SMACD: Semi-supervised Multi-Aspect Community Detection

Ekta Gujral
University of California Riverside
egujr001@ucr.edu

Evangelos E. Papalexakis
University of California Riverside
epapalex@cs.ucr.edu

ABSTRACT

Community detection in real-world graphs has been shown to benefit from using multi-aspect information, e.g., in the form of "means of communication" between nodes in the network. An orthogonal line of work, broadly construed as semi-supervised learning, approaches the problem by introducing a small percentage of node assignments to communities and propagates that knowledge throughout the graph. In this paper we introduce SMACD, a novel semi-supervised multi-aspect community detection. To the best of our knowledge, SMACD is the first approach to incorporate multi-aspect graph information and semi-supervision, while being able to discover communities. We extensively evaluate SMACD's performance in comparison to state-of-the-art approaches across six real and two synthetic datasets, and demonstrate that SMACD, through combining semi-supervision and multi-aspect edge information, outperforms the baselines.

KEYWORDS

Community Detection, Tensors, Tensor Decomposition, Coupled Matrix Tensor Factorization, Semi-supervised Learning

1 INTRODUCTION

Community detection in real graphs is a widely pervasive problem with applications in social network analysis and collaboration networks, to name a few. There have been continuing research efforts in order to solve this problem. Traditionally, research has focused plain graphs where the only piece of information present is the nodes and the edges [17] . In most real applications, however, the information available usually goes beyond a plain graph that captures relations between different nodes. For instance, in an on-line social network such as Facebook, relations and interactions between users are inherently *multi-aspect* or *multi-view*, i.e., they are naturally represented by a set of edge types rather than a single type of edge. Such different edge-types can be "who messages whom", "who pokes whom", "who-comments on whose timeline" and so on. There exists a significant body of work [5, 8, 10, 20] that uses this multi-view nature of real graphs for community detection.
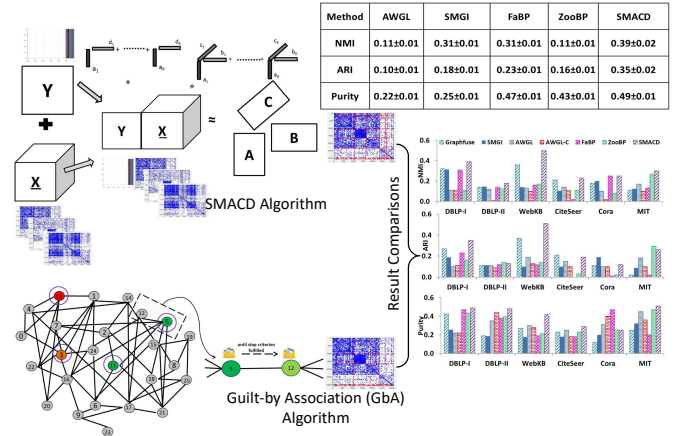
Figure 1: SMACD vs state-of-art techniques: Our proposed method SMACD successfully combines multi-view graph information and semi-supervision and outperforms state-of-the-art techniques.

Another line of work leverages partial ground truth information that may be available to us. Such partial ground truth information manifests as a small percentage of nodes for which we know the community where they belong. These partial node labels may be obtained via questionnaires or by leveraging domain expert opinion, however, since the process of obtaining those labels may be costly and time-consuming, we assume that they represent a small percentage of the nodes in our graph. The most popular school of thought that takes such partial ground truth into account are the so called "Guilt-by-Association" or label propagation techniques where the main idea is that affinity between nodes implies affiliation with the same community and those techniques iteratively propagate the known node labels throughout the graph estimating the unknown labels.

The problem that we solve is the following:

> **PROBLEM 1.** **Given** *(a) a multi-view or multi-aspect graph, and (b) a p% of node labels to R communities,* **find** *an assignment of all nodes of the graph to one (or more) of the R communities.*

Our main contributions are:

- **Novel Approach**: We introduce SMACD, a semi-supervised multi-aspect community detection algorithm that leverages multiple views of a graph and an existing (small) percentage of node labels for community detection.
- **Algorithm**: Under the hood of SMACD runs our proposed algorithm for Non-Negative Sparse Coupled Matrix-Tensor Factorization (NNSCMTF) which jointly decomposes a tensor that represents a multi-view graph, and a matrix which

contains partial node label information. NNSCMTF introduces latent sparsity and non-negativity constraints to the Coupled Matrix-Tensor Factorization model [1], which are well suited for community detection.

