Unveiling Anomalous Edges and Nominal Connectivity of Attributed Networks

Konstantinos D. Polyzos[*,1], Costas Mavromatis[*,2], Vassilis N. Ioannidis[1], and Georgios B. Giannakis[1]

[1]Department of Electrical and Computer Engineering, University of Minnesota, USA
[2]Department of Computer Science, University of Minnesota, USA

Abstract—Uncovering anomalies in attributed networks has recently gained popularity due to its importance in unveiling outliers and flagging adversarial behavior in a gamut of data and network science applications including the Internet of Things (IoT), finance, security, to list a few. The present work deals with uncovering anomalous edges in attributed graphs using two distinct formulations with complementary strengths, which can be easily distributed, and hence efficient. The first relies on decomposing the graph data matrix into low rank plus sparse components to markedly improve performance. The second broadens the scope of the first by performing robust recovery of the unperturbed graph, which enhances the anomaly identification performance. The novel methods not only capture anomalous edges linking nodes of different communities, but also spurious connections between any two nodes with different features. Experiments conducted on real and synthetic data corroborate the effectiveness of both methods in the anomaly identification task.

Index Terms—Anomaly identification, Low-Rank and Sparse Decomposition, Feature Smoothness, Laplacian Recovery

I. INTRODUCTION

Unveiling rare or anomalous data is of paramount importance in numerous applications such as recommender systems, finance, security, health care, autonomous driving and insurance [1]. Anomaly identification methods can spot network intrusions and failures, spam reviews, credit card frauds, malware and rare disease outbreaks.

The present work focuses on anomalies that appear in data adhering to a graph structure. Consider a graph that consists of nodes and edges, where each node is associated with an attributed vector. In an IoT network for example, the nodes represent mobile devices and the edges capture the communication among devices. The graph connectivity may be perturbed e.g. by an adversary who adds edges possibly at random. Examples of such anomalous edges include links among nodes belonging to different communities or among nodes with uncorrelated attributes. Leveraging the graph structure and the correlation of nodal features will enable identification of anomalous edges.

Anomalies can emerge in either static or dynamic graphs. Dynamic ones comprise sequences of graphs and the corresponding methods leverage temporal correlation too [2], [3],

[4], [5], [6], [7], [8]. The focus here will be on static graphs, but dynamic graphs are also included in our future research agenda.

Most anomaly detection methods for static graphs pursue community-based approaches. They primarily rely on matrix factorization techniques that decompose the graph matrix into a low rank component and a residual matrix [9], [10]. Matrix factorization is widely employed for dimensionality reduction [11], [12] and (graph) clustering [13]. The approach in [10] spots anomalous connections if they appear in the residual matrix, after the non-negative matrix factorization.

Albeit interesting, the aforementioned methods do not account for attributes. On the other hand, approaches have been reported that incorporate nodal features to unveil anomalous nodes [14], [15], [16], [17], [18], [19], [20]. These are suitable to spot anomalous nodes within a certain community, but not anomalous edges across communities [21].

Contribution. The present work proposes two novel approaches that rely on low rank and sparse factorization, and judiciously account for nodal attributes to spot anomalous edges in a given perturbed graph. The second approach further recovers the ground-truth topology, which turns out to markedly enhance identification performance. Accounting for smoothness across nodal features also boosts identification performance. Different from existing approaches, the novel methods not only identify anomalous links between different communities, but also mendacious edges between any two nodes based on the nodal features. Finally, two optimization methods with complementary strengths are developed to unveil the anomalous edges and robustly recover the underlying nominal topology. Although the first method is computationally more efficient, the second unveils more accurately the perturbed edges. Complexity analysis of the developed methods is also provided.

II. UNVEILING ANOMALOUS EDGES

Consider a graph that consists of N nodes and E edges connecting pairs of nodes. Connectivity is captured by the $N \times N$ adjacency matrix $\mathbf{A}$. The initial (a.k.a. nominal) graph is perturbed by inserting (anomalous) edges linking nodes with correlated or uncorrelated nodal features. Let $\tilde{\mathbf{A}}$ denote the perturbed adjacency, and $\tilde{\mathcal{E}}$ the set of perturbed edges. The $N \times N$ Laplacian matrix is

