LEARNING SPARSE GRAPHS WITH A CORE-PERIPHERY STRUCTURE

Sravanthi Gurugubelli and Sundeep Prabhakar Chepuri

Indian Institute of Science, Bangalore, India

ABSTRACT

In this paper, we focus on learning sparse graphs with a core-periphery structure. We propose a generative model for data associated with core-periphery structured networks to model the dependence of node attributes on core scores of the nodes of a graph through a latent graph structure. Using the proposed model, we jointly infer a sparse graph and nodal core scores that induce dense (sparse) connections in core (respectively, peripheral) parts of the network. Numerical experiments on a variety of real-world data indicate that the proposed method learns a core-periphery structured graph from node attributes alone, while simultaneously learning core score assignments that agree well with existing works that estimate core scores using graph as input and ignoring commonly available node attributes.

Index Terms— Core-periphery networks, graphical lasso, graph learning, structured graphs, topology inference.

1. INTRODUCTION

Mesoscale properties of graphs are often used to capture the structure of complex networks. A prevalent mesoscale feature in real-world networks is the core-periphery structure [1]. Core-periphery property is ubiquitous in social networks [2], trade and transport networks [3], citation networks, and communication networks [4]. They are useful in analyzing biological networks like brain networks [5,6], genome-scale metabolic networks of organisms [7], and network of protein–protein interactions [8], to name a few. For instance, in brain networks, the core-periphery structure explains cognitive learning processes [5] and in social networks [2] (contact networks [9]), the most influential spreaders of information (respectively, disease) are observed to be in the core part of the network. Therefore, identifying the core and peripheral vertices helps in analyzing the central processes in complex networks.

A core-periphery structure in graphs refers to the presence of densely connected groups of core vertices and sparsely connected periphery vertices. Core vertices are those vertices that have cohesive connections among them. Peripheral vertices, on the other hand, are not well connected to each other but are relatively well connected to core vertices. An example graph with a core-periphery structure is shown in Fig. 1(a). The dark nodes in the figure with sparse connections are peripheral nodes, while the lighter ones with dense connections are core nodes. Fig. 1(b) shows the adjacency matrix corresponding to the graph with vertices ordered in the descending order of coreness, where the lower-right block of the matrix corresponds to periphery-periphery connections and can be observed to be very sparse compared to the upper-left block corresponding to core-core connections.

Existing algorithms estimate core scores of the nodes of a graph given the network topology [1,11–15], but ignore node attributes that
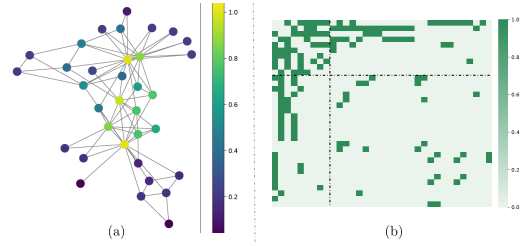
Fig. 1: (a) A network with core-periphery structure and (b) its adjacency matrix ordered according to decreasing vertex core score.
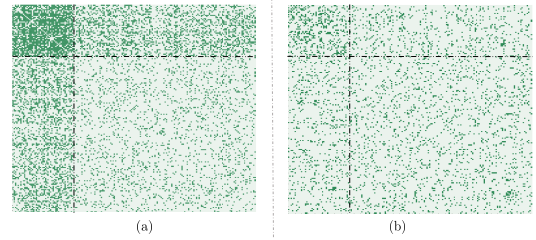

Fig. 2: Adjacency matrices estimated from (a) the proposed method and (b) graphical lasso on a webpage network dataset [10].

