SMACD: Semi-supervised Multi-Aspect Community Detection

Ekta Gujral
UC Riverside
egujr001@ucr.edu

Evangelos E. Papalexakis
UC 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 method along with an automated parameter tuning algorithm which essentially renders SMACD parameter-free. 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 overlapping and non-overlapping communities. We extensively evaluate SMACD's performance in comparison to state-of-the-art approaches across eight real and two synthetic datasets, and demonstrate that SMACD, through combining semi-supervision and multi-aspect edge information, outperforms the baselines.

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 [18] .

In most real applications, however, the information available usually goes beyond a plain graph that captures relations between different nodes. For instance, in an online 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 that uses this multi-view nature of real graphs for community detection. Indicatively, [23, 5, 10, 8] proposed algorithm combines multiple views of a graph in order to detect communities more accurately.

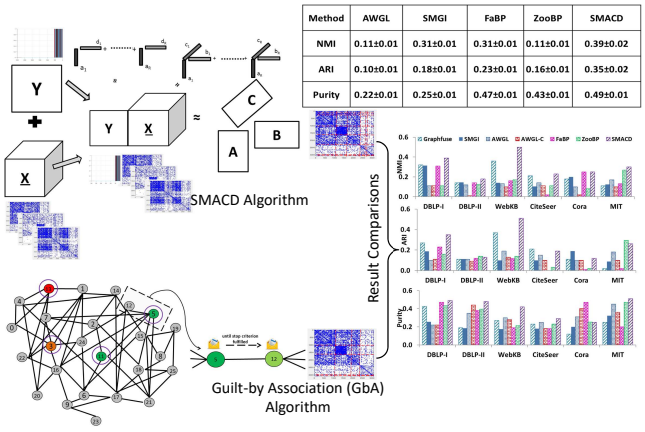


Figure 1: SMACD vs state-of-art techniques: Our proposed method SMACD successfully combines multi-view graph information and semi-supervision and outperforms state-of-the-art techniques.

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. Belief propagation [17] is one of the widely used "Guilt-by-Association" method and has been very successful in various real life scenarios including community detection. In view of that method which only gives non

zero weight to every graph Karsuyama et al. [15] proposed another multiple graph learning method (SMGI), where weight can be sparse. Auto-weighted Multiple-Graph Learning (AWGL) [19] framework learn the set of weights automatically for all graphs and classify graphs into different classes.

In this paper we propose a new approach to the problem of community detection that effectively integrates and leverages both (a) the multi-view nature of real graphs, and (b) partial supervision in the form of community labels for a small number of the nodes. To the best of our knowledge, this is the first approach that combines (a) and (b), and detects overlapping and non-overlapping communities.

Our main contributions are:

- **Novel Approach**: We introduce SMACD, a semi-supervised multi-aspect[1] community [2] detection algorithm. To the best of our knowledge, this is the first principled method for leveraging multiple views of a graph and an existing (small) percentage of node labels for community detection and is able to handle *overlapping* communities.
- **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.
- **Automated Parameter Turning**: Sparsity introduced by NNSCMTF is controlled by a parameter which if chosen arbitrarily may not yield the best possible performance. In Section 2 we introduce SelSPF, an automated parameter tuning algorithm that does not rely on the partial node labels and selects a value for the sparsity parameter which yields performance in terms of community detection accuracy which is on par with the one obtained when doing an exhaustive search for that parameter based on all the ground truth available to us.
- **Evaluation on Real Data**: We conduct extensive experiments in order to evaluate SMACD's performance in comparison to state-of-the-art methods.

[1]Note that in the paper we use the terms multi-view and multi-aspect interchangeably

[2]Note that in the paper we use the terms community and cluster interchangeably

Reproducibility: We encourage reproducibility and extension of our results by making our Matlab implementation and the synthetic data we used available at link[3]. Note also that all the datasets we use for evaluation are publicly available.

2 Problem Formulation & Proposed Method

2.1 Problem Definition Graphs are effective way to represent a large variety of data and relations between data entities. Each entity represented by node or vertex (V) and relation between entities are defined by weighted or unweighted edges (E_i, w_i). In this paper we focus on multi-view or multi-aspect graphs, i.e., a collection of graphs for the same set of nodes and different set of edges per view or layer. In the remainder of the paper we use the terms "view", "aspect", and "layer" interchangeably. Each graph can be represented using an adjacency matrix, a square node-by-node matrix that indicates an edge (and a potential weight associated to it) between two nodes. A multi-view graph can be, thus, represented as a collection of adjacency matrices.