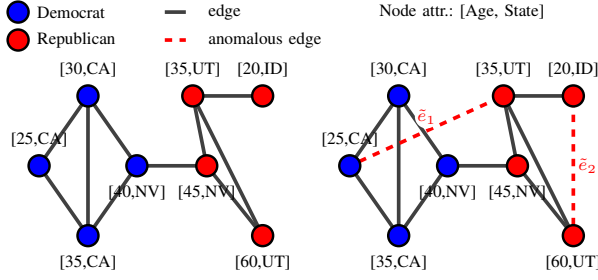$$\mathbf{L} := \mathrm{diag}(\mathbf{A1}) - \mathbf{A} \qquad (1)$$

Fig. 1: Original (left) and perturbed (right) friendship network of voters. Each voter is associated with attributes (age, state); and, same color voters belong to the same community (Democrat, Republic). Anomalous edges may appear between nodes of different communities (anomalous edge $\tilde{e}_1$) or between nodes with uncorrelated attributes (anomalous edge $\tilde{e}_2$).

where $\mathbf{1}$ is the all-ones vector and $\mathrm{diag}(\cdot)$ is a diagonal matrix.

Each node n also holds an $F \times 1$ feature vector $\mathbf{x}_n$, and all feature vectors are collected in the $N \times F$ feature matrix $\mathbf{X}$. For example, in a network that consists of voters, attributes may correspond to each voter's age and state of residency.
Goal. Given the perturbed Laplacian matrix $\tilde{\mathbf{L}} \in \mathbb{R}^{N \times N}$ and the feature matrix $\mathbf{X}$, this paper aims to unveil the anomalous edges in $\tilde{\mathcal{E}}$ and recover the unperturbed Laplacian matrix $\mathbf{L}$. Fig. 1 shows a perturbed social voting network.

A. Anomalous edges across communities

Anomalous edges often connect nodes belonging to different communities [1], [7], [9], [10]. For example, in a typical social voting network, each node is either Republican or Democrat, and connections between nodes across the isle are rare [1].

Factorizing the Laplacian matrix of a graph is known to reveal the community structure [7], [13], [10], [22]. Specifically, the low-rank approximation of the Laplacian yields the communities, while the low rank captures the number of communities. To retrieve anomalous edges connecting different clusters, it is possible to augment the decomposition with a sparse matrix containing edges that connect different communities, and hence do not obey the low-rank model [3], [7], [10]. Upon accounting for the low-rank approximation matrix $\mathbf{R}$ and the sparse matrix $\mathbf{S}$, our formulation amounts to

$$\min_{\mathbf{S},\mathbf{R}} \|\tilde{\mathbf{L}} - \mathbf{S} - \mathbf{R}\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mu\|\mathbf{R}\|_* \quad (2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm; $\|\cdot\|_1$ the ℓ_1 norm; $\|\cdot\|_*$ the nuclear norm; and, $\lambda, \mu > 0$ are properly chosen regularization parameters. The ℓ_1 norm controls sparsity of $\mathbf{S}$, while the nuclear norm equals the ℓ_1 norm applied to the vector of singular values of $\mathbf{R}$, which promotes $\mathbf{R}$ to be low rank [23]. The nonzero entries of $\mathbf{S}$ reveal the anomalous edges that do not adhere to the low rank model.

Albeit interesting, the approach in (2) does not leverage the feature matrix $\mathbf{X}$. In the social voting network paradigm, two individuals with opposing political views may be friends

because they live in the same state or belong to the same age group (cf. Fig. 1). The next section outlines our modeling approach that incorporates nodal features when available.

B. Smoothness over features

Smoothness of signal models has been widely adopted by several learning methods in the form of regularization for e.g., (semi-)supervised learning over graphs [24]. Intuitively, $\mathbf{X}$ is deemed smooth if features of connected vertices have similar values. This principle suggests, for instance, estimating one person's age by looking at their friends' age [25], [26].

Smooth features among neighboring nodes are captured by the so-termed Laplacian regularizer [25], [26], [27]

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \mathrm{Tr}(\mathbf{X}^T\mathbf{L}\mathbf{X}) \quad (3)$$

where $a_{ij} \neq 0$ only if nodes i and j are connected, Tr denotes matrix trace, and ^{T} stands for transposition. The value of (3) is small if neighboring vertices have similar signal values and high otherwise. Thus, minimizing (3) penalizes differences between features of connected nodes.