- **Evaluation on Real Data**: We conduct extensive experiments in order to evaluate SMACD's performance in comparison to state-of-the-art methods.

2 PROPOSED METHOD

2.1 Preliminary Definitions

A multi-view graph with K views is a collection of K adjacency matrices $X_1, \cdots X_K$ with dimensions $I \times I$ (where I is the number of nodes). This collection of matrices is naturally represented as a tensor $\underline{X}$ of size $I \times I \times K$. A tensor is a higher order generalization of a matrix. In order to avoid overloading the term "dimension", we call an $I \times J \times K$ tensor a three "mode" tensor, where "modes" are the numbers of indices used to index the tensor. We refer the interested reader to several surveys that provide more details and a wide variety of tensor applications [15] . One of the most popular and widely used tensor decompositions is the Canonical Polyadic (CP) or CANDECOMP/PARAFAC decomposition [4, 12], henceforth referred to this decomposition as CP. In CP, the tensor is decomposed into a sum of rank-one tensors, i.e., a sum of outer products of three vectors (for three-mode tensors): $\underline{X} \approx \sum_{r=1}^{R} A(:, r) \circ B(:, r) \circ C(:, r)$ where $A \in \mathbb{R}^{I \times R}, B \in \mathbb{R}^{J \times R}, C \in \mathbb{R}^{K \times R}$, and the outer product is given by $(A(:, r) \circ B(:, r) \circ C(:, r))(i, j, k) = A(i, r)B(j, r)C(k, r)$ for all i, j, k.

There are cases where in addition to the tensor, we also have a matrix whose rows (without loss of generality) have one-to-one correspondence with one of the modes of the tensors. We refer to this matrix and tensor as "coupled" and we can jointly analyze them using the Coupled Matrix-Tensor Factorization [1], a model which will be the basis for our proposed method.

2.2 SMACD: Semi-supervised Multi-Aspect Community Detection

As [10] has demonstrated, using higher-order information for the edges of a graph, such as the "means of communication", results in more accurate community detection. What if we additionally have semi-supervision in the form of community labels for a small subset of the individuals? In this Section we introduce SMACD which formulates this problem as a matrix-tensor couple, where the matrix contains the community labels for the small subset of users that are known, and missing values for the rest of its entries. The key rationale behind SMACD is the following: Using the coupled matrix that contains partial label information for each node will provide a *soft guide* to the tensor decomposition with respect to the community structure that it seeks to identify. Thus, using this side information we essentially guide the decomposition to compute a solution which bears a community structure as close to the partial labels as possible (in the least squares sense).

In [3] the authors propose semiBAT, where they follow a different approach of incorporating semi-supervision in the context of matrix-tensor coupling: instead of a bilinear decomposition for Y (the partial label matrix) which provides soft guidance to the

structure discovery, semiBAT explicitly uses a classification loss in the objective function. In [3] the goal classification of brain states, rather than discovering community structure, thus explicitly using the classification loss instead of taking a low-rank factorization of the label matrix seems more appropriate.

At a high level, SMACD takes as input a tensor $\underline{X} \in \mathbb{R}^{I \times I \times K}$ which contains the multi-view graph, a matrix $Y \in \mathbb{R}^{I \times R}$ containing the node assignments to communities, and the number of communities R (which is given implicitly through matrix Y. SMACD consists of the following two steps.

Step 1: Decomposition Given $\underline{X}, Y$ compute an $R - 1$ component Sparse and Non-negative MTF (as shown below in Section 2.2.1). The columns of A and B contain soft assignments of each node to one of $R-1$ communities. Both matrices contain similar information (which in practice ends up being almost identical, especially in cases where we have symmetric tensors in the first two modes).

Step 2: Hard Assignment In this step we assign each node to a single community by finding the community with maximum membership. This translates to finding the maximum column index for each row (which corresponds to each node). In the previous step we have computed a sparse decomposition which causes a number of the nodes to have all-zero rows in A, i.e., they have no assignment to any of the $R - 1$ communities. We assign those nodes to the R-th community which essentially is meant for capturing all remaining variation that our CP model in the CMTF decomposition was unable to capture. Step 2 is necessary only in the case where we have *non-overlapping* communities. However, SMACD works for overlapping communities as well, simply by eliminating Step 2 and computing Step 1 for R communities instead of $R - 1$.