The goal of our paper is to identify communities in that multi-view graph, which essentially boils down to assigning each node into one of R community labels. In order to simplify our problem definition, we assume that R is given to us. (there exist, however, heuristics in tensor literature [20] that can deal with an unknown R). 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.*

2.2 Preliminary Definitions A multi-view graph with K views is a collection of K adjacency matrices $\mathbf{X}_1, \cdots \mathbf{X}_K$ with dimensions $I \times I$ (where I is the number of nodes). This collection of matrices is naturally represented as a tensor $\underline{\mathbf{X}}$ 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. Table 1 contains the symbols used throughout the paper. We refer the interested reader to several surveys that provide more details and a wide variety of tensor applications [16] . In the interest of space, we also refer the reader to [16] for the definitions of Kronecker and Khatri-Rao products which

[3]`http://www.cs.ucr.edu/~egujr001/ucr/madlab/src/SHOCD.zip`

are not essential for following the basic derivation of our approach but are needed for fully appreciating the math behind it. One of the most popular and widely used

Symbols	Definition
$\underline{\mathbf{X}}, \mathbf{X}, \mathbf{x}, x$	Tensor, Matrix, Column vector, Scalar
$\mathbb{R}$	Set of Real Numbers
$\circ$	Outer product
$[\mathbf{A}; \mathbf{B}]$	Vertical stacking of $\mathbf{A}, \mathbf{B}$
$[\mathbf{A} \quad \mathbf{B}]$	Horizontal stacking of $\mathbf{A}, \mathbf{B}$
$\|\mathbf{A}\|$	Frobenius norm
$\mathbf{X}(:, r)$	r^{th} column of $\mathbf{X}$
$\mathbf{X}(r, :)$	r^{th} row of $\mathbf{X}$
$\mathbf{x}(r)$	r^{th} element of $\mathbf{x}$
$\otimes$	Kronecker product
$\odot$	Khatri-Rao product (column-wise Kronecker product [16])

Table 1: Table of symbols and their description

tensor decompositions is the Canonical Polyadic (CP) or CANDECOMP/PARAFAC decomposition [4, 13], 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): $\underline{\mathbf{X}} \approx \sum_{r=1}^{R} \mathbf{A}(:, r) \circ \mathbf{B}(:, r) \circ \mathbf{C}(:, r)$ where $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}$, and the outer product is given by $(\mathbf{A}(:, r) \circ \mathbf{B}(:, r) \circ \mathbf{C}(:, r))(i, j, k) = \mathbf{A}(i, r)\mathbf{B}(j, r)\mathbf{C}(k, r)$ for all i, j, k. The connection of CP to community detection is found in [10] where the authors use each column of factor matrices $\mathbf{A}$ and $\mathbf{B}$ as a community membership indicator for each node.

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.

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

Step 1: Decomposition Given $\underline{\mathbf{X}}, \mathbf{Y}$ compute an $R-1$ component Sparse and Non-negative MTF (as shown below in Section 2.3.1). The columns of $\mathbf{A}$ and $\mathbf{B}$ 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 $\mathbf{A}$, 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 $R-1$ as we show in Section 2.4.

2.3.1 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 $\mathbf{A}, \mathbf{B}$ 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 im-

pose element-wise non-negativity constraints (denoted as $\mathbf{A} \geq 0$) 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:

$$\min_{\mathbf{A}\geq\mathbf{0},\mathbf{B}\geq\mathbf{0},\mathbf{C}\geq\mathbf{0},\mathbf{D}\geq\mathbf{0}} \|\underline{\mathbf{X}} - \sum_r \mathbf{A}(:,r)\circ\mathbf{B}(:,r)\circ\mathbf{C}(:,r)\|_F^2 + \|\mathbf{Y} - \mathbf{A}\mathbf{D}^T\|_F^2 + \lambda\sum_{i,r}|\mathbf{A}(i,r)| + \lambda\sum_{j,r}|\mathbf{B}(j,r)| + \lambda\sum_{k,r}|\mathbf{C}(k,r)| + \lambda_d\sum_{l,r}|\mathbf{D}(l,r)| \tag{2.1}$$

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. 2.1. 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 [16] for a discussion on matricization), then because of properties of the CP/PARAFAC model [16], fixing $\mathbf{B}, \mathbf{C}, \mathbf{D}$ we have

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

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

$$\min_{\mathbf{A}\geq\mathbf{0}} \|[\mathbf{L} - \mathbf{A}\mathbf{M}]\|_F^2 + \lambda\sum_{i,r}|\mathbf{A}(i,r)| \tag{2.2}$$

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 $\mathbf{A}$ [24] 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.

