile, these hidden factor matrices are all about community structures of the same set of employees in the online ESNs information source, which should be consistent with the real-world community structure of the employees. To achieve such a goal, in this paper, an *consistent hidden factor matrix* $\mathbf{U} \in \mathbb{R}^{|\mathcal{U}| \times K}$ is introduced to represent the consistent community structure of the employees in online ESNs. To infer $\mathbf{U}$, we add a set of extra regularization terms to minimize the differences of community structure described by $\mathbf{U}$ from those described by $\mathbf{U}_s, \mathbf{U}_g, \mathbf{U}_p$ respectively:

$$J_R(\mathbf{A}_s^g, \mathbf{A}_g^g, \mathbf{A}_p^g) = \|\mathbf{U}_s - \mathbf{U}\|_F^2 + \|\mathbf{U}_g - \mathbf{U}\|_F^2 + \|\mathbf{U}_p - \mathbf{U}\|_F^2 .$$

Formally, the process of inferring the consistent community structure matrix $\mathbf{U}$ is called the *intra-fusion* phrase in the HUMOR framework. Based on the above remarks, the optimal consistent community structure matrix $\mathbf{U}^*$ which can minimize the matrix decomposition costs and regularization terms simultaneously can be obtained be resolving the following objective equation:

$$\mathbf{U}^* = \arg_{\mathbf{U}} \min_{\mathbf{U}, \mathbf{U}_s, \mathbf{U}_g, \mathbf{U}_p} \sum_{i \in \{s,g,p\}} J_C(\mathbf{A}_i^g) + \alpha \cdot J_R(\mathbf{A}_s^g, \mathbf{A}_g^g, \mathbf{A}_p^g),$$

where α represents the weight of the consistency regularizer and it is assigned with value 1 in the experiments.

B. Company Internal Information based Enterprise Community Detection

Besides the information available in online ESNs, employees also have a large amount of information available in the offline company information, which can also help identify the social community structures of the employees.

1) Internal Information based Enterprise Intimacy: In this section, we will measure the enterprise social intimacy among employees based on the company internal information, which include the organizational chart, job titles and workplaces. Due to the limited space, the detailed measure calculation materials are deleted, which is available in [5].

2) Internal Information based Community Detection and Intra-Fusion: Based on the above enterprise social intimacy measures, we can represent the intimacy scores among employees calculated based on organizational chart, job title and workplace information in the offline enterprise as adjacency matrices $\mathbf{A}_c^t, \mathbf{A}_t^t, \mathbf{A}_l^t \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$ respectively. Similarly, the non-negative matrix factorization (NMF) method can be applied to partition the intimacy matrices to obtain their corresponding community hidden factor matrices, and the decomposition cost functions are listed as follows:

$$J_C(\mathbf{A}_c^t) = \left\|\mathbf{A}_c^t - \mathbf{V}_c\mathbf{V}_c^\top\right\|_F^2, J_C(\mathbf{A}_t^t) = \left\|\mathbf{A}_t^t - \mathbf{V}_t\mathbf{V}_t^\top\right\|_F^2,$$
$$J_C(\mathbf{A}_l^t) = \left\|\mathbf{A}_l^t - \mathbf{V}_l\mathbf{V}_l^\top\right\|_F^2.$$

Meanwhile, these community hidden factor matrices $\mathbf{V}_c$, $\mathbf{V}_t$ and $\mathbf{V}_l$ all describe the *micro enterprise community structures* of the employees in the offline enterprise from different perspectives, which should be consistent with each other as well. To define the consistency regularization term, a new consistent community hidden factor matrix $\mathbf{V} \in \mathbb{R}^{|\mathcal{N}| \times K}$ is introduced, and the consistency regularization term between $\mathbf{V}$ and $\mathbf{V}_c$, $\mathbf{V}_t$, $\mathbf{V}_l$ can be represented as follows:

$$J_R(\mathbf{A}_c^t, \mathbf{A}_t^t, \mathbf{A}_l^t) = \|\mathbf{V}_c - \mathbf{V}\|_F^2 + \|\mathbf{V}_t - \mathbf{V}\|_F^2 + \|\mathbf{V}_l - \mathbf{V}\|_F^2 .$$