2.2.1 Non-negative Sparse Coupled Matrix-Tensor Factorization (NNSCMTF). In this section we describe our model along with an Alternating Least Squares algorithm that computes a locally optimal solution. We propose two constraints on top of the CMTF model , motivated by community detection:

Non-negativity Constraint: SMACD uses the factor matrices A, B as community assignments. Such assignments are inherently non-negative numbers (a negative assignment to a community is hard to interpret and is not natural). Thus, in NNSCMTF we impose element-wise non-negativity constraints (denoted as $A \geq 0$) to all factor matrices.

Latent Sparsity Constraint: In order to (a) further enhance interpretability and (b) suppress noise, we impose latent sparsity to the factors of the model. Intuitively, we would like the coefficients of the factor matrices to be non-zero only when a node belongs to a particular community, thus eliminating the need for ad-hoc thresholding. To that end we introduce ℓ_1 norm regularization for all factors which promotes a sparse solution.

The proposed model is:

$$\min_{A \geq 0, B \geq 0, C \geq 0, D \geq 0} \|\underline{X} - \sum_r A(:, r) \circ B(:, r) \circ C(:, r)\|_F^2 + \|Y - AD^T\|_F^2$$

$$+ \lambda \sum_{i, r} |A(i, r)| + \lambda \sum_{j, r} |B(j, r)| + \lambda \sum_{k, r} |C(k, r)| + \lambda_d \sum_{l, r} |D(l, r)| \quad (1)$$

where λ is the sparsity regularizer penalty. The above objective function is highly non-convex and thus hard to directly optimize.

However, we use Alternating Least Squares (ALS), a form of Block Coordinate Descent (BCD) optimization algorithm, in order to solve the problem of Eq. 1. The reason why we choose ALS over other existing approaches, such as Gradient Descent [1], , is the fact that ALS offers ease of implementation and flexibility of adding constraints and regularizers, does not introduce any additional parameters that may influence convergence, and as a family of algorithms has been very extensively studied and used in the context of tensor decompositions. The main idea behind ALS is the following: when fixing all optimization variables except for one, the problem essentially boils down to a constrained and regularized linear least squares problem which can be solved optimally. Thus, ALS cycles over all the optimization variables and updates them iteratively until the value of the objective function stops changing between consecutive iterations. In ALS/BCD approaches, such as the one proposed here, when every step of the algorithm is solved optimally, then the algorithm decreases the objective function monotonically.

In the following lines we demonstrate the derivation of one of the ALS steps. Let us denote $\mathbf{X}_{(i)}$ the i-th mode matricization or unfolding of $\underline{\mathbf{X}}$, i.e., the unfolding of all slabs of $\underline{\mathbf{X}}$ into an $I \times JK$ matrix (we refer the interested reader to [15] for a discussion on matricization), then because of properties of the CP/PARAFAC model [15], fixing $\mathbf{B}, \mathbf{C}, \mathbf{D}$ we have

$$\min_{\mathbf{A} \geq \mathbf{0}} \|\mathbf{X}_{(1)} - \mathbf{A}[(\mathbf{B} \odot \mathbf{C})^T\|_F^2 + \|\mathbf{Y} - \mathbf{A}\mathbf{D}^T\|_F^2 + \lambda \sum_{i,r} |\mathbf{A}(i,r)| \Rightarrow$$

$$\min_{\mathbf{A} \geq \mathbf{0}} \|[\ \mathbf{X}_{(1)} \ ; \mathbf{Y}] - \mathbf{A}[(\mathbf{B} \odot \mathbf{C})^T \quad \mathbf{D}^T]\|_F^2 + \lambda \sum_{i,r} |\mathbf{A}(i,r)| \Rightarrow$$

$$\min_{\mathbf{A} \geq \mathbf{0}} \|[\mathbf{L} - \mathbf{A}\mathbf{M}]\|_F^2 + \lambda \sum_{i,r} |\mathbf{A}(i,r)| \tag{2}$$

where $\mathbf{L} = [\ \mathbf{X}_{(1)} \ ; \mathbf{Y}]$, and $\mathbf{M} = [(\mathbf{B} \odot \mathbf{C})^T \quad \mathbf{D}^T]$. This problem is essentially a Lasso regression on the columns of $\mathbf{A}$ [21] and we use coordinate descent to solve it optimally . The update formulas for $\mathbf{B}, \mathbf{C}, \mathbf{D}$ follow the same derivation after fixing all but the matrix that is being updated. We omit the full listing of the algorithm due to space restrictions.