2.4 Overlapping Communities Our goal is to design an algorithm which consumes tensor $\underline{\mathbf{X}} = \{X_1, X_2.....,X_N\}$ along with small amount of labels $\mathbf{Y}$ and it outputs the set of collection of subsets of Nodes V which we considered as overlapping clusters. Thus, we will refer to nodes with multiple classes as overlapping nodes. In real-world networks, these nodes represent bridges between different communities. For this reason, the ability to identify these bridges or overlapping nodes, although often neglected, is necessary for evaluating the accuracy of any community detection algorithms. Given $\underline{\mathbf{X}}$ and $\mathbf{Y}$, CP decomposition is used to learn latent factors which detect community structure.

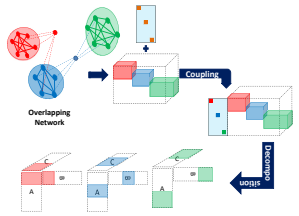


Figure 2: SMACD successfully combines multi-view graph information and semi-supervision.

Multi-view connectivity of tensor and coupling with label matrix can increase the robustness of community detection in the case of highly-mixed communities. Overlapping communities amounts to soft clustering over the nodes, as opposed to hard clustering which forces each node to belong to a unique community. The advocated approach only requires slight modifications in Step 2 to yield soft community assignments, that is, by treating A_{nk} as the normalized affiliation of node n to community k, and requiring $\mathbf{A}_{nk} \geq t$ per node n.

Once the algorithm returns the solution for NNSCMTF, the rows of factor matrix $\mathbf{A}$ provides the community association in networks with overlapping communities where a node can be associated with more than one community. To evaluate SMACD's result with ground truth communities, we compared resultant $A_{i,j}$ with threshold t and node is assigned with community 'r' if $\mathbf{A}(i,j) \geq t$. Each node's predicted label (or labels) is ordered incrementally based on corresponding value of $\mathbf{A}(i,j)$.

$$PredictedLabel(s)\{i\} = \begin{cases} indices(A(i,j)), & \text{if } \mathbf{A}(i,j) \geq t \\ R+1, & \text{otherwise} \end{cases}$$

2.4.1 SelSPF: Automated Selection of the Sparsity Penalty λ The SMACD model contains the λ sparsity penalty, which if not chosen correctly may have an impact on the final result. Traditionally, such parameters are chosen via trial-and-error, and in fact, all the baseline methods that we compare against in Section 3 follow this empirical approach for their parameter tuning. On the other hand, as part of SMACD we introduce SelSPF (based on principle of Armijo-Goldstein's rule [4] for selecting step size in backtracking line search methods), an automated algorithm that selects a "good" value of λ which achieves accuracy which is on-par with a brute force selection of λ based on community detection accuracy, which obviously entails knowing *all* community labels.

The intuition behind SelSPF is simple: Start with a very high λ which gives all-zero community assignments. Start decreasing λ on a logarithmic scale until a solution is reached for which all communities have at least one node assigned to them. Subsequently focus the search on a grid that starts from the previous stopping point and increase λ to the last point before at least one of the communities is empty again. SelSPF is based on the fact that λ and sparsity levels in the latent factors are directly related. We provide a detailed outline of SelSPF in Algorithm 1. Essentially with the introduction of SelSPF, there is no need for hand-tuning SMACD via trial-and-error.

Algorithm 1: SelSPF for automated selection of λ

Input: Tensor $\underline{\mathbf{X}}$, Shared matrix $\mathbf{Y}$, R, initial λ_{high}, τ = Step Size.
Output: Best λ value for our SMACD.
1: Set $\lambda = \lambda_{high}$ and iteration counter j=0.
2: Until $Rank\{f(\underline{\mathbf{X}}, \mathbf{Y}, \lambda_j)\} \geq R$ is satisfied, increment j and set $\lambda_j = \tau\, \lambda_{j-1}$
3: Break (λ_j,λ_{j-1}) into a grid of $\frac{1}{\tau}$ values and repeat step 2 with the λ_{j-1}.
4: **return** λ as the solution.

3 Experimental Evaluation

In this section we extensively evaluate the performance of SMACD on two synthetic and eight 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.

