UNSUPERVISED CLUSTERING ON SIGNED GRAPHS WITH UNKNOWN NUMBER OF CLUSTERS

Thomas Dittrich and Gerald Matz
Institute of Telecommunications, TU Wien
email: firstname.lastname@tuwien.ac.at

***Abstract*—We consider the problem of unsupervised clustering on signed graphs, i.e., graphs with positive and negative edge weights. Motivated by signed cut minimization, we propose an optimization problem that minimizes the total variation of the cluster labels subject to constraints on the cluster size, augmented with a regularization that prevents clusters consisting of isolated nodes. We estimate the unknown number of clusters by tracking the change of total variation with successively increasing putative cluster numbers. Simulation results indicate that our method yields excellent results for moderately unbalanced graphs.**

I. Introduction

Clustering is one of the most fundamental data science problems, with applications in image segmentation, social network analysis, marketing research, and genre classification for music and films. The goal is to partition a data set into disjoint subsets so that each subset is maximally homogeneous (i.e., objects within the subset are as similar as possible) while distinct subsets are maximally separated (i.e., objects from these subsets are as different as possible). Often, the data set has the structure of a graph (or network) or a graph is constructed via distance metrics on the original data. The clustering is then performed on the graph. Our main interest here is in unsupervised scenarios, where neither the number of clusters nor any cluster labels (affiliation) is known [1]. However, the methods we develop can also be used in a semi-supervised scenario with partially known cluster labels [2], [3], [4].

A typical approach in finding clusters in an (unsigned) graph is to find a partitioning of the node set that minimizes the graph cut, i.e., the number of edges that connect the different partitions. Obviously, the trivial partitioning with one partition being the full set yields a cut value of zero. This problem can be tackled by imposing size or balancing constraints on the partitioning, as done, e.g., with spectral clustering [5], [1] or with total variation (TV) minimization [6]. As an alternative, some applications allow for semi-supervised clustering, where the cluster association is known beforehand for a (typically small) set of nodes, thereby rendering the trivial single-cluster solution infeasible. Examples of such approaches include clustering by harmonic functions [2] and manifold regularization [7].

Some of the previously mentioned approaches have been extended to signed graphs, e.g., signed spectral clustering [8], [9] and manifold regularization [10]. In our preceding work [11], [12], we demonstrated that semi-supervised TV minimization on signed graphs yields tighter relaxations of the cut minimization problem.

In this paper, we develop an unsupervised extension of the signed TV minimization algorithm for clustering from [12]. Specifically, we incorporate upper bounds for the size of the individual clusters, similar in spirit to [13]. We use a neighborhood regularization to prevent clusters consisting of isolated nodes and we introduce a recursive cluster refinement strategy which allows us to estimate the unknown number of clusters by tracking the change in TV. Simulation results with signed stochastic block models confirm the excellent performance of our approach on moderately unbalanced graphs.

II. Signed Graph Clustering

Let $\mathcal{G}$ be a graph with vertex set $\mathcal{V} = \{1, \dots, N\}$ and signed, weighted adjacency matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$. In such graphs, a positive weight $W_{ij} > 0$ indicates similarity between nodes i and j while a negative weight $W_{ij} > 0$ indicates dissimilarity.

Consider a partition of $\mathcal{V}$ into K disjoint sets $\mathcal{V}_1, \dots, \mathcal{V}_K$. The signed graph cut is defined as

$$\gamma\{\mathcal{V}_1, \dots, \mathcal{V}_K\} = \sum_{k=1}^{K} \tilde{\gamma}\{\mathcal{V}_k\} \tag{1}$$

with (here, $(W)_+ = \max\{0, W\}$)

$$\tilde{\gamma}\{\mathcal{V}_k\} = \sum_{i \in \mathcal{V}_k} \left(\sum_{j \in \mathcal{V} \setminus \mathcal{V}_k} (W_{ij})_+ + \sum_{j \in \mathcal{V}_k} (-W_{ij})_+ \right).$$

Thus, $\gamma\{\mathcal{V}_1, \dots, \mathcal{V}_K\}$ measures the sum of negative edge weights within the individual sets and the sum of positive edges between distinct sets. A signed graph is called k-balanced [14] if there exists a partition $\mathcal{V}_1, \dots, \mathcal{V}_K$ such that $\gamma\{\mathcal{V}_1, \dots, \mathcal{V}_K\} = 0$, i.e., there are no negative edges within each set and no positive edges between distinct sets. In other words, k-balanced graphs are perfectly clusterable.