3　EXPERIMENTAL EVALUATION

In this section we extensively evaluate the performance of SMACD on two synthetic and six real datasets, and compare its performance with state-of-the-art approaches which either use multi-view graphs or semi-supervision (but not both) for community detection. We implemented SMACD in Matlab using the functionality of the Tensor Toolbox for Matlab [2] which supports efficient computations for sparse tensors.

3.1　Data-set description

3.1.1　Synthetic data generation. In order to fully control the community structure in our experiments we generate synthetic multi-view graphs which different cluster density. We generally follow the synthetic data creation of [10]. We partition the adjacency matrices corresponding to different graph views into different blocks, each one corresponding to a community. We further corrupt those datasets with random Gaussian noise with variance 0.05. We construct two synthetic datasets: Synthetic-1 has 5 views and 5 communities and has very few "cross-edges", whereas Synthetic-2

has 3 views and higher number of "cross-edges", making it a harder dataset. We include those synthetic datasets in our code package.

3.1.2　Real Data Description. In order to truly evaluate the effectiveness of SMACD, we test its performance against six real datasets that have been used in the literature. Those datasets are: DBLP-I, DBLP-II, Cora, CiteSeer, WebKB, and MIT reality mining dataset. DBLP-I and DBLP-II datasets are collected from the DBLP online database and were used in [10]. In DBLP-I and DBLP-II, the first graph view represents citations of one author to another. The second view represents co-authorship relations. Finally the third view relates two authors if they share any three terms in a title or in abstract of their publication. The Cora dataset [19] was collected from the LINQS online database, and consists of 2708 machine learning publications and citations. This network consists of 5429 edges and 7 different communities. CiteSeer dataset [19] consists of 3312 publications related to AI, DB, IR, ML and HCI research categories. The WebKB dataset [19] is small dataset of 878 web pages of Washington universities which belong to 5 categories, namely courses, facilities, student, project and staff. We considered these categories as ground truth classes. Finally, the MIT reality mining [7], collected by researchers at MIT, consists of 87 mobile users information collected on campus. Ground truth is the self-reported affiliation of the users.

3.2　Evaluation Measures

We evaluate the community detection performance in terms of three different quality measures: Normalized Mutual Information (NMI), Adjacent Random Index (ARI) and Purity. These measures provide a quantitative way to compare the obtained communities $\Omega = w_1, w_2, \ldots, w_r$ to ground truth classes C= $c_1, c_2, \ldots, c_r$.

More Specifically, $NMI(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]}$ where I(Ω,C) is mutual information between cluster Ω and C, H(Ω) and H(C) are entropy of cluster and classes. Next, Purity is defined as the ratio of number of nodes correctly extracted to total number of nodes. Formally, $Purity(\Omega, C) = \frac{1}{N} \sum_{k=0}^{N} max|w_k \cap c_k|$ where w_k and c_k are the number of objects in a community and a class respectively. $|w_k \cap c_k|$ is the interaction of objects of w_k and c_k. Finally when interpreting communities as binary decisions of each object pair, Adjacent Random Index(ARI) is defined as: $ARI(\Omega, c) = \frac{tp+tn}{tp+fp+fn+tn}$, where tp, tn, fp and fn are true positive, true negative, false positive and false negative, respectively. . NMI, Purity and ARI are defined on the scale $[0, 1]$ and the higher the score, the better the community quality is.

3.3　Baselines for Comparison

Here we briefly present the state-of-the-art baselines. For each baseline we use the *reported parameters* that yielded the best performance in the respective publications. For fairness, we also compare against the parameter configuration for SMACD that yielded the best performance in terms of NMI. All comparisons were carried out over 50 iterations each, and each number reported is an average with a standard deviation attached to it.