III. LEVERAGING SPARSITY, SMOOTHNESS AND LOW RANK

Incorporating feature smoothness, (2) boils down to (cf. (3))

$$\min_{\mathbf{S},\mathbf{R}} \|\tilde{\mathbf{L}} - \mathbf{S} - \mathbf{R}\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mu\|\mathbf{R}\|_* + \gamma\mathrm{Tr}(\mathbf{X}^T\mathbf{R}\mathbf{X}) \quad (4)$$

where $\gamma > 0$ tunes the degree of smoothness. The non-zero entries of $\mathbf{S}$ reveal edges that do not adhere to the low rank model with smooth features. The next subsections provide alternative formulations of (4), and corresponding optimization solvers with complementary benefits.

A. Anomalous edge identification via low rank factorization

As both the nuclear and ℓ_1 norms are non-differentiable at the origin, the optimization problem (4) is convex but non-smooth. Although iterative solvers of (4) are available, they rely on performing a costly singular value decomposition per iteration, which incurs prohibitive computational complexity, especially for large-scale graphs [2], [28].

A prudent strategy to bypass this hurdle relies on the factorization $\mathbf{R} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}, \mathbf{V}$ are $N \times R$ matrices with R denoting an upper bound on the rank of $\mathbf{R}$. Specifically, it holds that [29], [30]

$$\|\mathbf{R}\|_* = \min_{\mathbf{U},\mathbf{V}} \|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 \quad \text{s.t. } \mathbf{R} = \mathbf{U}\mathbf{V}^T. \quad (5)$$

Combining (5) with (4), the optimization in (4) reduces to

$$\min_{\mathbf{S},\mathbf{U},\mathbf{V}} f(\mathbf{U},\mathbf{V},\mathbf{S}) :=$$
$$\|\tilde{\mathbf{L}} - \mathbf{S} - \mathbf{U}\mathbf{V}^T\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mu(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2)$$
$$+ \gamma\mathrm{Tr}(\mathbf{X}^T(\mathbf{U}\mathbf{V}^T)\mathbf{X}). \quad (6)$$

Although (6) is non-convex, alternating minimization solvers can be employed with closed-form updates. Different from

727

(4), problem (6) can be solved in a distributed fashion. We developed an alternating least-squares (ALS) solver of (6) whose steps are listed under Algorithm 1. Our ALS is provably convergent, but the proof is omitted due to space limitations.

B. Anomalous edge identification aided by Laplacian recovery

The formulation in (4), focused solely on retrieving the anomalous edges. However, one can also approximate the *unperturbed* graph structure, and thus aid the identification of anomalous edges. This subsection deals with robust estimation of the nominal graph topology too, and further improving identification of the anomalous edges.

Graph connections capture correlations among nodal feature vectors; see [31] for a comprehensive survey. Learning the graph structure by leveraging the statistical properties of the features will endow estimation with increased robustness to anomalies. Popular approaches to estimating graph structure are closely related to the sparse inverse covariance matrix estimation [32], [33], [34]. The graph Laplacian can be approximated similarly by estimating the sparse inverse covariance matrix [35], [32], [36]. Specifically, the sparse inverse covariance selection problem [35] is formulated as follows

$$\min_{\mathbf{\Theta} \succeq \mathbf{0}} \; \mathrm{Tr}(\mathbf{X}\mathbf{X}^T\mathbf{\Theta}) - \log\det(\mathbf{\Theta}) + \beta\|\mathbf{\Theta}\|_1 \qquad (7)$$

where $\mathrm{Tr}(\mathbf{X}\mathbf{X}^T\mathbf{\Theta}) - \log\det(\mathbf{\Theta})$ minimizes the negative log-likelihood, while $\|\mathbf{\Theta}\|_1$ promotes sparsity. The term $\mathrm{Tr}(\mathbf{X}\mathbf{X}^T\mathbf{\Theta})$ is closely related to the smoothing term $\mathrm{Tr}(\mathbf{X}^T\mathbf{R}\mathbf{X})$ of (4), and holds the feature correlation information. Although $\mathbf{\Theta}$ captures feature correlations ($\mathbf{\Theta}_{ij} \neq 0$ if nodes i and j have conditionally correlated features), it is not necessarily a valid Laplacian.