If a signed graph is not balanced, it is natural to try to find the most reasonable clusters by minimizing the signed cut, which leads to the following combinatorial optimization problem:

$$\begin{aligned} \min_{\mathcal{V}_1, \dots, \mathcal{V}_K} \ & \gamma\{\mathcal{V}_1, \dots, \mathcal{V}_K\}, \\ \text{s.t. } & \mathcal{V}_k \cap \mathcal{V}_l = \emptyset \text{ for } k \neq l, \\ & \bigcup_{k=1}^{K} \mathcal{V}_k = \mathcal{V}, \\ & \mathcal{L}_k \subset \mathcal{V}_k, \ k = 1, \dots, K. \end{aligned}$$

The last line enables a semi-supervised approach by incorporating the sets $\mathcal{L}_k$, which comprise those nodes which are known a priori to belong to cluster k. In an unsupervised setting, these sets are empty and the last constraint becomes void.

III. Size-Constrained TV Clustering

We next formulate a computationally tractable relaxation of the clustering problem via signed TV. We use a binary-valued one-hot encoding $\mathbf{x}_i^{\mathrm{T}} = (X_{i1}, \ldots, X_{iK}) \in \{-1, 1\}^K$ for the cluster affiliation of node i. If node i belongs to cluster k then $X_{ik} = 1$ and $X_{il} = -1$, $l \neq k$. Equivalently, the cluster label vector is normalized as $\mathbf{x}_i^{\mathrm{T}}\mathbf{1} = 2 - K$ and has the single $+1$ entry at position k, i.e., $\mathbf{x}_i = 2\mathbf{e}_k - \mathbf{1}$ (here, $\mathbf{e}_k$ is the kth unit vector). With $S_{ij} = \operatorname{sign}(W_{ij})$, we measure the similarity/dissimilarity of cluster labels $\mathbf{x}_i^{\mathrm{T}}$ and $\mathbf{x}_j^{\mathrm{T}}$ across positive/negative edges by

$$D(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{K} (X_{ik} - S_{ij} X_{jk})_+.$$

Nodes that have the same cluster label and are connected by a positive edge and nodes that have distinct cluster labels and are connected by a negative edge have $D(\mathbf{x}_i, \mathbf{x}_j) = 0$. By contrast, nodes that have the same cluster label and are connected by a negative edge and nodes that have distinct cluster labels and are connected by a positive edge have $D(\mathbf{x}_i, \mathbf{x}_j) = 2$. The TV for all cluster labels $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^{\mathrm{T}}$ is then defined as [12]

$$\mathrm{TV}(\mathbf{X}) = \sum_{i=1}^{N} \sum_{j=1}^{N} |W_{ij}|\, D(\mathbf{x}_i, \mathbf{x}_j). \tag{2}$$

With the relation $\mathcal{V}_k = \{i : \ X_{ik} = 1\}$, it follows that for $i \in \mathcal{V}_l$ and $j \in \mathcal{V}_m$

$$|W_{ij}| D(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 2(W_{ij})_+, & l \neq m, \\ 2(-W_{ij})_+, & l = m. \end{cases}$$

By that, $\mathrm{TV}(\mathbf{X}) = 2\,\gamma\{\mathcal{V}_1, \ldots, \mathcal{V}_K\}$ and thus minimizing the TV over all admissible cluster labels is equivalent to minimizing the signed cut.

In order to avoid vastly different cluster sizes, we impose size constraints inspired by [13]. In that paper, the unsigned TV objective was augmented with piece-wise linear penalty terms for exceeding upper and lower bounds on cluster size. We rather advocate incorporating upper size limits via hard side constraints. More specifically, we postulate that the kth cluster consists of at most u_k vertices. In terms of the one-hot encoded cluster labels X_{ik}, these constraints read $\sum_{i=1}^{N} (X_{ik})_+ \leq u_k$, $k = 1, \ldots, K$.