GraphFuse [10]: GraphFuse is a tensor decomposition based approach which can be seen as a special case of SMACD when there is no semi-supervision. The sparsity penalty factor λ for DBLP-I,
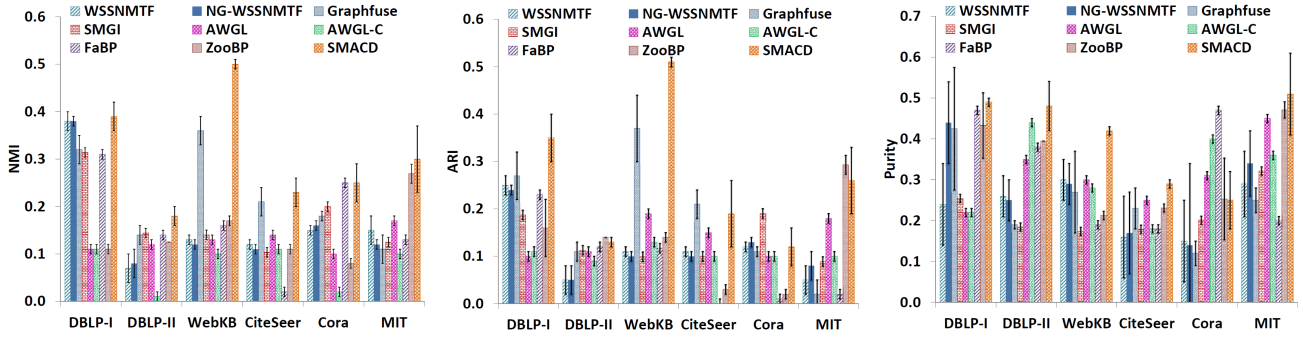
Figure 2: Experimental results for NMI, ARI and Purity. SMACD mostly outperforms baselines and, in particular, works better in very hard scenarios such as the MIT dataset.

DBLP-II, CiteSeer, Cora, WebKB and MIT is set for λ= 0.000001, 0.0001, 0.000001, 0.1, 0.00005 and 0.00001, respectively and a maximum of 150 iterations was used for convergence.

WSSNMTF and NG-WSSNMTF [11]: The details for the methods are described in [11]. We used the SVD matrix initialization. The sparsity penalty parameter η for WSSNMTF and NG-WSSNMTF are chosen to lead to best clustering performance and max 100 iterations are used for reaching the convergence.

Fast Belief Propagation (FaBP) [16]: FaBP is a fast, iterative Guilt-by-Association technique, in particular conducting Belief Propagation. A belief in our case is a community label for each node. We used one-vs-all technique for multi-clustering.

ZooBP [9]: ZooBP works on any undirected heterogeneous graph with multiple edge types. As in FaBP, a belief here is a community label for each node.

SMGI [14]: Sparse Multiple Graph Integration method is another method of integrating multiple graphs for label propagation, which introduces sparse graph weights which eliminate the irrelevant views in the multi-view graph.

AWGL [18]: Parameter-Free Auto-Weighted Multiple Graph Learning is the latest auto-weighted multiple graph learning framework, which can be applied to multi-view unsupervised (AWGL-C) as well as semi-supervised (AWGL) clustering task.

Parameter Tuning In order to be on-par with the baselines, we tuned SMACD's parameter λ so that we obtain the maximum performance. We provided ≤10% labels in matrix and rest of labels are empty. The maximum number of iterations for SMACD is set to 10^3. We perform experiments with various values of λ ranging from 10^{-8} to 10^6 on all real multi-view networks to explore the behaviour of our algorithm. λ is chosen to give best clustering results in terms of NMI, for DBLP-I, DBLP-II, CiteSeer, Cora, WebKB and MIT values for λ= 0.3, 0.09, 0.0001, 1, 0.9 and 600, respectively. For both the synthetic data, penalty factor is set to 1.

3.4 Experimental Results

Below we extensively evaluate SMACD and compare it against baseline methods.