Based on the above remarks and the *intra-fusion* strategy, the optimal consistent community structure matrix $\mathbf{V}^*$ of employees in the offline enterprise can be obtained by resolving the following objective equation:

$$\mathbf{V}^* = \arg_{\mathbf{V}} \min_{\mathbf{V}, \mathbf{V}_c, \mathbf{V}_t, \mathbf{V}_l} \sum_{i \in \{c,t,l\}} J_C(\mathbf{A}_i^t) + \alpha \cdot J_R(\mathbf{A}_c^t, \mathbf{A}_t^t, \mathbf{A}_l^t),$$

where the same parameter α (with value 1 in the experiments) is also used to represents the weight of the consistency regularization term here.

C. Inter-Fusion of Enterprise Communities

Generally, the company contains the complete information about all the employees, and some of whom can get involved in the online ESNs as well. In this section, we will introduce the joint optimization objective function for detecting the communities via the *inter-fusion* of information available in the online ESNs and offline company.

For the common employees shared by the online ESNs and offline company internal information sources, the detected community structures about them should be consistent with each other. To achieve such a goal, we propose to regularize the detected community structures from online ESNs and offline company internal information sources. Before introducing the regularization term, we first define the binary employee transition matrix $\mathbf{T} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{N}|}$, where entry $T(i,j) = 1$ iff employee u_i and u_j are actually the same employee in the online ESNs and offline enterprise respectively. Based on the transitional matrix, we can map the employees as well as their hidden community matrix from the online ESN to the offline enterprise by simply multiplying it with the transition matrix. For instance, given the hidden community structure matrix $\mathbf{U}$ of the employees online, we can represent the mapped community structure matrix to the offline company as $\mathbf{T}^\top\mathbf{U}$. The difference of the employees' community structure mapped from the online ESNs (i.e., $\mathbf{T}^\top\mathbf{U}$) and that obtained from the offline enterprise (i.e., $\mathbf{V}$) can be represented as inter-source community regularization term

$$J_D(\mathbf{U}, \mathbf{V}) = \left\|(\mathbf{T}^\top\mathbf{U})(\mathbf{T}^\top\mathbf{U})^\top - \mathbf{V}\mathbf{V}^\top\right\|_F^2,$$

where the non-zero entries in matrices $(\mathbf{T}^\top\mathbf{U})(\mathbf{T}^\top\mathbf{U})^\top$ and $\mathbf{V}\mathbf{V}^\top$ denote the confidence scores for the corresponding employees to be in the same community based on the information available in online ESNs and offline company internal data respectively.

By adding the inter-source community regularization term, we can redefine the enterprise community detection objective equation. Formally, the process of adding the inter-source community regularization term $J_D(\mathbf{U}, \mathbf{V})$ to the objective function is named as the *inter-fusion* procedure in this paper. The new objective equation can be represented as:

$$\min_{\mathbf{U}, \mathbf{V}} \sum_{i \in \{s,g,p\}} J_C(\mathbf{A}_i^g) + \alpha \cdot J_R(\mathbf{A}_s^g, \mathbf{A}_g^g, \mathbf{A}_p^g)$$
$$+ \sum_{i \in \{c,t,l\}} J_C(\mathbf{A}_i^t) + \alpha \cdot J_R(\mathbf{A}_c^t, \mathbf{A}_t^t, \mathbf{A}_l^t) + \beta \cdot J_D(\mathbf{U}, \mathbf{V}),$$

where parameter β denotes the weight of the inter-source community regularization term, whose sensitivity analysis is also available in [5].

By replacing the cost and regularization functions with the matrix representations, we can obtain the final joint objective function involving 8 hidden factor matrix variables, each of which can represent the community structure of the employees from different aspects. However, simultaneous inference of the optimal values for the matrix variables can be very computational hard and time consuming.

