COMMUNITY DETECTION IN ATTRIBUTED NETWORKS USING GRAPH WAVELETS

Meiby Ortiz-Bouza and Selin Aviyente

Department of Electrical and Computer Engineering, Michigan State University
East Lansing, 48824
ortizbou@msu.edu, aviyente@egr.msu.edu

2022 56th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5906-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEECONF56349.2022.10051935

ABSTRACT

Many real-world systems can be represented as networks where the different entities in the system are presented by nodes and their interactions by edges. An important task in the study of networks is community detection, where nodes in the same community are more densely connected to each other than they are to the rest of the network. While there has been a lot of work on community detection using the connectivity between the nodes, i.e., adjacency matrix, many real-world networks also have attribute information for each node. Community detection in attributed graphs requires joint modeling of graph structures and node attributes to make full use of available data. In this paper, we introduce a graph signal processing based approach to community detection in attributed networks. The proposed algorithm uses spectral graph wavelets to filter the attributes and constructs a new network from the graph filtered attributes across different scales. In this manner, both the graph connectivity information and the node attributes are taken into account in the community detection task. The proposed method is evaluated on multiple attributed social networks and is shown to perform well on networks with both binary and numerical attributes.

Index Terms— Community Detection, Graph Spectral Wavelets, Attributed networks

1. INTRODUCTION

Networks provide a powerful tool for representing real-world systems such as social networks, citation and coauthor networks, biological systems, among others [1]. An important aspect of analyzing networks is the discovery of communities. In most real-world networks, both the graph connectivity and node attributes are available.

Classical clustering methods focus on detecting communities using the attributes of the nodes [2] ignoring the relationships between the nodes and creating communities only based on their similarities. Another group of methods focus only on the topology of the network [3, 4]. However, these methods usually fall short in attributed graph clustering, as they do not exploit informative node features such as user profiles in social networks and document contents in citation networks. Therefore, in recent years, several methods have been proposed to detect communities by combining the node attributes and link information [5, 6, 7], such that the nodes in a community are more densely connected to each other than they are to the rest of the network, but also share some similar attributes.

The first class of methods for attributed graph clustering focus on combining link and node information by defining a new objective function that integrates the two types of similarity. In particular, the adjacency matrix that captures link information and the similarity matrix that quantifies the affinity between the attributes are combined in formulating the objective function. For example, authors in [8] proposed a method based on non-negative matrix factorization (NMF) combining graph connectivity and attributes. Authors in [9] propose a modified label propagation algorithm that uses both link strength and node attribute information to improve the quality of the detected communities. In [7], authors propose two methods to combine link and node information. The first method uses a weighting strategy to use a graph constructed from the attributes, G, to create a new graph that combines attributes and edge information. The second method linearly combines two similarity matrices, one created from the attributes and the other from the edges. In [5], authors propose a method where they obtain a k-NN graph by using a set of node attributes. The k-NN graph is then combined with the original network to strengthen the community structure of the network.

A second group of methods uses graph neural networks for community detection. In [10], authors propose a two phase method based on graph convolution network (GCN) to detect communities based on nodes and link information. A label sampling model is proposed to generate labels to validate and train the GCN. They use structural centers as initial labels and then update these labels using the nodes attributes information. Authors in [6] propose an adaptive graph convolution method for attributed networks using k-order low-pass filters. In general, GCN based methods rely on prior knowledge and a large number of labels. Since labels are often not available, these methods have limited applicability in practice.

In this paper, we propose a framework for detecting communities using spectral graph wavelets [11]. Following the approach in [11], we define the graph wavelets and scaling functions across multiple scales. By construction, a wavelet associated to a node is centered at this node and captures the structure of local connectivity in the neighborhood of this node. These graph wavelets were initially used for community detection in [3], where authors propose a method for multi-scale community detection (MS-CD). However, their method is based only on structural information as the wavelet basis is used to construct a graph for the subsequent community discovery, without taking the attributes of the network into consideration. In this work, we propose to filter the attributes using the wavelet basis at different scales to construct new graphs that will contain both link and node information. This approach is a more generalized way of implementing a graph filter since multiple filters corresponding to different scales are applied, allowing us to select the one that gives the best community structure. As the scale increases the spanned neighborhood increases and we might detect communities with larger number of nodes. In other words, wavelets on graphs provide a view of how a node sees the network.