might also have information about the coreness of nodes. In many applications, we have access only to attributes of entities, and the underlying graph structure may not always be available. For example, in brain network analysis, we may have functional magnetic resonance imaging (fMRI) data of different subjects without information about the underlying structural connectivity. Therefore, in this work, we develop an approach that learns a core-periphery structured graph from node attributes alone so that the coreness of nodes are revealed implicitly. Conventional approaches to network topology inference [16–18] are based on graphical lasso, which assumes a Gaussian Markov random field model for data and estimates a conditional independence graph determined by the estimated sparse inverse covariance matrix [19]. However, the sparsity pattern recovered from graphical lasso does not readily incorporate a core-periphery structure in networks. To incorporate such a core-periphery structure, it is important to model the generative process of the edges in a graph through core scores of the vertices of the graph. See an illustration of networks estimated using the proposed method (described later on) and graphical lasso in Fig. 2(a) and Fig. 2(b), respectively. It can be observed that graphical lasso does not capture the core-periphery structure.

In this work, we propose a generative data model to relate the node attributes to the graph as well as to the nodal core scores. The proposed probabilistic generative model for node attributes models the dependence of the node attributes on the core scores through a latent graph structure. In particular, we model core scores as variables that influence the sparsity of the graph. Though often ignored

in graph analysis tasks, spatial distances between nodes play a vital role in differentiating the core nodes from the peripheral ones [14]. For example, countries far apart in a world trade network are less likely to be connected. To this end, we also incorporate spatial information into our model. Using the proposed position-aware probabilistic model, which promotes a core-periphery network structure, we jointly estimate a sparse graph and core scores of every node in the graph. Specifically, the proposed estimator jointly learns a sparse graph structure and node core score assignments that induce dense (sparse) connections in core (respectively, peripheral) parts of the network while accounting for the spatial distances between the nodes whenever available. We evaluate the proposed method through a number of numerical experiments on real-world data from various domains like brain, social, and transportation networks. We verify the correctness of the core scores learnt using the proposed method by comparing them with existing core score estimation algorithms, which use only the underlying known graph. The results indicate that the proposed method estimates core scores of the vertices from node attributes alone and are on par with existing methods. We also apply our method on fMRI data and report interesting observations about the differences in interactions between the brain regions in healthy subjects and individuals with attention deficit hyperactivity disorder (ADHD).

2. BACKGROUND: GAUSSIAN GRAPHICAL MODEL

Consider a weighted and undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{v_1, \cdots v_N\}$ is the vertex set with N vertices and $\mathcal{E}$ is the edge set. Let us collect the vertex attributes in the feature matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_d] \in \mathbb{R}^{N \times d}$, where the ith row of $\mathbf{X}$ contains d features of the entity associated to the ith vertex of $\mathcal{G}$.

In a Gaussian graphical model, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d$ are modeled as independent and identical observations drawn from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^N$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ is a positive definite matrix. The sparsity structure of the precision matrix $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ encodes all the conditional dependencies between the N variables associated with the vertices of $\mathcal{G}$. Specifically, any (i, j)th entry of $\boldsymbol{\Theta}$ being zero implies conditional independence of variables associated with vertices i and j, given the rest, and that there is no edge between the two vertices. Graphical lasso learns the sparsity pattern in $\boldsymbol{\Theta}$ by solving an ℓ_1-regularized Gaussian maximum log-likelihood problem [19]:

$$\underset{\boldsymbol{\Theta} \succeq 0}{\text{maximize}} \quad \log \det \boldsymbol{\Theta} - \text{tr}(\mathbf{S}\boldsymbol{\Theta}) - \lambda \|\boldsymbol{\Theta}\|_1, \quad (1)$$

where $\mathbf{S}$ is the empirical covariance matrix and $\lambda > 0$ is a regularization parameter that controls the sparsity in $\boldsymbol{\Theta}$. Although graphical lasso recovers sparse graphical models, it does not readily incorporate any specific sparsity structure such as the core-periphery structure of interest as the ℓ_1-penalty is uniformly applied on all the edges.

3. MODEL DESCRIPTION

In this section, we propose a prior that induces a sparsity pattern in graphs determined by the core scores of its vertices and the spatial distances between the vertices. Then using this probabilistic model, we propose an estimator for learning sparse Gaussian graphical models with a core-periphery structure.