To cope with this challenge, we couple (2) and (7), and introduce due constraints to recover the unperturbed graph Laplacian and the anomalies

$$\min_{\mathbf{R},\mathbf{S},\mathbf{\Theta} \succeq \mathbf{0}} \; \|\tilde{\mathbf{L}} - \mathbf{S} - \mathbf{R}\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mu\|\mathbf{R}\|_*$$
$$+ \alpha(\mathrm{Tr}(\mathbf{X}\mathbf{X}^T\mathbf{\Theta}) - \log\det(\mathbf{\Theta}) + \beta\|\mathbf{\Theta}\|_1)$$
$$+ \kappa\|\mathbf{R} - \mathbf{\Theta}\|_F^2 \qquad (8a)$$

subject to

$$\mathrm{Tr}(\mathbf{R}) \geq m \qquad (8b)$$
$$\mathbf{R}\mathbf{1} = \mathbf{0} \qquad (8c)$$
$$\mathbf{R}_{ij} \leq 0, \; i \neq j \qquad (8d)$$

where (8b), (8c), (8d) constrain the learned $\mathbf{R}$ to be positive semidefinite and thus a valid Laplacian, while $\|\mathbf{R} - \mathbf{\Theta}\|_F^2$ seeks a Laplacian that stays close to $\mathbf{\Theta}$ but also couples (2) with (7) (closeness is tuned by the scalar $\kappa > 0$). The number of edges in the sought Laplacian is benchmarked by m (cf. (8b)), while $\alpha > 0$ weighs the degree of smoothness in (2) and (7).

The problem in (8) is convex and can be solved using either well known optimization solvers (e.g interior point schemes), or, the alternating direction method of multipliers (ADMM)

method that can also afford for distributed implementation; we will henceforth refer to this ADMM solver of (8)) as Algorithm 2. Note also that upon recovering the Laplacian, the adjacency matrix can be readily obtained.

Algorithm 1 ALS for (6)

1: **Initialize**
   $l = 0, \mathbf{U}^{(l)}, \mathbf{V}^{(l)}, \mathbf{S}^{(l)}$
2: **while** not converged **do**
3: $\mathbf{U}^{(l+1)} = \underset{\mathbf{U}}{\arg\min} \, f(\mathbf{U}, \mathbf{V}^{(l)}, \mathbf{S}^{(l)})$
4: $\mathbf{V}^{(l+1)} = \underset{\mathbf{V}}{\arg\min} \, f(\mathbf{U}^{(l+1)}, \mathbf{V}, \mathbf{S}^{(l)})$
5: $\mathbf{S}^{(l+1)} = \underset{\mathbf{S}}{\arg\min} \, f(\mathbf{U}^{(l+1)}, \mathbf{V}^{(l+1)}, \mathbf{S})$
6: $l = l + 1$
7: **end while**

Computational complexity. The $\mathbf{U}$ and $\mathbf{V}$ updates under Algorithm 1 are available in closed form, while $\mathbf{S}$ is updated entry wise using the soft threshold operator [37]. Considering $R \ll F \ll N$, the most computationally expensive operation is matrix multiplication between $N \times N$ and $N \times R$ matrices. Therefore, the overall complexity per step is $\mathcal{O}(N^2R)$ [38].

Algorithm 2 minimizes a convex function under (8b), (8c), and (8d) that can be expressed as linear constraints [39]. This implies that (8) can be efficiently solved in $\mathcal{O}(\log(\frac{1}{\epsilon}))$ interior point sub-iterations; see e.g., [40], [41], [42], [43], [44]. Accounting also for the SVD computations $\mathcal{O}(N^3)$, and matrix multiplications, the overall complexity is $\mathcal{O}(\log(\frac{1}{\epsilon})N^3)$.

IV. Numerical Tests

This section evaluates the anomaly identification performance of the novel methods with real and synthetic graph datasets.

Stochastic block model (SBM). SBMs are synthetic networks that consist of C communities. The feature matrix is formed by the F eigenvectors of the Laplacian matrix that correspond to the F smallest eigenvalues. Such features are smooth with respect to the graph [25]. The generated stochastic block model networks are denoted as SBM(C,F).

Aarhus network. The Aarhus dataset is a social network that consists of 32 employees (nodes) from the Computer Science Department of Aarhus University and represents their lunch relations (edges) [45]. The 32×26 feature matrix $\mathbf{X}$ is a binary matrix that represents the lunch relationships between the 32 employees from Computer Science department and another group of 26 people.