The rest of the paper is organized as follows. Section 2 provides background on graph convolution, Graph Fourier transform and graph wavelets. Section 3 presents the proposed community detection algorithm while Section 4 illustrates results on real-world attributed networks. Finally, Section 5 provides conclusions and discussion on future work.

2. BACKGROUND

In this section, we review the basics of graph Fourier transform (GFT), graph convolution, and spectral graph wavelets.

2.1. Graph Fourier Transform

Given a graph, $\mathcal{G} = \{V, E, \mathbf{A}\}$, where V, E and $\mathbf{A}$ are the set of nodes, edges, and adjacency matrix of the graph, respectively, the graph Laplacian is $\boldsymbol{\mathcal{L}} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal degree matrix defined as $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. The normalized Laplacian matrix $\boldsymbol{\mathcal{L}}_n$ is defined as $\boldsymbol{\mathcal{L}}_n = \mathbf{D}^{-1/2}(\mathbf{D}\text{-}\mathbf{A})\mathbf{D}^{-1/2} = \mathbf{I_N} - \mathbf{D^{-1/2}AD^{-1/2}}$, where $\mathbf{I_N}$ is the identity matrix of size N. The spectrum of $\boldsymbol{\mathcal{L}}$ is composed of its set of eigenvalues, $\{\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N\}$, and by the matrix $\mathbf{U}$ of its normalized eigenvectors $\mathbf{U} = (u_1|u_2|...|u_N)$ [12]. In graph signal processing, $\mathbf{U}$ is considered as the matrix of the graph's Fourier modes, and $(\sqrt{\lambda_i})_{i=1,..,N}$ as the set of associated frequencies [13]. The graph Fourier transform of a signal f defined on the nodes of the graph is then defined as $\tilde{f} = \mathbf{U}^\top f$.

2.2. Graph Convolution

A graph signal matrix can be represented as a vector $f \in \mathbb{R}^{N\times p}$, where N is the number of nodes in the graph and p is the number of attributes for each node. The symmetrically normalized graph Laplacian can be eigen-decomposed as $\boldsymbol{\mathcal{L}}_n = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1}$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, ..., \lambda_n)$ are the eigenvalues in increasing order, and $\mathbf{U} = (u_1|u_2|...|u_N)$ are the associated orthogonal eigenvectors. A linear graph filter can be represented as a matrix $\mathbf{G} = \mathbf{U}p(\boldsymbol{\Lambda})\mathbf{U}^{-1}$, where $p(\boldsymbol{\Lambda}) = \text{diag}(p(\lambda_1), ..., p(\lambda_n))$ is called the frequency response function of $\mathbf{G}$. Graph convolution is defined then as the product of a graph signal f with a graph filter $\mathbf{G}$, $\tilde{f} = \mathbf{G}f$, where $\tilde{f}$ is the filtered graph signal.

2.3. Spectral Graph Wavelets

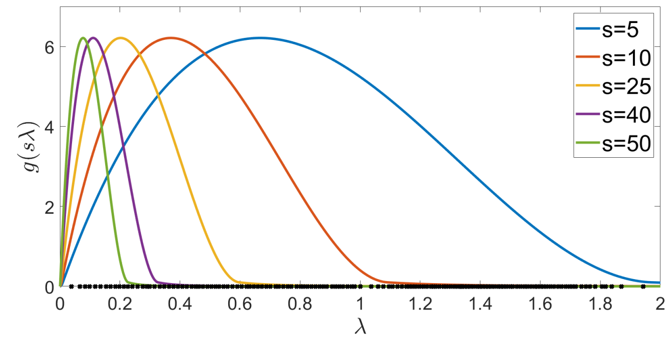
Spectral graph wavelets were defined in [11] using the graph Fourier modes, $\mathbf{U}$. The wavelet at scale s centered at node a is denoted as $\psi_{s,a}$. The construction of $\psi_{s,a}$ is based on band-pass filters defined in the graph Fourier domain, generated by stretching a band-pass wavelet filter kernel $g(\cdot)$ with a scale parameter $s > 0$. The matrix representation of the kernel filter at a scale s is given by a diagonal filter applied to the N eigenvalues of $\mathcal{L}$, $\mathbf{G}_s = \text{diag}(g(s\lambda_1), g(s\lambda_2), ..., g(s\lambda_N))$. The wavelet basis at a scale s is then given by

$$\boldsymbol{\Psi}_s = \{\psi_{s,1}|\psi_{s,2}|...|\psi_{s,N}\} = \mathbf{U}\mathbf{G}_s\mathbf{U}^\top.$$