We can now reformulate the signed cut minimization problem—augmented with cluster size constraints—as a TV minimization problem. The main remaining issue is the fact that the label domain $\mathbf{X} \in \{-1, 1\}^{N \times K}$ is discrete and thus not convex. We therefore use box constraint $\mathbf{X} \in \mathcal{X} = [-1, 1]^{N \times K}$ as relaxation to arrive at the convex signed TV clustering problem

$$\begin{aligned} \min_{\mathbf{X} \in \mathcal{X}} \ & \mathrm{TV}(\mathbf{X}), \\ \text{s.t. } & \sum_{k=1}^{K} X_{ik} = 2 - K, \ i = 1, \ldots, N, \\ & X_{ik} = 1 \text{ for } i \in \mathcal{L}_k, \\ & \sum_{i=1}^{N} (X_{ik})_+ \leq u_k, \ k = 1, \ldots, K. \end{aligned} \tag{3}$$

The final cluster label vectors are obtained as the signs of the solution $\mathbf{X}$ of this problem. Note that in order for (3) to yield a complete partition of $\mathcal{V}$, it is necessary that $\sum_{k=1}^{K} u_k \geq N$.

In an unsupervised setting, we choose to associate an arbitrary node i with the first cluster, i.e., $\mathcal{L}_1 = \{i\}$, while all other known label sets remain empty, $\mathcal{L}_2 = \cdots = \mathcal{L}_K = \emptyset$. This also prevents the trivial solution $\mathbf{X} = \mathbf{0}$ in the case $K = 2$ and all-negative labels in the general case. Furthermore, the cluster size constraints are particularly instrumental in the unsupervised case to avoid extremely unbalanced clustering results.

IV. ADMM Solution

A. Basic Algorithm

We next derive an augmented ADMM [15] algorithm to solve the signed TV minimization problem. To this end, we capture the cluster size constraints by

$$\mathcal{P} = \left\{ \mathbf{X} \in \mathbb{R}^{N \times K} : \sum_{i=1}^{N} (X_{ik})_+ \leq u_k, \ k = 1, \ldots, K \right\},$$

and the remaining label vector constraints via

$$\mathcal{Q} = \left\{ \mathbf{X} \in [-1, 1]^{N \times K} : \ X_{ik} = 1 \text{ for } i \in \mathcal{L}_k, \ \sum_{k=1}^{K} X_{ik} = 2 - K, \ i = 1, \ldots, N \right\}.$$

We can then rephrase (3) in ADMM standard form,

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{Z}} \ & \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{K} ([\mathbf{Z}_k]_{ij})_+ + \chi_{\mathcal{P} \cap \mathcal{Q}}(\mathbf{X}), \\ \text{s.t. } & \nabla_{\mathcal{G}} \tilde{\mathbf{x}}_k = \mathbf{Z}^{(k)}, \ k = 1, \ldots, K, \end{aligned} \tag{4}$$

where $\tilde{\mathbf{x}}_k$ denotes the kth column of $\mathbf{X}$, $\chi_{\mathcal{P} \cap \mathcal{Q}}(\mathbf{X})$ is the indicator function of $\mathcal{P} \cap \mathcal{Q}$, and we used the signed graph gradient [12]

$$[\nabla_{\mathcal{G}} \mathbf{x}_k]_{ij} = (X_{ik} - S_{ij} X_{jk}) |W_{ij}|.$$

The steps of the augmented ADMM are summarized in Algorithm 1. The update in line 7 uses the graph divergence, which is the adjoint of the graph gradient, defined as [12]

$$-[\operatorname{div}_{\mathcal{G}} \mathbf{Z}_k]_i = \sum_{j=1}^{N} |W_{ij}| [\mathbf{Z}_k]_{ij} - S_{ji} |W_{ji}| [\mathbf{Z}_k]_{ji}.$$

Algorithm 1 — Multiclass signed TV clustering

Input: $\mathbf{W}$, $\mathcal{L}_1, \dots, \mathcal{L}_K$

Initialization

$d_i = 2\sum_{j=1}^{N}(W_{ij}^2 + W_{ji}^2)$

$\mathbf{D} = \operatorname{diag}\{d_1, \dots, d_N\}$

$\mathbf{x}_i^{(0)} = \begin{cases} 2\mathbf{e}_k - \mathbf{1}, & i \in \mathcal{L}_k, \\ \mathbf{0}, & \text{otherwise} \end{cases}$