[4] https://en.wikipedia.org/wiki/Backtracking_line_search

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 with 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 then assign different nodes to each block with a probability which is a function of the density of the block (i.e., community) we desire; if this probability is not close to 1, then there will be a considerable amount of nodes falling outside of those blocks, effectively acting as noise. 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 [22] 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 [22]consists of 3312 publications related to AI, DB, IR, ML and HCI research categories. The WebKB dataset [22] 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.

3.1.3 Overlapping community real dataset description The "Insight Resources (IR) Repository" (a.k.a IRR) consists of five multi-view datasets with manual annotation of user stances (e.g., political or sports). Our interest is in Rugby Union dataset[12]. It

is a collection of 854 international Rugby Union players, clubs , and organizations active on Twitter. The ground truth consists of communities corresponding to 15 countries. The communities are overlapping, as players can be assigned to both their home nation and the nation in which they play club rugby. SNOW2014G dataset is first introduced in [21] and author extracted largest connected component and retweet social interactions to form the graph edges from the tweet collection. It consists of top 10992 users, 3 views and clustered them into 90 classes.

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 $\Omega = w_1, w_2, ..., w_r$ to ground truth classes C= $c_1, c_2, ..., c_r$.

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: $ARI(\Omega, c) = \frac{tp+tn}{tp+fp+fn+tn}$, $\omega(c_1, c_2) = \frac{\omega_u(c_1,c_2)-\omega_e(c_1,c_2)}{1-\omega_e(c_1,c_2)}$ where tp, tn, fp and fn are true positive, true negative, false positive and false negative, respectively. Omega index (ω) is the overlapping version of the Adjusted Random Index (ARI). Omega index considers the number of nodes pairs belong together in no clusters, how many are belong together in exactly single cluster or exactly two clusters, and so on. NMI, Purity and ARI (or Omega index for overlapping community) 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. However, moving one step further, we also evaluate SELSPF and demonstrate that an unsupervised selection of parameters yields qualitatively the same performance for SMACD as the brute force selection. 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, 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 as DBLP-I and DBLP-II ($\eta_1 = \eta_2 = 0.01$), for CiteSeer ($\eta_1 = \eta_2 = 1$) , Cora ($\eta_1 = 0.01$, $\eta_2 = 10$) ,WebKB ($\eta_1 = \eta_2 = 0.01$) and MIT ($\eta_1 = 1$, $\eta_2 = 1000$). These η values are chosen to lead to best clustering performance and max 100 iterations are used for reaching the convergence.

Fast Belief Propagation (FaBP) [17]: 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 [15]: 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 [19]: 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 $\leq$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. For overlapping communities, we use $\mathbf{t}$=0.1 for both datasets.

3.4 Experimental Results Below we extensively evaluate SMACD and compare it against baseline

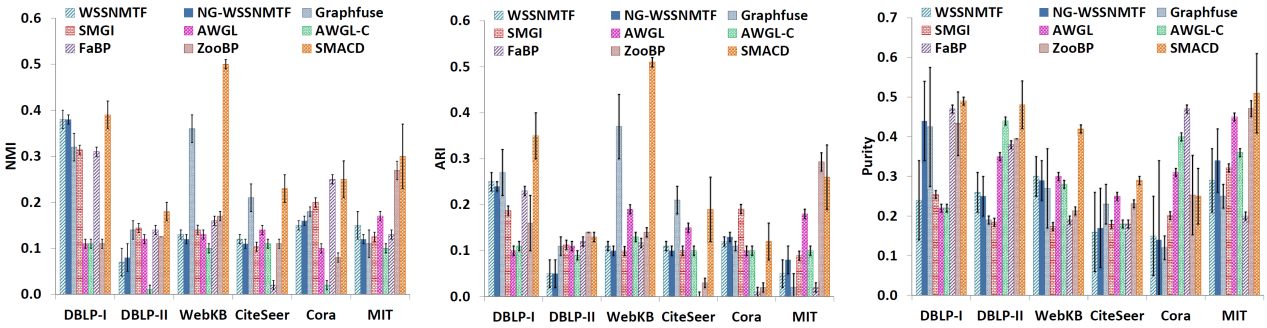


Figure 3: Experimental results for NMI, ARI and Purity. SMACD mostly outperforms baselines and, in particular, works better in very hard scenarios such as the MIT dataset.

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.