Let $\mathbf{c} = [c_1, c_2, \ldots, c_N]^{\mathrm{T}} \in \mathbb{R}^N$ denote a vector containing the core scores with $c_i \in [0, 1]$ denoting the core strength of vertex i. In other words, the likelihood of vertex i belonging to the core part of the network increases with the value of c_i. Also, let d_{ij} denote the spatial distance between vertices i and j. We now propose a

probabilistic generative model that relates the node attributes in $\mathbf{X}$ to its core scores through $\boldsymbol{\Theta}$.

We model the node attributes based on a Gaussian graphical model. That is, the conditional probability distribution of $\mathbf{X}$ given the precision matrix $\boldsymbol{\Theta}$, is given by (up to constants)

$$p(\mathbf{X}|\boldsymbol{\Theta}) = \det \boldsymbol{\Theta} \exp(-\text{tr}(\mathbf{S}\boldsymbol{\Theta})). \quad (2)$$

In networks with a core-periphery structure, we have sparser connections between vertices in the periphery, relatively denser connections between the vertices in the core and periphery, and very dense connections between vertices in the core [cf. Fig. 1]. Further, vertices that are spatially well separated have sparser connections between them. To promote a sparsity pattern determined by the core-periphery structure in the graph, we therefore model the edges of the graph such that the value of Θ_{ij} is very small when $c_i + c_j - e \log(d_{ij})$ is small, i.e., if vertices i and j both belong to the periphery or if they are spatially far apart. Here, parameter e controls the dependence of d_{ij} on Θ_{ij}. The parameter e is set to 0 if the spatial information is not available or accounted for. To satisfy our above requirements, we model the generative process for each entry of Θ_{ij} using a Laplace distribution $p(\Theta_{ij}; c_i, c_j)$ with inverse diversity parameterized by the latent variables c_i and c_j as

$$w_{ij} = 1 - c_i - c_j + e \log(d_{ij}). \quad (3)$$

Specifically, the prior distribution of $\boldsymbol{\Theta}$ parameterized by $\mathbf{c}$ is

$$p(\boldsymbol{\Theta}; \mathbf{c}) = \prod_{i,j=1}^{N} p(\Theta_{ij}; c_i, c_j) = Z \prod_{i,j=1}^{N} \exp\left(-\lambda w_{ij} |\Theta_{ij}|\right),$$

where Z is the normalization constant and $\lambda > 0$ controls the overall impact of c_i and c_j on Θ_{ij}. For $p(\Theta_{ij}; \mathbf{c})$ for, $i, j, 1, \ldots, N$ to be a valid probability distribution, the inverse diversity parameter w_{ij} should be nonnegative, i.e., $1 - c_i - c_j + e \log(d_{ij}) > 0$ for i, j, $1, \ldots, N$. Next, we aim to infer the model parameters $\boldsymbol{\Theta}$ and $\mathbf{c}$ based on the observed node attributes $\mathbf{X}$.

4. THE PROPOSED LEARNING ALGORITHM

In this section, we present an algorithm to jointly learn the core score vector $\mathbf{c}$ and a sparse graph represented by the zero pattern in $\boldsymbol{\Theta}$. We estimate the model parameters $\boldsymbol{\Theta}$ and $\mathbf{c}$ by maximizing the posterior distribution given data, i.e., by maximizing

$$l(\boldsymbol{\Theta}, \mathbf{c}) = \log p(\mathbf{X}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}; \mathbf{c}) \quad (4)$$

with respect to the parameters $\boldsymbol{\Theta}$ and $\mathbf{c}$. The log-likelihood function is $\log p(\mathbf{X}|\boldsymbol{\Theta}) = \log \det \boldsymbol{\Theta} - \text{tr}(\mathbf{S}\boldsymbol{\Theta})$. The prior distribution $\log p(\boldsymbol{\Theta}; \mathbf{c}) = -\lambda \sum_{i,j=1}^{N} w_{ij} |\Theta_{ij}|$ is a *weighted* ℓ_1-penalty on $\boldsymbol{\Theta}$ with the weights determined by $\mathbf{c}$. Thus the proposed optimization problem for learning the model parameters $(\boldsymbol{\Theta}, \mathbf{c})$ is