A. Evaluation results

In the SBM, anomalous edges are randomly added between nodes that belong to different communities, while in the Aarhus dataset, anomalous edges are inserted randomly between any two nodes. After the optimization, edges with negative (nonzero) values in $\mathbf{S}$ are collected in the set $\hat{\mathcal{E}}$. Edges in $\hat{\mathcal{E}}$ are sorted in ascending order so that possible anomalous edges are ordered by probability of correct detection. The tests are repeated 10 times and the average percentage values of the following metrics are reported.

728

TABLE I: Best averaged metrics achieved on each graph configuration for each model. Bold font indicates the highest value of each metric in a given network.

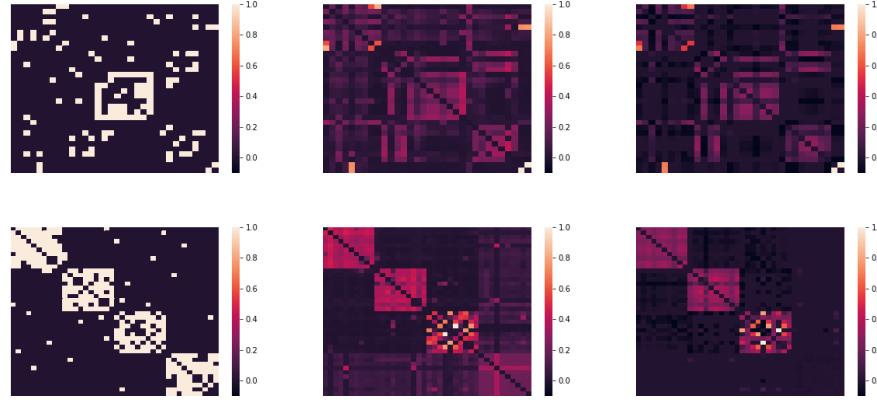| Dataset	Random guess		Baseline		Algorithm 1		Algorithm 2	
	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR
SBM(4,4)	6.6	0.7	24.0	34.8	**57.0**	58.6	55.9	**72.6**
SBM(6,3)	4.0	0.4	14.0	22.1	**46.7**	37.6	42.6	**64.6**
SBM(8,4)	0.3	∼ 0.0	11.0	20.5	24.0	24.2	**42.0**	**55.5**
Aarhus	13.7	1.4	38.0	25.6	**72.0**	46.9	66.0	**88.4**

Fig. 2: Aarhus (top) and SBM$(4, 4)$ (bottom) networks heatmap of the original (left) adjacency. The estimated one (middle) from Method 2 is closer to the original than the baseline estimation (right).

Hit rate@10 (H@10), which indicates how many edges are correctly identified as anomalous, given the first 10 candidate anomalous edges of $\hat{\mathcal{E}}$.

Mean reciprocal rank (MRR). Each edge of $\hat{\mathcal{E}}$ receives a reciprocal rank score (if correctly identified as anomalous) based on its position in $\hat{\mathcal{E}}$: 1 for first place, 1/2 for second place, 1/3 for third place, and so on. MRR is the average of reciprocal ranks of all correctly identified anomalous edges.

Table I illustrates the *best* evaluation results for different methods after systematic parameter search. The following baseline methods are considered. The method in (2) that is employed by various approaches, e.g [7], and does not consider feature smoothness and a random-guess method that selects a random subset of the edges as anomalous.

Both proposed methods significantly outperform the baseline, which speaks for the importance of smoothness constraints for anomaly identification. The improvements on H@10 accuracy metric compared to the baseline ranges from 1.9 times higher (Aarhus, Algorithm 1) to 3.8 times higher (SBM(8,4), Algorithm 2). Improvements on MRR ranges from 1.2 times (SBM(8,4), Algorithm 1) to 3.5 times higher (Aarhus, Algorithm 2). The highest values of H@10 and MRR are observed in the Aarhus dataset ($\sim$ 72% and 88.4% respectively). In all cases, our second algorithm outperforms the first one in the MRR metric due to its ability to approximate the exact topology of the original graph, which enhances the anomaly identification performance. On the other hand, although Algorithm 2 achieves better anomaly identification performance, it is less computationally efficient compared to Algorithm 1.