$\mathbf{V}_k^{(0)} = \mathbf{0}$

$\rho = 0.1$

$t = 0$

Iterations

repeat

$\mathbf{Y}_k^{(t+1)} = -\operatorname{div}_{\mathcal{G}}\left(2\mathbf{V}_k^{(t)} - \mathbf{V}_k^{(t-1)}\right)$

$\tilde{\mathbf{x}}_k^{(t+1)} = \mathbf{x}_k^{(t)} - \frac{1}{\rho}\mathbf{D}^{-1}\mathbf{y}_k^{(t+1)}$

$\mathbf{X}^{(t+1)} = \pi_{\mathcal{P}\cap\mathcal{Q}}\left(\tilde{\mathbf{X}}^{(t+1)}\right)$

$\tilde{\mathbf{V}}_k^{(t+1)} = \mathbf{V}_k^{(t)} + \rho\nabla_{\mathcal{G}}\mathbf{x}_k^{(t+1)}$

$\mathbf{V}_k^{(t+1)} = \pi_{[0,1]^{N\times N}}\left(\tilde{\mathbf{V}}_k^{(t+1)}\right)$

$t = t + 1$

until stopping criterion is satisfied

Output: $\hat{\mathbf{X}} = \mathbf{X}^{(t)}$

We choose the augmenting variable (cf. line 9) as

$$\mathbf{D} = \operatorname{diag}\{d_1, \dots, d_N\}, \quad d_i = 2\sum_{j=1}^{N}(W_{ij}^2 + W_{ji}^2). \tag{5}$$

Note that $\mathbf{D} \succeq -\operatorname{div}_{\mathcal{G}}\nabla_{\mathcal{G}}$. Furthermore, $\pi_{\mathcal{P}\cap\mathcal{Q}}$ in line 10 denotes the projection onto the constraint set $\mathcal{P}\cap\mathcal{Q}$, which can be computed via Dykstra's algorithm [16] by alternately projecting the columns and rows of $\mathbf{X}$ onto the simplices $\mathcal{P}$ and $\mathcal{Q}$, e.g., using the algorithm from [17].

Analogously to [12], we use the stopping criteria $\sum_{k=1}^{K}\|\mathbf{R}_k^{(t+1)}\|_{\mathrm{F}} \le \epsilon^{\mathrm{p}}$ and $\sum_{k=1}^{K}\|\mathbf{s}_k^{(t+1)}\|_2 \le \epsilon^{\mathrm{d}}$, with the primal and dual residuals

$$\begin{aligned}\mathbf{R}_k^{(t+1)} =& \frac{1}{\rho}\left(\mathbf{V}_k^{(t+1)} - \mathbf{V}_k^{(t)}\right),\\ \mathbf{s}_k^{(t+1)} =& -\rho\operatorname{div}_{\mathcal{G}}\nabla_{\mathcal{G}}(\mathbf{x}_k^{(t+1)} - \mathbf{x}_k^{(t)})\\ & - \operatorname{div}_{\mathcal{G}}(2\mathbf{V}_k^{(t)} - \mathbf{V}_k^{(t-1)} - \mathbf{V}_k^{(t+1)}),\end{aligned}$$

and

$$\begin{aligned}\epsilon^{\mathrm{p}} =& \epsilon^{\mathrm{abs}}\sqrt{KN} + \epsilon^{\mathrm{rel}}\sum_k \|\nabla_{\mathcal{G}}\mathbf{x}_k^{(t+1)}\|_{\mathrm{F}},\\ \epsilon^{\mathrm{d}} =& \epsilon^{\mathrm{abs}}\sqrt{KN} + \epsilon^{\mathrm{rel}}\sum_k \|\operatorname{div}_{\mathcal{G}}\mathbf{V}_k^{(t+1)}\|_2.\end{aligned}$$

Algorithm 2 — Regularized signed TV clustering

Input: $\mathbf{W}$, $\mathcal{L}_1, \dots, \mathcal{L}_K$, $x_{\min}$

Initialization

$\widetilde{\mathbf{W}} = \mathbf{W}$

$\lambda_i^+ = \lambda_i^- = 1$ for $i \in \mathcal{L}_k$

Iterations

repeat

$\mathbf{X}$ =Algorithm 1($\widetilde{\mathbf{W}}$, $\mathcal{L}_1, \dots, \mathcal{L}_K$)