$$\underset{\boldsymbol{\Theta} \succeq 0, \mathbf{c}}{\text{maximize}} \quad \log \det \boldsymbol{\Theta} - \text{tr}(\mathbf{S}\boldsymbol{\Theta}) - \lambda \sum_{i,j=1}^{N} w_{ij} |\Theta_{ij}|$$

$$\text{s. to} \quad w_{ij} = 1 - c_i - c_j + e \log(d_{ij})$$

$$w_{ij} > 0, \quad i, j = 1, 2, \ldots, N$$

$$\sum_{i=1}^{N} c_i = M, \quad c_i \in [0, 1], \quad i = 1, 2, \ldots, N. \quad (5)$$

The ℓ_1-penalty in the above optimization problem is relatively smaller if either of the vertices belongs to the core and is the smallest when they both belong to the core as explained by the data. As
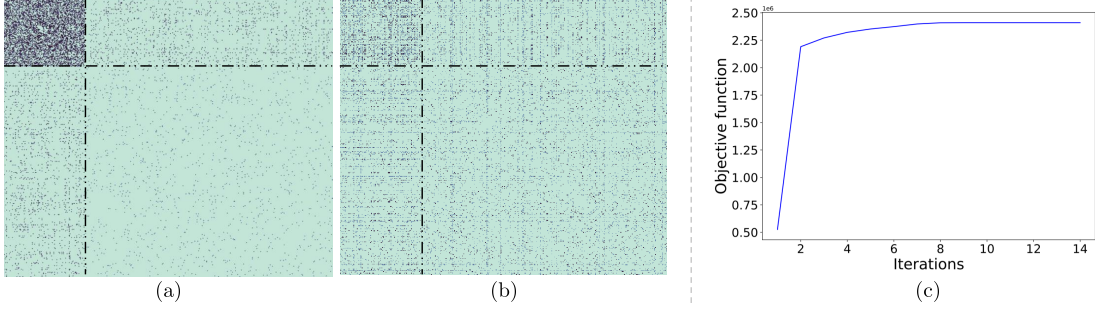
Fig. 3: (a) Estimated and (b) ground truth networks of the Twitter dataset ordered in the descending order of the estimated core scores. (c) Convergence of the algorithm for *Celegans* dataset.

the value of e is increased (while satisfying the inequality $w_{ij} > 0$), the fraction of edges between spatially distant vertices decreases. Further, to prevent the case where all the weights tend to zero, we constrain the sum of the core scores to a real-valued positive number M, while $c_i \in [0, 1]$ fixes the scale of the core scores.

The problem in (5) is a non-convex optimization problem in the variables Θ and $\mathbf{c}$. In what follows, we propose a solver based on block coordinate ascent to solve (5). This decomposes the above non-convex optimization problem into a set of convex subproblems. At each iteration of the algorithm, we update Θ by fixing $\mathbf{c}$ and then update $\mathbf{c}$ while fixing Θ.

4.1. Updating the graph

For a fixed $\mathbf{c}$, the problem in (5) simplifies to the following graphical lasso problem with a weighted ℓ_1-regularization

$$\underset{\Theta \succeq 0}{\text{maximize}} \quad \log \det \Theta - \text{tr}(\mathbf{S}\Theta) - \lambda \sum_{i,j=1}^{N} w_{ij}|\Theta_{ij}| \qquad (6)$$

with known weights that depend on $\mathbf{c}$. This is a convex program that can be solved using existing solvers, e.g., QUIC [20].