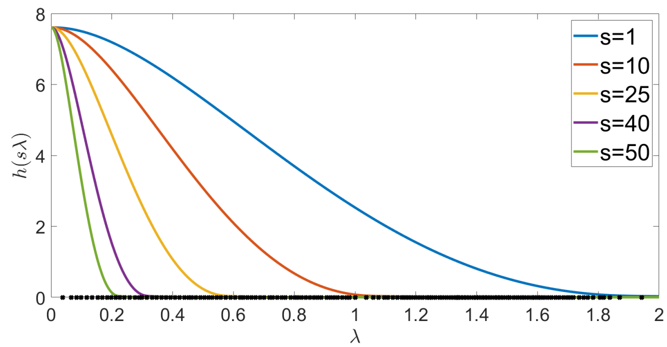
The wavelet coefficients at scale s for a graph signal f are defined as $\tilde{f}_\Psi = \boldsymbol{\Psi}_s^\top f$. This wavelet transform will be used later in Section 3.

By this definition, a wavelet associated to a node a is centered around this node, i.e., wavelets on graphs provide a view of the network from the node perspective. At small scales, the filter is stretched out and lets through high frequency modes. Therefore, the corresponding wavelet extends only to the close neighborhood of the node in the graph. At large scales, the filter function is compressed around low frequency modes and the corresponding wavelet spans a larger neighborhood. The parameters and the shape of the kernel filter $g(\cdot)$ used in this work are defined in the next section.

Similarly, the scaling basis can be generated using a low-pass scaling filter kernel $h(\cdot)$ as



(a)



(b)

Fig. 1: (a) Band-pass wavelet filter and (b) Low-pass scaling filter functions for different scales. The eigenvalues of the example network are indicated with crosses on the x-axis.

$$\boldsymbol{\Phi}_s = \mathbf{U}\text{diag}(h(s\lambda_1), h(s\lambda_2), \ldots, h(s\lambda_N))\mathbf{U}^\top = \mathbf{U}\mathbf{H}_s\mathbf{U}^\top$$

and thus, we can obtain the scaling coefficients, $\tilde{f}_\Phi = \boldsymbol{\Phi}_s^\top f$. The scaling function kernel $h(\cdot)$ is a low-pass filter designed to smoothly represent the low frequency content on the graph.

2.4. Graph Wavelet Filter

We use the band-pass filter kernel proposed in [11, 3] defined as follows,

$$g(x; \alpha, \beta, x_1, x_2) = \begin{cases} x_1^{-\alpha} x^\alpha & \text{for } x < x_1, \\ p(x) & \text{for } x_1 \leq x \leq x_2, \\ x_2^\beta x^{-\beta} & \text{for } x > x_2, \end{cases} \quad (1)$$

where $p(x)$ is a cubic spline such that g and g' are continuous. The filter g is parametrized by the integers α and β, and x_1 and x_2 are the transition points. These parameters were set in [3] following the argument that the eigenvector associated with the smallest non-zero eigenvalue (Fiedler vector) is important for community detection as it contains information on the coarsest description of the graph [14]. Following the analysis of authors in [3], these parameters are set to $\alpha = 2$, $\beta = 1/log_{10}(\frac{\lambda_2}{\lambda_3})$, $x_1 = 1$, $x_2 = x_1/\lambda_2$, and the minimum and maximum scales to $s_{min} = x_1/\lambda_1$ and $s_{max} = x_2/\lambda_2$, respectively. The wavelet scales s are selected to be logarithmically spaced between the minimum and maximum scales s_{min} and s_{max}, since, in