Finally, Figure 2 illustrates the estimated graph topology by the second method, compared to the ground truth topology on two datasets. As a baseline method, method (7) is used, which is employed by various approaches, e.g. [46]. These qualitative results confirm that the second method captures the main topological structure of the original graph, while its topology estimation is closer to the ground truth topology than the baseline method.

V. CONCLUSIONS AND FUTURE WORK

The present work aims at identifying anomalous edges on attributed networks. Two novel approaches are formulated to jointly leverage the low rank plus sparse decomposition of the Laplacian of the perturbed graph, along with feature smoothness. Extensive numerical tests corroborate that the novel adaptation of feature smoothness markedly improves the anomaly identification performance and the proposed methods outperform the baseline. Future research directions include robust topology identification, robust community detection by employing ego-tensors [47], and anomaly identification on dynamic and knowledge graphs.

REFERENCES

[1] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: A survey," *Data Mining and Knowledge Discovery*, vol. 29, p. 626–688, May 2015.
[2] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, pp. 50–66, 2 2013.

[3] V. N. Ioannidis, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, "Coupled graph and tensor factorization for recommender systems and community detection," *IEEE Transactions on Knowledge and Data Engineering*, 1 2019.

[4] E. Ceci, Y. Shen, G. B. Giannakis, and S. Barbarossa, "Graph-based learning under perturbations via total least-squares," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2870–2882, 2020.

[5] E. Ceci and S. Barbarossa, "Graph signal processing in the presence of topology uncertainties," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1558–1573, 2020.

[6] E. Ceci, Y. Shen, G. B. Giannakis, and S. Barbarossa, "Signal and graph perturbations via total least-squares," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 747–751, IEEE, 2018.

[7] B. Baingana and G. B. Giannakis, "Joint community and anomaly tracking in dynamic networks," *IEEE Transactions on Signal Processing*, vol. 64, pp. 2013–2025, 4 2016.

[8] D. Koutra, E. E. Papalexakis, and C. Faloutsos, "Tensorsplat: Spotting latent anomalies in time," in *2012 16th Panhellenic Conference on Informatics*, pp. 144–149, Oct 2012.

[9] D. Chakrabarti, "Autopart: Parameter-free graph partitioning and outlier detection," in *Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases*, PKDD '04, (Berlin, Heidelberg), p. 112–124, Springer-Verlag, 2004.

[10] H. Tong and C. Lin, "Non-negative residual matrix factorization with application to graph anomaly detection," in *Proceedings of the 11th SIAM International Conference on Data Mining, SDM 2011*, pp. 143–153, 2011.

[11] V. Nikulin and T.-H. Huang, "Unsupervised dimensionality reduction via gradient-based matrix factorization with two adaptive learning rates," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning* (I. Guyon, G. Dror, V. Lemaire, G. Taylor, and D. Silver, eds.), vol. 27 of *Proceedings of Machine Learning Research*, (Bellevue, Washington, USA), pp. 181–194, PMLR, 02 Jul 2012.

[12] M. Ambai, N. P. Utama, and Y. Yoshida, "Dimensionality reduction for histogram features based on supervised non-negative matrix factorization," *IEICE Transactions on Information and Systems*, vol. E94.D, no. 10, pp. 1870–1879, 2011.

[13] D. Kuang, H. Park, and C. H. Q. Ding, "Symmetric nonnegative matrix factorization for graph clustering," in *12th SIAM International Conference on Data Mining (SDM '12), pp. 106–117*, 2012.

[14] X. Gao, B. Xiao, D. Tao, and X. Li, "A survey of graph edit distance," *Pattern Analysis and Applications*, vol. 13, pp. 113–129, 2008.

[15] E. Müller, P. I. Sánchez, Y. Mülle, and K. Böhm, "Ranking outlier nodes in subspaces of attributed graphs," in *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pp. 216–222, April 2013.

[16] L. Akoglu, M. McGlohon, and C. Faloutsos, "oddball: Spotting anomalies in weighted graphs," in *Advances in Knowledge Discovery and Data Mining* (M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi, eds.), (Berlin, Heidelberg), pp. 410–421, Springer Berlin Heidelberg, 2010.

[17] V. N. Ioannidis, D. Berberidis, and G. B. Giannakis, "Graphsac: Detecting anomalies in large-scale graphs," *arXiv preprint arXiv::1910.09589*, 2019.