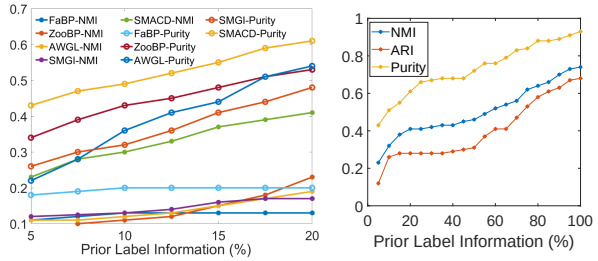
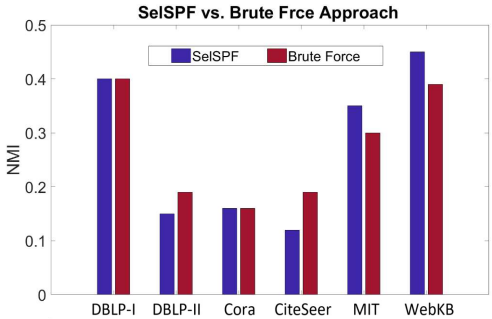


Figure 4: (a) SMACD vs. Guilt-by-Association(FaBP and ZooBP), AWGL and SMGI for different degrees of semi-supervision for DBLP-I. (b) Performance of SMACD as a function of the number of labels. These results confirm the intuition, since performance improves as the number of labels increases.

The most interesting comparison, however, is on the real datasets, since they present more challenging cases than the synthetic ones, shown in Figure 3. 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.



λ	DBLP-I	DBLP-II	Cora	Citeseer	MIT	WebKB
SelSPF	0.3	0.09	1	0.001	600	0.9
Brute force	0.5	0.001	1	0.0001	100	0. 1

Figure 5: λ selection using SelSPF vs. brute force approach. SelSPF is able to choose a value for λ which yields similar accuracy as the expensive and grossly impractical brute force approach, effectively rendering SMACD parameter-free.

3.4.2 SMACD Performance on overlapping communities We report the accuracy of our method for real world Rugby and SNOW2014G datasets with overlapping communities. Figure 6 shows the results. SMACD outpeforms all other state-of-art methods in accuracy (or purity) measures. In particular our method has always the highest purity score and in all but one case it has the best NMI score with small (i.e. 2.5%) number of prior label information. This shows that coupled tensor and matrix allows the overlapping community structure to be more easily and accurately detectable. We ran SMACD and other state-of-art methods for 20 times using 2.5-30% prior label information. In Figure 6 we present our clustering accuracy on both data-sets. For SNOW2014G dataset our algorithm outperformed by predicting cluster with $\approx$ **3x, 4x, 2x and 5x** more accuracy than AWGL, SMGI, ZooBP and FaBP respectively.

3.4.3 Performance vs. Degree of Semi-supervision Next, we evaluate the performance of

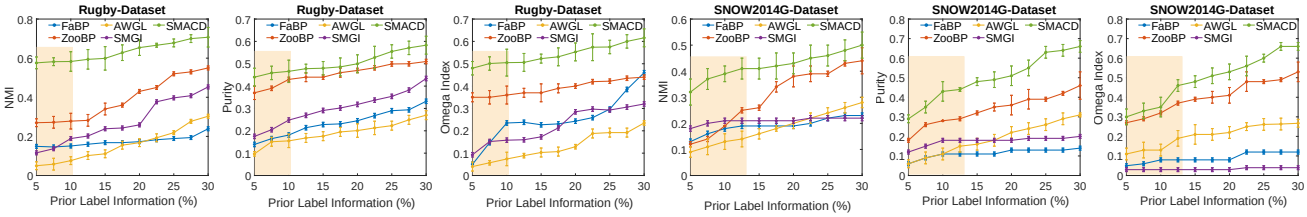


Figure 6: Experimental results for NMI, Purity and Omega index . SMACD consistently outperforms the baseline with an upward trend as the number of available labels increases and works better in *small amounts of labels*.

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 4(a) 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(b) 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.

3.4.4 Evaluation of SelSPF We evaluate the effectiveness of SelSPF in choosing a λ that yields good community quality. We compare SMACD's performance with respect to the λ chosen using a brute force evaluation (on 50 iterations per λ) of the performance according to NMI (using all the labels), against the selection made by SelSPF. Figure 5 demonstrates that, in terms of NMI, both parameter selections in fact yield very comparable (if not identical) performance. This result indicates that SMACD can be used by practitioners as a black box, without the need for specialized and tedious trial-and-error tuning.