$\mathcal{M}_i^+ = \{j \in \mathcal{N}_i^+ : X_{jk} > 0\}$

$\mathcal{M}_i^- = \{j \in \mathcal{N}_i^- : X_{jk} < 0\}$

if $|\mathcal{M}_i^+| = \emptyset$ or $\min\{X_{jk} : j \in \mathcal{M}_i^+\} \le x_{\min}$

$\widetilde{W}_{ij} = \begin{cases} W_{ij} + \lambda_i^+, & j \in \mathcal{N}_i^+, \\ W_{ij}, & \text{otherwise.} \end{cases}$

$\lambda_i^+ = 2\lambda_i^+$

if $|\mathcal{M}_i^-| = \emptyset$ or $\min\{X_{jk} : j \in \mathcal{M}_i^-\} \ge -x_{\min}$

$\widetilde{W}_{ij} = \begin{cases} W_{ij} - \lambda_i^-, & j \in \mathcal{N}_i^+, \\ W_{ij}, & \text{otherwise.} \end{cases}$

$\lambda_i^- = 2\lambda_i^-$

until no more regularization needed

Output: $\mathbf{X}^{(m)}$

Furthermore, convergence is improved with the varying penalty strategy [15], [18]

$$\rho = \begin{cases} \eta\rho, & \sum_k \epsilon^{\mathrm{d}}\|\mathbf{R}_k^{(t+1)}\|_{\mathrm{F}} - \mu\epsilon^{\mathrm{p}}\|\mathbf{s}_k^{(t+1)}\|_2 \ge 0, \\ \eta^{-1}\rho, & \sum_k \epsilon^{\mathrm{d}}\|\mathbf{R}_k^{(t+1)}\|_{\mathrm{F}} - \mu^{-1}\epsilon^{\mathrm{p}}\|\mathbf{s}_k^{(t+1)}\|_2 \le 0, \end{cases}$$

where η and μ are suitable constants.

B. Regularization

In the semi-supervised setting, there is the danger that pre-labeled nodes are isolated as individual clusters. This can be prevented via a regularization strategy similar to [12], however, here we allow for different weights for each labeled node and we add regularization parameters for dissimilar neighbors. Define the neighborhood of a labelled node $i \in \mathcal{L}_k$ as

$$\begin{aligned}\mathcal{N}_i^+ &= \{j : W_{ij} > 0 \text{ and } \forall \tilde{k} \ne k, l \in \mathcal{L}_{\tilde{k}} : W_{il} \le 0\}\\ \mathcal{N}_i^- &= \{j : W_{ij} < 0 \text{ and } \forall l \in \mathcal{L}_k : W_{il} \le 0\}\end{aligned} \tag{6}$$

It is defined as a penalty term

$$\begin{aligned}\sum_{k=1}^{K}\sum_{i\in\mathcal{L}_K}\sum_{j\in\mathcal{N}_i^+}\lambda_i^+(X_{ik} - X_{jk})\\ \sum_{j\in\mathcal{N}_i^-}\lambda_i^-(X_{ik} + X_{jk}),\end{aligned} \tag{7}$$

which can easily be incorporated into Algorithm 1 by adding the individual λ_i^+ and λ_i^- to the weights of the edges (i, j) for $j \in \mathcal{N}_i^+$ and $j \in \mathcal{N}_i^-$, respectively. The algorithm for this regularization is summarized in Algorithm 2.