4.2. Updating the vertex core scores

For a fixed Θ, the problem in (5) simplifies to

$$\underset{c_1,\cdots,c_N}{\text{maximize}} \quad \sum_{i,j=1}^{N} |\Theta_{ij}|(c_i + c_j)$$

$$\text{s. to} \quad \sum_{i=1}^{N} c_i = M, \ c_i \in [0, 1],$$

$$c_i + c_j < 1 + e\log(d_{ij}), \ i,j = 1,\cdots,N. \qquad (7)$$

which is a linear program that can be solved using standard off-the-shelf solvers. We can clearly see from the objective function that the core scores are influenced by the edge weights, which are in turn learnt from data. The subproblem in (7) can also be used to estimate core scores given a graph.

The proposed procedure of block coordinate ascent is initialized with an arbitrary $\mathbf{c}$ (we use a scaled all-one vector in our experiments) and is repeated till convergence.

5. NUMERICAL EXPERIMENTS

In this section, we evaluate the graph inference and core score learning capabilities of our framework on real-world datasets from biological, social, and transportation networks. We compare the core

scores estimated from existing algorithms, namely, MINRES [21], Rombach [11], RandomWalk [13] and k-cores [4]. MINRES learns the core scores $\mathbf{c}$ such that the adjacency matrix is approximated by $\mathbf{cc}^{\text{T}}$. Rombach is an extension of the continuous formulation of [1], which proposes an ideal block model for core-periphery structured networks and estimates core scores by comparing the given adjacency matrix with the ideal block model. RandomWalk estimates the core scores in a network by developing the behavior of a random walker. k-cores is a method for partitioning the nodes in a network recursively from the periphery to the more central ones. The input to these existing methods is the ground truth graph. In contrast, the input to our method is just the node attributes.

5.1. Model evaluation

We first apply our method on Celegans [22], Cora [23], London underground [24], Twitter [25], and WebKB [10] datasets.

Cora is a citation network dataset. In addition to the network of citations, it also contains a binary matrix of size $N \times d$, where $N = 2708$ is the number of papers in the dataset and $d = 1433$ is the size of the vocabulary. The (i, j)th entry of $\mathbf{X}$ indicates if the jth word in the vocabulary is present in the ith paper. The Twitter dataset consists of data related to 464 Twitter users, covering athletes and organizations involved in the London 2012 summer olympics. The node attribute matrix $\mathbf{X} \in \mathbb{R}^{464 \times 3097}$ corresponds to the lists of the users, and the ground truth network is formed by the followers' information, with an edge between two users if either of them follow the other. The WebKB [10] dataset contains webpages collected from computer science departments of four universities. The node attribute matrix $\mathbf{X} \in \mathbb{R}^{4518 \times d}$ is a binary matrix indicating whether each word in the vocabulary is present or absent in the webpage. The vocabulary size is $d = 1703$ words. The spatial distance information for Cora, Twitter, and WebKB datasets is not available.

The network in Celegans dataset is a network of neurons and synapses in a type of worm, called *C. elegans*. The dataset also contains the spatial coordinates of the nodes in the network. In addition to using them as d_{ij}, we also consider the coordinates as the data matrix $\mathbf{X} \in \mathbb{R}^{N \times 2}$, where $N = 131$ is the number of nodes in the network. London underground dataset is a network of an underground transportation system. The spatial coordinates of tube stations form the data matrix $\mathbf{X} \in \mathbb{R}^{N \times 2}$, where $N = 315$ is the number of tube stations. For Cora and Twitter datasets, the hyperparameter e is set to 0 and to 0.09 for Celegans and London underground datasets. We fix M to $N/8$ for all the datasets. Increasing M increases the fraction of nodes with high core scores, whereas λ can be tuned according to the required percentage of edges in the network.