3.4.1 Comparison with Baselines. For all datasets we compute Normalized Mutual Information, Purity and Adjacent Random Index. For SMACD, AWGL , SMGI, ZooBP and FaBP we use labels for 10% of the nodes in each dataset. We observe that SMACD performed better than other approaches when applied on SYN-I and SYN-II. SYN-I is designed with high cluster density in layer 2 and 3,

and noisy links, and has high number of cross-community edges between nodes. Given that, we found that SMACD achieved the highest NMI, ARI and Purity. We omit the figure of the results due to space restrictions.
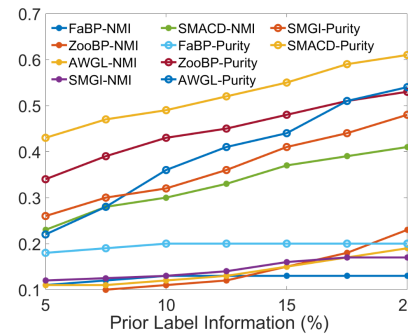

Figure 3: SMACD vs. Guilt-by-Association(FaBP and ZooBP), AWGL and SMGI for different degrees of semi-supervision for DBLP-I.

The most interesting comparison, however, is on the real datasets, since they present more challenging cases than the synthetic ones, shown in Figure 2. SMACD outperforms the other state-of-the-art approaches in most of the real multi-view networks, with the exception of Cora. In the cases of DBLP-I and DBLP-II, SMACD gave better results compared to the baselines , specifically in terms of NMI and Purity. For Citeseer, SMACD has comparable behavior with the baselines in terms of NMI. Most importantly, however, SMACD achieves the highest NMI, ARI, and Purity for WebKB and MIT, arguably the hardest of the six real datasets we examined and have been analyzed in the literature.

3.4.2 Performance vs. Degree of Semi-supervision. Next, we evaluate the performance of SMACD compared to Guilt-by-Association as a function of the degree of semi-supervision, i.e., the percentage of available labels. We performed experiments for the DBLP-I dataset for 5%, 10% and 20% labeled nodes and we show the results in Figure 3 showing a consistent trend between the two methods. We further measure the performance of SMACD as the number of labels grows, and summarize the results in Figure 4 where we can see that what we would expect intuitively holds true: the more labels we have the higher the community accuracy. Due to limited space, we show the trend only for DBLP-I but we observe similar behavior for the rest of the datasets.

From figure 4, it is observed that when all the nodes are labeled, SMACD achieved the Purity ≈ 0.93 and NMI ≈ 0.75. The reason for
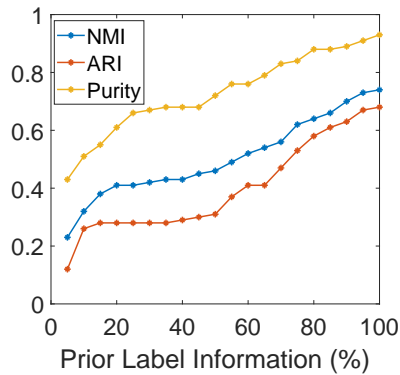
Figure 4: Performance of SMACD for DBLP-I as a function of the number of labels. These results confirm the intuition, since performance improves as the number of labels increases.

this behavior is the fact that we are modeling each community as a rank-1 tensor, and there appears to exist community structure that we cannot model as such. We reserve further investigation for future work.

4 RELATED WORK

We provide review of work related to our problem.

Multi-vew Clustering/Community Detection: There is work in the literature (such as some of the baselines we compare against) that leverages multiple graph views for community detection, including Weighted Simultaneous Symmetric Non-negative Matrix Trifactorization (WSSNMTF) and Natural Gradient Weighted Simultaneous Symmetric Non-negative Matrix Trifactorization (NG-WSSNMTF) [11] and GraphFuse [10].

Heterogeneous Information Networks (HIN): Heterogeneous Information Networks are versatile representations of networks that involve multiple typed objects (or nodes) and multiple typed links denoting different relations (or edges). There is a fairly rich body of work in the literature working on related problems to ours [13, 22], however, we were unable to find an implementation directly applicable to the problem at hand for experimental comparison.

Guilt-by-Association techniques: Prior work that has leveraged label propagation for community detection, including FaBP [16] and ZooBP [9], with the latter also leveraging the multi-view nature of the graph.

Tensor and Coupled Models: To the best of our knowledge the NNSCMTF model has not been previously proposed. Most relevant to our proposed framework, Cao *et al.* [3], propose a semi-supervised learning framework, based on matrix-tensor coupling. We were unable to directly compare the method of [3] as released because the focus of [3] is 4-mode tensors.