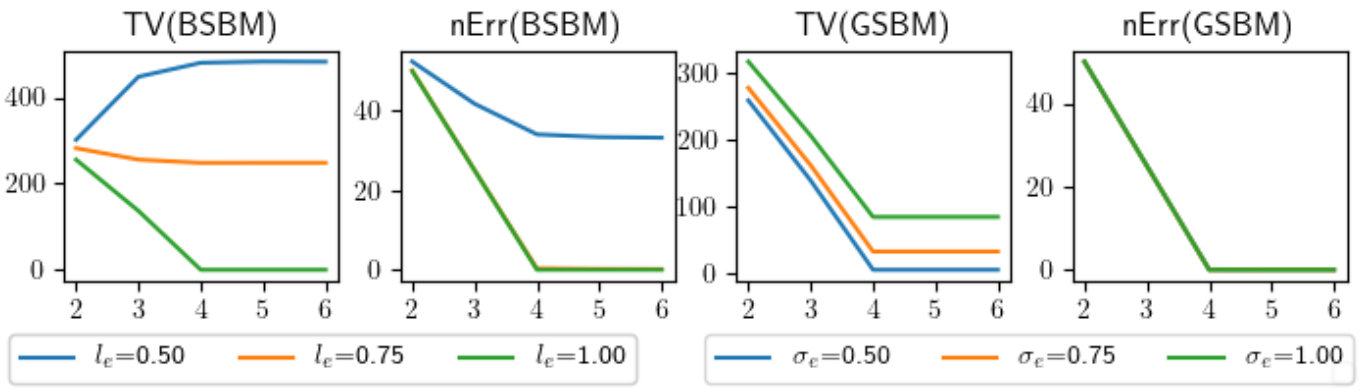

Fig. 1. TV and number of falsely clustered nodes versus K for BSBM with $l_e \in \{0.5, 0.75, 1\}$ (left) and GSBM with $\sigma_e \in \{0.5, 0.75, 1\}$ (right).

C. Detecting the Number of Clusters

To motivate our approach, consider a balanced graph with K_0 clusters. If we cluster such a graph using $K = K_0$ classes, the minimum TV (by definition) is zero. For $K < K_0$, we will incur a larger TV since some clusters necessarily will suffer from negative intra-cluster edges. Similarly, for $K > K_0$ some of the true clusters will start to get separated (by cutting positive inter-cluster edges), again leading to higher TV (assuming clusters may not be empty). Evidently, the minimum TV as a function of K will have a minimum for $K = K_0$. If empty clusters are admissible, no further partitioning will happen for $K \geq K_0$ and the TV will remain constant. A similar behavior is conjectured to hold true for (weakly) unbalanced graphs (for signed stochastic block models, we verified this conjecture experimentally).

Based on this observation, we propose to estimate the true number of clusters via a cluster refinement scheme in which we repeatedly run Algorithm 2 with successively increasing K and choose the estimate $\hat{K}$ of the number of clusters as that K for which the TV is smallest. In practice, we stop as soon as increasing K does no longer decrease the TV.

With an unknown number of clusters we are in an unsupervised scenario without pre-existing labels. However, when running our clustering algorithm with K tentative clusters, we have the results for $K-1$ at our disposal. Assume we have a known label for each of the clusters 1 to $K-1$. We then augment the known label set by picking the strongest label for cluster K, i.e.,

$$\begin{aligned} \mathcal{L}_K &= \{i(K)\}, \quad \text{with} \\ i(K) &= \arg\max_i \quad X_{iK}, \\ &\text{s.t.} \quad X_{jK} > 0 \text{ for } W_{ij} > 0. \end{aligned}$$

When starting the clustering with $K = 2$, we use the (arbitrary) initialization $\mathcal{L}_1 = \arg\max_i \sum_{j=1}^{N} \mathrm{I}\{W_{ij} < 0\}$.

V. Experiments

In our experiments we used a generalization of the stochastic block model from [19], which is characterized by the edge probabilities p_i ("intra") and p_e ("extra") for edges within clusters and between clusters, respectively. Our extension allows for multiple clusters of different size and for signed edge weights. In the Bernoulli stochastic block model (BSBM) intra- and inter-cluster edges have negative edge weight ($W = -1$) with respective probability l_i and l_e. In the Gaussian stochastic block model (GSBM), the distribution for intra- and inter-cluster edges weights respectively are $\mathcal{N}(\mu_i, \sigma_i^2)$ and $\mathcal{N}(\mu_e, \sigma_e^2)$. Our results have been obtained by averaging over 1000 model realizations.