Once the core scores are learnt, we order the nodes of both the learnt and the ground truth networks in the decreasing order of core

		Proposed	MINRES	Rombach	RandomWalk	k-cores
Celegans	$\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	41.940	41.821	39.076	40.877	39.051
	$\|\|\mathbf{\Theta}\| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	32.642	32.841	32.538	32.707	32.748
Cora	$\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	55.488	55.434	54.690	55.326	54.909
	$\|\|\mathbf{\Theta}\| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	47.626	55.722	47.884	47.983	47.625
London underground	$\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	79.216	79.249	78.7563	79.338	79.169
	$\|\|\mathbf{\Theta}\| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	78.818	78.905	78.811	78.858	78.856
Twitter	$\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	134.692	137.142	124.112	131.278	129.221
	$\|\|\mathbf{\Theta}\| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$	110.526	111.837	111.427	111.429	110.526

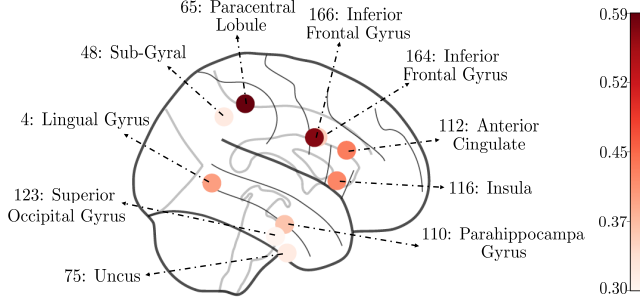**Table 1**: Comparison of the proposed method with the existing core score estimation algorithms.

Fig. 4: Regions with significant differences in the core scores of healthy individuals and subjects with ADHD.

scores. We often observe a perfect core-periphery structure in the estimated network. Although the proposed method is agnostic of the ground truth network, we observe that the ground truth network, when ordered according to the core scores learnt using the node attributes of the network alone, reveals the core-periphery structure. As an example, the adjacency matrices of the learnt and the ground truth networks of Twitter dataset are shown in Fig. 3. The adjacency matrices computed using the proposed method and graphical lasso on a subset of WebKB data related to Texas University are shown in Fig. 2(a) and Fig. 2(b), respectively.

To compare the proposed method with existing works that estimate core scores given a network, we apply existing core score estimation algorithms on the ground truth networks of the considered datasets, whereas for the proposed method we compute core scores from node attributes. We then order the networks according to the core scores given by the respective algorithms. We compare the ordered networks $\mathbf{\Theta}_0$ with the ideal core-periphery block model [1], given by

$$\mathbf{\Theta}_{\text{ideal}} = \left[\begin{array}{c|c} \mathbf{1}_{tt} & \mathbf{0}_{t(N-t)} \\ \hline \mathbf{0}_{(N-t)t} & \mathbf{0}_{(N-t)(N-t)} \end{array} \right], \qquad (8)$$

where $\mathbf{1}_{mn}$ and $\mathbf{0}_{mn}$ are $m \times n$ dimensional matrices with all ones and all zeros, respectively. We compute $\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ for different algorithms. In (8), t fixes the proportion of core nodes in the considered network. We fix t to $N/4$ for all the experiments in this section. Comparison of $\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ and $\|\|\mathbf{\Theta}\| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ for different algorithms including the proposed method are shown in Table 1. We observe that the values given by the proposed method are similar to those obtained from the other methods. This indicates that the core-periphery partitioning of the networks by the proposed method is similar to the others, in spite of not knowing the network directly. Furthermore, $\|\mathbf{\Theta}_0 - \mathbf{\Theta}_{\text{ideal}}\|_F^2 > \|\|\mathbf{\Theta}\| - \mathbf{\Theta}_{\text{ideal}}\|_F^2$ for all the datasets, indicating a more prominent core-periphery structure in the estimated graph than the ground truth, which is also evident from

Fig. 3(a). This suggests that the graph estimated using the proposed method by itself can be used to differentiate the core nodes from the peripheral ones.

Finally, Fig. 3(c) shows the convergence plot for the Celegans dataset. We observe that the value of the objective function monotonically increases till convergence. The proposed algorithm converges in less than 10 iterations for this dataset.