5 CONCLUSIONS

We introduce SMACD, a novel approach on semi-supervised multi-aspect community detection based on a novel coupled matrix-tensor model. We propose an automated parameter tuning algorithm, which effectively renders SMACD *parameter-free*. We extensively

evaluate SMACD's effectiveness over the state-of-the-art, in a wide variety of real and synthetic datasets, demonstrating the merit of leveraging semi-supervision and higher-order edge information towards high quality overlapping and non-overlapping community detection.

6 ACKNOWLEDGEMENTS

REFERENCES

[1] Evrim Acar, Tamara G Kolda, and Daniel M Dunlavy. 2011. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422* (2011).

[2] Brett W. Bader, Tamara G. Kolda, et al. 2015. MATLAB Tensor Toolbox Version 2.6. Available online. (2015). http://www.sandia.gov/~tgkolda/TensorToolbox/

[3] Bokai Cao, Chun-Ta Lu, Xiaokai Wei, S Yu Philip, and Alex D Leow. 2016. Semi-supervised Tensor Factorization for Brain Network Analysis. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 17–32.

[4] J Douglas Carroll and Jih-Jie Chang. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of âĂIJEckart-YoungâĂİ decomposition. *Psychometrika* 35, 3 (1970), 283–319.

[5] Wei Cheng, Xiang Zhang, Zhishan Guo, Yubao Wu, Patrick F Sullivan, and Wei Wang. 2013. Flexible and robust co-regularized multi-domain graph clustering. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 320–328.

[6] Vin De Silva and Lek-Heng Lim. 2008. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.* 30, 3 (2008), 1084–1127.

[7] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, and Nikolai Nefedov. 2012. Clustering with multi-layer graphs: A spectral perspective. *IEEE Transactions on Signal Processing* 60, 11 (2012), 5820–5831.

[8] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, and Nikolai Nefedov. 2014. Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds. *IEEE Transactions on signal processing* 62, 4 (2014), 905–918.

[9] Dhivya Eswaran, Stephan Günnemann, Christos Faloutsos, Disha Makhija, and Mohit Kumar. 2017. ZooBP: belief propagation for heterogeneous networks. *Proceedings of the VLDB Endowment* 10, 5 (2017).

[10] Evangelos E. Papalexakis, Leman Akoglu, and Dino Ienco. [n. d.]. Do more views of a graph help? community detection and clustering in multi-graphs. In *IEEE FUSION'13*.

[11] Vladimir Gligorijevic, Yannis Panagakis, and Stefanos Zafeiriou. [n. d.]. Fusion and Community Detection in Multi-layer Graphs. ([n. d.]).

[12] R.A. Harshman. 1970. Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis. (1970).

[13] He Jiang, Yangqiu Song, Chenguang Wang, Ming Zhang, and Yizhou Sun. [n. d.]. Semi-supervised Learning over Heterogeneous Information Networks by Ensemble of Meta-graph Guided Random Walks.

[14] Masayuki Karasuyama and Hiroshi Mamitsuka. 2013. Multiple graph label propagation by sparse integration. *IEEE transactions on neural networks and learning systems* 24, 12 (2013), 1999–2012.

[15] T.G. Kolda and B.W. Bader. 2009. Tensor decompositions and applications. *SIAM review* 51, 3 (2009).

[16] Danai Koutra, Tai-You Ke, U Kang, Duen Horng Polo Chau, Hsing-Kuo Kenneth Pao, and Christos Faloutsos. 2011. Unifying guilt-by-association approaches: Theorems and fast algorithms. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 245–260.

[17] Jure Leskovec, Kevin J Lang, and Michael Mahoney. 2010. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*. ACM, 631–640.

[18] Feiping Nie, Jing Li, Xuelong Li, et al. [n. d.]. Parameter-Free Auto-Weighted Multiple Graph Learning: A Framework for Multiview Clustering and Semi-Supervised Classification.

[19] Prithviraj Sen, Galileo Mark Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. 2008. Collective Classification in Network Data. *AI*

Magazine 29, 3 (2008), 93–106.

[20] Wei Tang, Zhengdong Lu, and Inderjit S Dhillon. 2009. Clustering with multiple graphs. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*. IEEE, 1016–1021.

[21] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.

[22] Mengting Wan, Yunbo Ouyang, Lance Kaplan, and Jiawei Han. 2015. Graph regularized meta-path based transductive regression in heterogeneous information network. In *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 918–926.