A. Cluster Refinement

The first experiment illustrates the iterative cluster refinement procedure for a BSVM/GSBM with $N = 100$ nodes and $K_0 = 4$ clusters of size 25 and intra-/extra-cluster edge probabilities $p_i = 4/5$ and $p_e = 1/15$ (on average, a node is connected to 20 nodes from the same cluster and to 5 nodes from a different cluster). All intra-cluster edge weights equal 1. The extra-cluster edge weights are controlled by l_e for BSBM and by σ_e for GSBM ($\mu_e = -1$). Note that l_e and σ_e determine how unbalanced the graphs are ($l_e = 1$ and $\sigma_e = 0$ yield balanced graphs).

The TV and the number of incorrectly clustered nodes obtained with our iterative cluster refinement procedure (cf.IV-C) are shown in Fig. 1. It can be seen that for high l_e and low σ_e (i.e., weakly unbalanced graphs) the TV and the number of errors has a minimum at $K = K_0 = 4$. Obviously, for $K < 4$ some nodes necessarily have to be misclassified. For $K > 4$, the curves flatten out, indicating that clusters 5 and 6 essentially remain empty. For $l_e = 0.5$, the graphs are strongly unbalanced and thus no longer meaningfully clusterable — this is reflected in poor clustering results and the lack of a minimum in the TV.

B. Estimation of K

The second experiment analyzes the performance of estimating the number of clusters for different ground truth K. We use the GSBM with clusters of size 25, $p_i = 4/5$, $p_e = 2/(5(K-1))$. All intra-cluster edges have weight 1 and extra-cluster edges have $\mu_e = -1$ and varying σ_e.

Fig. 2 shows the relative estimation error (mean and standard deviation) for the number of clusters versus the graph

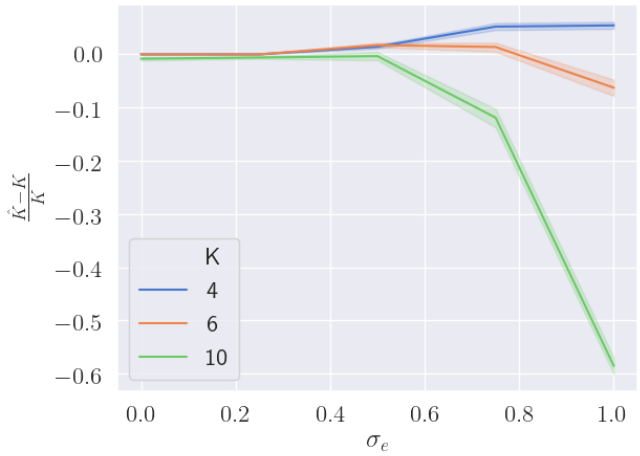

Fig. 2. Relative estimation error $(\hat{K} - K)/K$ versus graph imbalance parameter σ_e for $K \in \{4, 6, 10\}$.

imbalance parameter σ_e for $K \in \{4, 6, 10\}$. The estimation of K is highly accurate for small σ_e but starts to deteriorate beyond $\sigma_e = 0.5$, specifically for larger K. Note that in this regime of strongly unbalanced graphs with many clusters, our method tends to underestimate K since there is a large number of positive edges connecting different clusters that make it difficult to tell the clusters apart.

C. Comparison with Spectral Clustering

In Table I we compare our Algorithm 2 and signed spectral clustering (SC) [8] in terms of misclassification rate (number of incorrectly clustered nodes). We consider a GSBM with $N = 1000$ nodes whose cluster structure is mainly determined by negative edges, i.e., $p_i < p_e$. Specifically, we choose $p_i = 0.01$, $p_e = 0.01$, $K = 2$, $\mu_e = -1$ and different σ_e. We consider cases with equal and strongly different cluster size.

When the graph is almost balanced (small σ_e) TV and SC both achieve almost perfect clustering results. Also, for the case of equal cluster size SC and TV feature essentially the same performance. However, when the graph is more strongly unbalanced (larger σ_e) and cluster sizes are substantially different, our method maintains accurate performance and beats SC (which eventually breaks down completely) by large margins.

VI. Conclusion

We introduced a TV based algorithm for unsupervised clustering in signed graphs without knowledge of the number of clusters. The method controls cluster size via explicit upper bounds and neighborhood regularization. We estimate the number of clusters by iteratively refining the graph partition and tracking the change in TV. Our experiments verified that our approach succeeds in estimating the number of clusters and in correctly clustering the nodes provided the graph is not too unbalanced. We also found that our algorithm outperforms signed spectral clustering for graphs with uneven cluster sizes.