5.2. Brain network analysis

We next apply the proposed method to examine differences between the core and the peripheral regions of healthy individuals and subjects with ADHD. For this purpose, we use fMRI data from the OHSU brain institute [26]. The dataset $\mathcal{X} \in \mathbb{R}^{190 \times 74 \times 79}$ consists of fMRI time series for the regions of interest in the cc200 parcellation for a total of 79 individuals, 42 of which correspond to healthy subjects and the others to subjects with ADHD.

We independently compute the core scores of different individuals from their fMRI data. We denote the average of the core score vectors of healthy subjects by $\bar{\mathbf{c}}_{\text{HC}}$ and that of subjects with ADHD by $\bar{\mathbf{c}}_{\text{ADHD}}$. The magnitude of difference between the normalized average core score vectors of the two groups, $|\bar{\mathbf{c}}_{\text{HC}} - \bar{\mathbf{c}}_{\text{ADHD}}|$, serves as a measure of the differences in coreness of different brain regions across the two groups. The highest difference in connectivity (denoting difference in interactions) in the estimated networks is observed in the following regions: *paracentral lobule, inferior frontal gyrus, anterior singulate* and *insula*. Fig. 4 shows 10 regions with the largest difference in connectivity, as measured by the 10 largest values of $|\bar{\mathbf{c}}_{\text{HC}} - \bar{\mathbf{c}}_{\text{ADHD}}|$. The darker nodes in the figure denote the regions with a larger difference in the cores scores of the two groups. These identified regions, namely, *paracentral lobule, inferior frontal gyrus, anterior cingulate* and *insula* coincide with those reported in [27] as the regions with differences in activation for healthy individuals and patients with ADHD.

6. CONCLUSIONS

We developed a generative model to relate node attributes to the core scores of vertices through a latent graph structure. Based on the proposed generative model, we presented a joint estimator to simultaneously infer the vertex core scores and a sparse graph whose sparsity pattern is determined by the core scores. The recovered graphs can be readily used to perform core-periphery detection. We presented a block coordinate ascent algorithm to solve the proposed estimation problem. We demonstrated via numerical experiments that the proposed method learns a core-periphery structured graph from only the node attributes while learning core scores on par with methods that use the ground truth network as input. We also applied our method to fMRI data to infer the regions that are the most affected in subjects with ADHD.

7. REFERENCES

[1] S. Borgatti and M. Everett, "Models of core/periphery structures." *Social networks*, vol. 21, no. 4, pp. 375–95, Oct. 2000.

[2] P. Barberá, N. Wang, R. Bonneau, J. T. Jost, J. Nagler, J. Tucker, and S. González-Bailón, "The critical periphery in the growth of social protests." *PloS one*, vol. 10, no. 11, p. e0143611, Nov. 2015.

[3] T. Verma, F. Russmann, N. Araújo, J. Nagler, and H. Herrmann, "Emergence of core–peripheries in networks." *Nature communications*, vol. 7, no. 1, pp. 1–7, Jan. 2016.

[4] J. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, "K-core decomposition of internet graphs: hierarchies, self-similarity and measurement biases." *arXiv preprint cs/0511007*, Nov 2005.

[5] D. Bassett, N. Wymbs, M. Rombach, M. Porter, P. Mucha, and S. Grafton, "Task-based core-periphery organization of human brain dynamics." *PLoS computational biology*, vol. 9, no. 9, p. e1003171, Sep. 2013.

[6] V. Harlalka, R. Bapi, P. Vinod, and D. Roy, "Atypical flexibility in dynamic functional connectivity quantifies the severity in autism spectrum disorder." *Frontiers in Human Neuroscience*, vol. 13, no. 6, Feb. 2019.

[7] M. Da Silva, H. Ma, and A. Zeng, "Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks." *Proceedings of the IEEE.*, vol. 96, no. 8, pp. 1411–20, Jul. 2008.