To simplify the problem, in this paper, we propose to constrain the *intra-fusion* regularization terms $J_R(\mathbf{A}_1^g, \mathbf{A}_2^g, \mathbf{A}_3^g)$ as well as $J_R(\mathbf{A}_c^t, \mathbf{A}_t^t, \mathbf{A}_l^t)$ to be both 0, i.e.,

$$\mathbf{U}_s = \mathbf{U}_g = \mathbf{U}_p = \mathbf{U},$$
$$\mathbf{V}_c = \mathbf{V}_t = \mathbf{V}_l = \mathbf{V}.$$

And the simplified objective function will contain two variables $\mathbf{U}$ and $\mathbf{V}$ only, which will be much more efficient to address. The simplified objective function can be represented as

$$\begin{aligned}
&\min_{\mathbf{U},\mathbf{V}} \left\|\mathbf{A}_s^g - \mathbf{U}\mathbf{U}^\top\right\|_F^2 + \left\|\mathbf{A}_g^g - \mathbf{U}\mathbf{U}^\top\right\|_F^2 + \left\|\mathbf{A}_p^g - \mathbf{U}\mathbf{U}^\top\right\|_F^2 \\
&+ \left\|\mathbf{A}_c^t - \mathbf{V}\mathbf{V}^\top\right\|_F^2 + \left\|\mathbf{A}_t^t - \mathbf{V}\mathbf{V}^\top\right\|_F^2 + \left\|\mathbf{A}_l^t - \mathbf{V}\mathbf{V}^\top\right\|_F^2 \\
&+ \beta \cdot \left\|(\mathbf{T}^\top\mathbf{U})(\mathbf{T}^\top\mathbf{U})^\top - \mathbf{V}\mathbf{V}^\top\right\|_F^2.
\end{aligned}$$

The objective equation is not actually jointly convex and no closed-form solution exists. In this paper, we propose to solve with an alternative updating approach. We will fix one variable (e.g., $\mathbf{V}$) and update the other variable (e.g., $\mathbf{U}$) iteratively and alternatively, and such a process continues until all the variables converge.

Let $\mathcal{L}(\mathbf{U}, \mathbf{V})$ denote the objective function. In iteration τ, the updating equations can be represented as

$$\begin{aligned}
&\mathbf{U}^{(\tau)} = \mathbf{U}^{(\tau-1)} - \eta_1 \cdot \frac{\partial \mathcal{L}(\mathbf{U}^{(\tau-1)}, \mathbf{V}^{(\tau-1)})}{\partial(U)} \\
&= \mathbf{U}^{(\tau-1)} - 2\eta_1\Big(6\mathbf{U}^{(\tau-1)}(\mathbf{U}^{(\tau-1)})^\top\mathbf{U}^{(\tau-1)} - (\mathbf{A}_s^g + \mathbf{A}_g^g + \mathbf{A}_p^g)\mathbf{U}^{(\tau-1)} \\
&- \big((\mathbf{A}_s^g)^\top + (\mathbf{A}_g^g)^\top + (\mathbf{A}_p^g)^\top\big)\mathbf{U}^{(\tau-1)} - 2\mathbf{T}\mathbf{V}^{(\tau-1)}(\mathbf{V}^{(\tau-1)})^\top\mathbf{T}^\top\mathbf{U}^{(\tau-1)} \\
&+ 2\mathbf{T}\mathbf{T}^\top\mathbf{U}^{(\tau-1)}(\mathbf{U}^{(\tau-1)})^\top\mathbf{T}\mathbf{T}^\top\mathbf{U}^{(\tau-1)}\Big).
\end{aligned}$$

$$\begin{aligned}
&\mathbf{V}^{(\tau)} = \mathbf{V}^{(\tau-1)} - \eta_2 \cdot \frac{\partial \mathcal{L}(\mathbf{U}^{(\tau)}, \mathbf{V}^{(\tau-1)})}{\partial(V)} \\
&= \mathbf{V}^{(\tau-1)} - 2\eta_2\Big(8\mathbf{V}^{(\tau-1)}(\mathbf{V}^{(\tau-1)})^\top\mathbf{V}^{(\tau-1)} - (\mathbf{A}_c^t + \mathbf{A}_t^t + \mathbf{A}_l^t)\mathbf{V}^{(\tau-1)} \\
&- \big((\mathbf{A}_c^t)^\top + (\mathbf{A}_t^t)^\top + (\mathbf{A}_l^t)^\top\big)\mathbf{V}^{(\tau-1)} - 2\mathbf{T}^\top\mathbf{U}^{(\tau)}(\mathbf{U}^{(\tau)})^\top\mathbf{T}\mathbf{V}^{(\tau-1)}\Big).
\end{aligned}$$