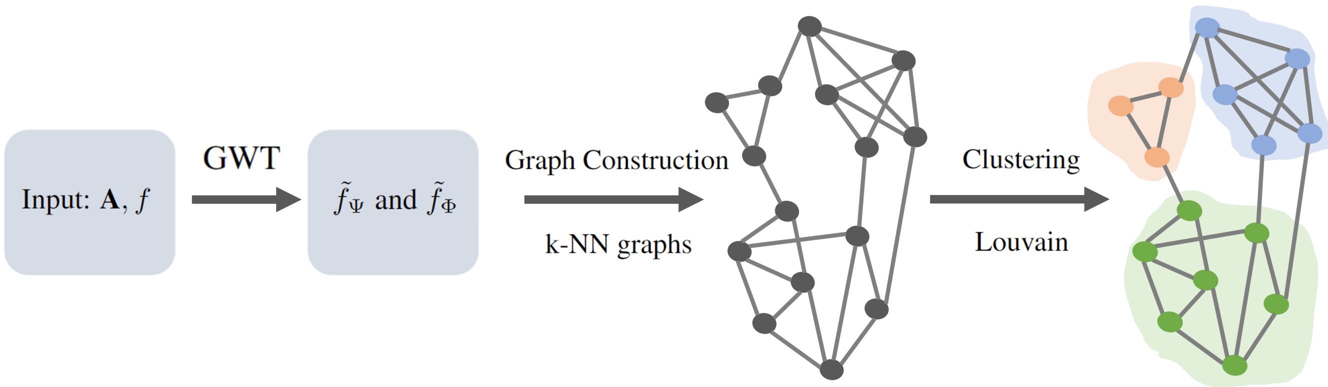


Fig. 2: Framework of the Proposed Attributed Network Community Detection Method

general, the density of the eigenvalues on the interval $[0, 2]$ is not uniform.

Note that although the filter kernel function is defined as a continuous function in the graph Fourier domain, only the values on the spectrum are needed. Thus, for each given scale parameter the filter is discrete. Fig. 1a shows the proposed wavelet filters at different scales and parameters for a toy example of a network with 300 nodes.

2.5. Graph Scaling Filter

The scaling functions in the definition of the graph scaling function are used to represent the low frequency content on the graph. They do not generate the wavelets by scale relation as in traditional orthogonal wavelets. Thus, the design of the scaling function is not coupled to the choice of wavelet kernel $g(\cdot)$. In this paper, we use the following low-pass filter kernel

$$h(x; \alpha, \beta, x_1, x_2) = \begin{cases} (1 - x_1^{-\alpha} x^{\alpha}) + q(x_1) & \text{for } x < x_1, \\ q(x) & \text{for } x_1 \leq x \leq x_2, \\ x_2^{\beta} x^{-\beta} & \text{for } x > x_2, \end{cases} \tag{2}$$

where $q(x)$ is sixth order spline such that h and h' are continuous. The parameters are set as in the graph wavelet filter previously discussed. Fig. 1b shows the proposed scaling filters at different scales and parameters for a toy example of a network with 300 nodes.

3. PROPOSED METHOD

In this work, we propose a community detection method for attributed networks using graph wavelet transform (GWT) that takes into account both the topology and the attributes of the network. We propose to create a network using the filtered attributes $\tilde{f}_\Psi = \mathbf{\Psi}_s^\top f$ ($\tilde{f}_\Phi = \mathbf{\Phi}_s^\top f$). The scaling function coefficients and spectral graph wavelet coefficients contain the approximate and detailed information of the graph signal f defined on each node of the graph, respectively. Given the nature of the wavelet and scaling basis functions, the filtered attributes capture the topology information of the network as well as the attribute information. A new graph is constructed from the filtered attributes using the k-nearest neighbors (k-NN) method. Once the graphs are constructed we can use a clustering algorithm to detect the communities. Fig. 2 illustrates the proposed framework.

3.1. k-NN Graph Construction

The k-Nearest Neighbor Graph (k-NNG) for a set of objects V is a undirected graph with an edge between two objects $v_i, v_j \in V$, if v_i is in the set of the k-nearest neighbors of v_j with respect to a given similarity measure, denoted as $N_k(v_j)$, or v_j is in the set of the k-nearest neighbors of v_i, $N_k(v_i)$.