3.4.5 Why SMACD? The ability to effectively leverage the multi-view nature of a graph stems from the model that SMACD uses under the hood. The underlying CP model has well-studied uniqueness properties [16] which have implications about the quality of the decomposition, and hence the community assignments. In short, CP is unique under mild conditions, which essentially guarantees that the computed decomposition is the only combination of factors (thus, community assignments) which can reconstruct the data, and not any rotated version thereof. On the other hand, matrix-based approaches, such as [11], typically suffer from rotational ambiguity (this is easy to see since, for a bilinear model, $\mathbf{X} \approx \mathbf{A}\mathbf{B}^T = \mathbf{A}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{B}^T = \tilde{\mathbf{A}}\tilde{\mathbf{B}}^T$ for any invertible $\mathbf{Q}$) and fail to guarantee that the computed community assignments are the best possible, and not any rotation thereof. Finally, coupling with the matrix containing partial community labels is a "soft" manner of imposing semi-supervision. Instead of making hard assignments of the nodes for which we have labels, SMACD is using the underlying structure of the $\mathbf{Y}$ label matrix in order to "guide" the low-rank structure discovered by the CP decomposition on the tensor $\underline{\mathbf{X}}$. Thus, in combination with CP's uniqueness, soft semi-supervision of $\mathbf{Y}$ guides the decomposition to a set of unique community assignments, as close as possible to the partially observed community assignments.

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 [14, 25], 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 [17] 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

The authors would like to thank Evrim Acar for discussions on the NNSCMTF model and Xiaowen Dong for sharing the MIT dataset. Research was supported by the Department of the Navy, Naval Engineering Education Consortium under award no. N00174-17-1-0005. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding parties.

References

[1] E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422*, 2011.

[2] B. W. Bader, T. G. Kolda, et al. Matlab tensor toolbox version 2.6. Available online, 2015.

[3] B. Cao, C.-T. Lu, X. Wei, S. Y. Philip, and A. D. Leow. Semi-supervised tensor factorization for brain network analysis. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 17–32. Springer, 2016.

[4] J. D. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika*, 35(3):283–319, 1970.

[5] W. Cheng, X. Zhang, Z. Guo, Y. Wu, P. F. Sullivan, and W. Wang. Flexible and robust co-regularized multi-domain graph clustering. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 320–328. ACM, 2013.

[6] V. De Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127, 2008.

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

[8] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov. Clustering on multi-layer graphs via subspace analysis on grassmann manifolds. *IEEE Transactions on signal processing*, 62(4):905–918, 2014.

[9] D. Eswaran, S. Günnemann, C. Faloutsos, D. Makhija, and M. Kumar. Zoobp: belief propagation for heterogeneous networks. *Proceedings of the VLDB Endowment*, 10(5), 2017.

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

[11] V. Gligorijevic, Y. Panagakis, and S. Zafeiriou. Fusion and community detection in multi-layer graphs.

[12] D. Greene and P. Cunningham. Producing a unified graph representation from multiple social network views. In *Proceedings of the 5th annual ACM web science conference*, pages 118–121. ACM, 2013.

[13] R. Harshman. Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis. 1970.

[14] H. Jiang, Y. Song, C. Wang, M. Zhang, and Y. Sun. Semi-supervised learning over heterogeneous information networks by ensemble of meta-graph guided random walks.

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

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

[17] D. Koutra, T.-Y. Ke, U. Kang, D. H. P. Chau, H.-K. K. Pao, and C. Faloutsos. Unifying guilt-by-association approaches: Theorems and fast algorithms. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 245–260. Springer, 2011.

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

[19] F. Nie, J. Li, X. Li, et al. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification.

[20] Papalexakis, Evangelos E. Automatic unsupervised tensor mining with quality assessment. In *SIAM SDM*, 2016.

[21] G. Rizos, S. Papadopoulos, and Y. Kompatsiaris. Multilabel user classification using the community structure of online networks. *PloS one*, 12(3):e0173347, 2017.

[22] P. Sen, G. M. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

[23] W. Tang, Z. Lu, and I. S. Dhillon. Clustering with multiple graphs. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, pages 1016–1021. IEEE, 2009.

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

[25] M. Wan, Y. Ouyang, L. Kaplan, and J. Han. Graph regularized meta-path based transductive regression in heterogeneous information network. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 918–926. SIAM, 2015.