[18] V. N. Ioannidis and G. B. Giannakis, "Defending graph convolutional networks against adversarial attacks," in *Proceedings IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 8469–8473, IEEE, 2020.

[19] V. N. Ioannidis, S. Chen, and G. B. Giannakis, "Efficient and stable graph scattering transforms via pruning," *IEEE Transactions on Pattern Anal. Mach. Intel.*, 2020.

[20] V. N. Ioannidis, A. G. Marques, and G. B. Giannakis, "Efficient and stable graph scattering transforms via pruning," *IEEE Trans. Sig. Process.*, 2020.

[21] J. Gao, F. Liang, W. Fan, C. Wang, Y. Sun, and J. Han, "On community outliers and their efficient detection in information networks," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, (New York, NY, USA), p. 813–822, Association for Computing Machinery, 2010.

[22] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), p. 1225–1234, Association for Computing Machinery, 2016.

[23] B. Recht, W. Xu, and B. Hassibi, "Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization," in *2008 47th IEEE Conference on Decision and Control*, pp. 3065–3070, Dec 2008.

[24] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, pp. 6160–6173, 2014.

[25] V. N. Ioannidis, D. Romero, and G. B. Giannakis, "Kernel-based reconstruction of space-time functions via extended graphs," in *50th Asilomar Conference on Signals, Systems and Computers*, pp. 1829–1833, Nov 2016.

[26] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines* (B. Schölkopf and M. K. Warmuth, eds.), (Berlin, Heidelberg), pp. 144–158, Springer Berlin Heidelberg, 2003.

[27] V. N. Ioannidis, M. Ma, A. Nikolakopoulos, G. B. Giannakis, and D. Romero, "Kernel-based inference of functions on graphs," in *Adaptive Learning Methods for Nonlinear System Modeling* (D. Comminiello and J. Principe, eds.), Elsevier, 2018.

[28] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Transactions on Information Theory*, vol. 59, pp. 5186–5205, Aug 2013.

[29] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, p. 471–501, Aug. 2010.

[30] B. Recht and C. Ré, "Parallel stochastic gradient algorithms for large-scale matrix completion," *Mathematical Programming Computation*, vol. 5, pp. 201–226, 2013.

[31] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proceedings of the IEEE*, vol. 106, pp. 787–807, May 2018.

[32] B. M. Lake and J. B. Tenenbaum, "Discovering structure by learning sparse graph," in *Proceedings of the 33rd Annual Cognitive Science Conference*, 2010.

[33] M. Yuan and Y. Lin, "Model selection and estimation in the gaussian graphical model," *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.

[34] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *J. Mach. Learn. Res.*, vol. 9, p. 485–516, June 2008.

[35] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, pp. 432–441, 12 2007.

[36] L. Zhao, Y. Wang, S. Kumar, and D. P. Palomar, "Optimization algorithms for graph laplacian estimation via admm and mm," *IEEE Transactions on Signal Processing*, vol. 67, pp. 4231–4244, Aug 2019.

[37] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, p. 127–239, Jan. 2014.

[38] A. Davie and A. Stothers, "Improved bound for complexity of matrix multiplication," *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, vol. 143, 04 2013.

[39] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, pp. 6160–6173, Dec 2016.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. USA: Cambridge University Press, 2004.

[41] S. Diamond and S. Boyd, "Cvxpy: A python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, pp. 2909–2913, Jan. 2016.

[42] P. Raghavendra, "Optimal algorithms and inapproximability results for every csp?," in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, (New York, NY, USA), p. 245–254, Association for Computing Machinery, 2008.

[43] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming (series B)*, vol. 95, p. 2003, 2001.

[44] E. De Klerk, *Aspects of semidefinite programming: interior point algorithms and selected applications*, vol. 65. Springer Science & Business Media, 2006.

[45] E. Bayram, D. Thanou, E. Vural, and P. Frossard, "Mask combination of multi-layer graphs for global structure inference," *arXiv preprint arXiv:1910.10114*, 2019.

[46] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, pp. 6160–6173, Dec 2016.

[47] F. Sheikholeslami and G. B. Giannakis, "Overlapping community detection via constrained parafac: A divide and conquer approach," in *IEEE International Conference on Data Mining (ICDM)*, (New Orleans, LA), pp. 127–136, Nov. 2017.