TABLE I
Average misclassification error for spectral clustering and TV clustering (Algorithm 2) for SSBM with different cluster sizes and different imbalance σ_e.

	[500, 500]		[200, 800]		[50, 950]	
σ_e	SC	TV	SC	TV	SC	TV
0.50	0.54	0.01	0.03	0.03	1.01	0.12
0.62	3.78	0.10	1.43	0.09	4.12	0.27
0.75	10.99	1.91	2.79	0.77	11.12	0.51
0.88	15.69	5.94	8.70	3.39	23.25	0.84
1.00	33.47	25.73	15.20	6.50	36.93	1.48

References

[1] U. von Luxburg, "A tutorial on spectral clustering," *Stat. and Comp.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[2] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Int. Conf. Mach. Learn. (ICML-03)*, Washington, DC, USA, Aug. 2003, pp. 912–919.

[3] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proc. Int. Conf. Mach. Learn.*, San Francisco, CA, USA, July 2001, pp. 19–26.

[4] W. Liu, J. Wang, and S. Chang, "Robust and scalable graph-based semisupervised learning," *Proc. IEEE*, vol. 100, no. 9, pp. 2624–2638, Sept. 2012.

[5] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Adv. in Neural Inf. Process. Sys.*, 2002, pp. 849–856.

[6] X. Bresson, T. Laurent, D. Uminsky, and J. von Brecht, "Multiclass total variation clustering," in *Adv. in Neural Inf. Process. Sys. 26*, Lake Tahoe (NV), Dec. 2013, pp. 1421–1429.

[7] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. of Mach. Learn. research*, vol. 7, no. Nov, pp. 2399–2434, 2006.

[8] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak, "Spectral analysis of signed graphs for clustering, prediction and visualization," in *Proc. SIAM Int. Conf. Data Mining*, Columbus (OH), May 2010, pp. 559–570.

[9] K.-Y. Chiang, J. J. Whang, and I. S. Dhillon, "Scalable clustering of signed networks using balance normalized cut," in *Proc. ACM Int. Conf. Inf. and Knowledge Management*, 2012, pp. 615–624.

[10] A. B. Goldberg, X. Zhu, and S. Wright, "Dissimilarity in graph-based semi-supervised classification," in *Proc. Int. Conf. on Art. Int. and Stat.*, San Juan (Puerto Rico), Mar. 2007, pp. 155–162.

[11] P. Berger, T. Dittrich, and G. Matz, "Semi-Supervised Clustering Based on Signed Total Variation," in *Proc. IEEE Global Conf. on Sig. and Inf. Process. (GlobalSIP)*, 2018, pp. 793–797.

[12] P. Berger, T. Dittrich, G. Hannak, and G. Matz, "Semi-Supervised Multiclass Clustering Based on Signed Total Variation," in *Proc. IEEE Int. Conf. Acoustics, Speech and Sig. Process. (ICASSP)*, Brighton, UK, May 2019, pp. 4953–4957.

[13] E. Bae and E. Merkurjev, "Convex variational methods on graphs for multiclass segmentation of high-dimensional data and point clouds," *J. of Math. Imaging and Vision*, vol. 58, no. 3, pp. 468–493, July 2017.

[14] J. A. Davis, "Clustering and structural balance in graphs," *Human relations*, vol. 20, no. 2, pp. 181–187, 1967.

[15] Y. Zhu, "An Augmented ADMM Algorithm With Application to the Generalized Lasso Problem," *J. Comp. Graph. Stat.*, vol. 26, no. 1, pp. 195–204, Feb. 2017.

[16] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in hilbert spaces," in *Adv. in Order Restricted Statistical Inference*, R. Dykstra, T. Robertson, and F. T. Wright, Eds. New York, NY: Springer New York, 1986, pp. 28–47.

[17] W. Wang and M. Á. Carreira-Perpiñán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," *CoRR*, vol. abs/1309.1541, 2013.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *FnT Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[19] S. Heimlicher, M. Lelarge, and L. Massoulié, "Community detection in the labelled stochastic block model," *arXiv:1209.2910*, 2012.