where parameters η_1 and η_2 denote the gradient descent steps in updating matrices $\mathbf{U}$ and $\mathbf{V}$ respectively. The optimal learning rates η_1 and η_2 in each iteration steps obtaining the minimum $\mathcal{L}(\mathbf{U}, \mathbf{V})$ can be represented as

$$\eta_1^{(\tau)} = \arg_{\eta_1} \min \mathcal{L}(\mathbf{U}^{(\tau)}, \mathbf{V}^{(\tau)}),$$
$$\eta_2^{(\tau)} = \arg_{\eta_2} \min \mathcal{L}(\mathbf{U}^{(\tau)}, \mathbf{V}^{(\tau)}).$$

The functions can be addressed by taking derivative of $\mathcal{L}(\cdot)$ with regards to η_i (or η_2) and make it equal to 0, we can obtain a cubic equation involving η_i (or η_2). Multiple roots may exist when addressing the equation and the representation of the roots is very complicated. In this paper, for simplicity, we propose to assign η_i and η_2 with a small constant value (i.e., 0.05 in the experiments). The above alternative updating scheme involves the multiplication of matrices. Let n be the number of employees in the company and τ be the required rounds to achieve convergence, the time complexity of the alternative updating method can be represented $O(\tau n^{2.373})$, where the Optimized Coppersmith-Winograd algorithm [1] is applied in the matrix product calculation. In addition, some pre-computation can be applied to effective reduce the real running time of the framework by storing the results of matrices $(\mathbf{A}_s^g)^\top + (\mathbf{A}_g^g)^\top + (\mathbf{A}_p^g)^\top$, $\mathbf{T}\mathbf{T}^\top$, etc., in advance.

To test the effectiveness of the proposed framework HUMOR in detecting the communities in companies, extensive experiments have been done on real-world enterprise datasets. For more information about the experiment settings, experiment results and parameter sensitivity analysis, please refer to [5] for more information.

IV. Conclusion

In this paper, we have studied the enterprise community detection problem based on the enterprise information about the employees in both company internal information sources and online ESNs involving the employees. An integrated community detection framework HUMOR has been proposed in this paper. HUMOR can detect the *micro community structures* about the employees based on each category of the *enterprise intimacy* scores calculated based on one type of enterprise information. In addition, the *micro community structures* about the employees are further fused in HUMOR with the *intra-fusion* and *inter-fusion* steps. Extensive experiments have been done on the real-world enterprise datasets (including both the company internal data and the online ESNs data), and the results have demonstrated the outstanding performance of framework HUMOR.

V. Acknowledgement

This work is supported in part by NSF through grants IIS-1526499, and CNS-1626432, and NSFC 61672313. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X GPU used for this research.

References

[1] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, 1990.

[2] J. Yang and J. Leskovec. Overlapping community detection at scale: A nonnegative matrix factorization approach. In *WSDM*, 2013.

[3] J. Zhang, Y. Lv, and P. Yu. Friend recommendation in enterprise social networks. In *CIKM*, 2015.

[4] J. Zhang, P. Yu, and Y. Lv. Organizational chart inference. In *KDD*, 2015.

[5] J. Zhang, P. Yu, and Y. Lv. Enterprise community detection via hierarchical information fusion. 2017. [Online; "https://www.cs.uic.edu/~jzhang2/files/2017_icde_paper_full.pdf"; accessed 22-January-2017].

[6] J. Zhang, P. Yu, Y. Lv, and Q. Zhan. Information diffusion at workplace. In *CIKM*, 2016.