[8] H. Jeong, S. Mason, A. Barabási, and Z. Oltvai, "Lethality and centrality in protein networks." *Nature*, vol. 411, no. 6833, pp. 41–2, May. 2001.

[9] M. Kitsak, L. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. Stanley, and H. Makse, "Identification of influential spreaders in complex networks." *Nature physics*, vol. 6, no. 11, pp. 888–93, Nov. 2010.

[10] M. Craven, A. McCallum, D. PiPasquo, T. Mitchell, and D. Freitag, "Learning to extract symbolic knowledge from the world wide web." *Carnegie-mellon univ pittsburgh pa school of computer Science*, 1998 (accessed Aug., 2021).

[11] M. Rombach, M. Porter, J. Fowler, and P. Mucha, "Core-periphery structure in networks." *SIAM J. Applied mathematics*, vol. 74, no. 1, pp. 167–90, Oct. 2014.

[12] P. Holme, "Core-periphery organization of complex networks." *Physical Review E.*, vol. 72, no. 4, p. 046111, Oct. 2005.

[13] R. Della, F. Dercole, and P. C, "Profiling core-periphery network structure by random walkers." *Scientific reports*, vol. 3, no. 1, pp. 1–8, Mar. 2013.

[14] J. Jia and A. Benson, "Random spatial network models for core-periphery structure." in *In Proc. 12th ACM Int. Conf. on Web Search and Data Mining*, New Orleans, USA, Jan. 2012.

[15] X. Zhang, T. Martin, and N. ME, "Identification of core-periphery structure in networks." *Physical Review E.*, vol. 91, no. 3, p. 032803, Mar. 2015.

[16] D. Xiaowen, T. Dorina, R. Michael, and F. Pascal, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, May 2019.

[17] S. Gurugubelli and S. P. Chepuri, "Learning multi-layer graphs and a common representation for clustering," in *Proc. of the European Signal Process. Conf. (EUSIPCO)*, Dublin, Aug. 2021.

[18] S. K. Kadambari and S. P. Chepuri, "Product graph learning from multi-domain data with sparsity and rank constraints," *IEEE Trans. Signal Process.*, Oct. 2021.

[19] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso." *Biostatistics*, vol. 9, no. 3, pp. 432–41, Jul. 2008.

[20] C. Hsieh, M. Sustik, I. Dhillon, and P. Ravikumar, "Quic: quadratic approximation for sparse inverse covariance estimation." *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 2911–47, Oct. 2014.

[21] J. Boyd, W. Fitzgerald, M. Mahutga, and D. Smith, "Computing continuous core/periphery structures for social relations data with minres/svd." *Social Networks*, vol. 32, no. 2, pp. 125–37, May 2010.

[22] K. Marcus and C. H. Claus, "Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems." *PLoS Comput Biol*, 2006 (accessed Aug., 2021). [Online]. Available: https://snap.stanford.edu/data/C-elegans-frontal.html

[23] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, "Collective classification in network data." *AI magazine*, 2008 (accessed Aug., 2021). [Online]. Available: https://snap.stanford.edu/data/C-elegans-frontal.html

[24] J. Jia and A. R. Benson, "Random spatial network models for core-periphery structure," in *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, 2019.

[25] D. Greene and P. Cunningham, "Producing a unified graph representation from multiple social network views." in *InProceedings of the 5th annual ACM web science conference*, Paris, Frace, May. 2013.

[26] Richards, E. John, and X. Wanze, "Brains for all the ages: structural neurodevelopment in infants and children from a life-span perspective." *Advances in child development and behavior*, pp. 1–52, 2015 (accessed Aug., 2021). [Online]. Available: https://www.nitrc.org/

[27] S. Dickstein, K. Bannon, C. F. Xavier, and M. Milham, "The neural correlates of attention deficit hyperactivity disorder: An ale meta-analysis." *J. Child Psychology and Psychiatry*, vol. 47, no. 10, pp. 1051–62, Nov. 2006.