This results in a graph for which every point is connected to its k-th nearest neighbors capturing the local information. In this paper, we constructed k-NN graphs from the filtered attributes $\tilde{f}_\Psi$ and $\tilde{f}_\Phi$ for each scale s. The similarity measure used was the Euclidean distance. And the similarity matrix is then defined as,

$$W(i,j) = \begin{cases} 1 & \text{if } \tilde{f}_{\Psi_i} \in N_k(\tilde{f}_{\Psi_j}) \text{ or } \tilde{f}_{\Psi_j} \in N_k(\tilde{f}_{\Psi_i}), \\ 0 & \text{otherwise.} \end{cases}$$

where $\tilde{f}_{\Psi_a}$ is a p length vector of the filtered attributes corresponding to node a.

3.2. Clustering

Once the graphs from the filtered attributes $\tilde{f}_\Psi$ and $\tilde{f}_\Phi$ are created, we can find the communities. In this work, we use a well-known community detection method based on modularity maximization, Louvain [15], to cluster the nodes into communities. Louvain is a method based on maximizing the modularity metric Q defined as

$$Q = \frac{1}{2m} \sum_{ij} [W_{ij} - \gamma \frac{k_i k_j}{2m}] \delta(c_i, c_j),$$

where $\mathbf{W}$ is the similarity matrix of the graph, k_i is the sum of the weights of the edges attached to vertex i, c_i is the community to which vertex i is assigned, $m = \frac{1}{2}\sum_{ij} W_{ij}$, and γ is a resolution parameter that needs to be set. In this work, we find partitions of a network using a range of values for $\gamma \in [0.1, 1.5]$ and select the value of γ that gives us the partition with the highest evaluation metric.

4. EXPERIMENTAL RESULTS ON REAL NETWORKS

In this section, we present the experimental results for our method evaluated on four real-world datasets.

Table 1: Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) results.

	Cora		Citeseer		Sinanet		Wiki	
Algorithms	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI
MS-CD	0.5072	0.3453	0.4064	0.3926	0.4064	0.1997	0.3548	0.2677
Louvain	0.5044	0.3433	0.3645	0.3457	0.2490	0.2044	0.4105	0.2369
k-means	0.2825	0.1621	0.3597	0.3279	**0.6413**	**0.5828**	0.3323	0.0578
SC	0.2856	0.1614	0.3244	0.2898	0.5395	0.3802	0.3974	0.0954
ARGA	0.4562	0.3865	0.2967	0.2781	0.4815	0.3781	0.3715	0.1129
ARVGA	0.4657	0.3895	0.3124	0.3022	0.4854	0.3993	0.3987	0.1084
Graph from $\tilde{f}_\Psi$	**0.5553**	**0.4068**	**0.4187**	**0.4015**	0.5201	0.4732	**0.4481**	**0.2978**
Graph from $\tilde{f}_\Phi$	**0.5874**	**0.5426**	**0.4391**	**0.4500**	**0.6503**	**0.6406**	**0.4946**	**0.2929**

4.1. Datasets

We evaluated our method on two groups of real networks. The first group of networks have binary attributes, while those in the second group have numerical attributes. The first group of data sets are:

- **Cora** is composed of 2,708 machine learning papers classified into seven classes: case based reasoning, genetic algorithms, neural networks, probabilistic methods, reinforcement learning, and rule learning theory. This network has 5,429 edges representing citations among the papers (nodes). Each document is described by a binary vector of 1,433 dimensions indicating the presence or absence of 1,433 words.
- **Citeseer** is a citation network of 3,312 machine learning publications classified into 6 classes: agents, artificial intelligence, database, information retrieval, machine learning, and human-computer interaction. This network has 4,732 links and each paper is described by a binary vector of 3,703 dimensions indicating the presence of 3,703 unique words.

The second group of datasets includes

- **Sinanet** is a microblogs users relationship network of 3,490 users from 10 major forums including finance and economics, literature and arts, fashion and vogue, current events and politics, sports, science and technology, entertainment, parenting and education, public welfare, and normal life. This network has 30,282 links between the users representing the followers/followees relationships. Each user is described by a 10 dimensional numerical attribute vector describing the users' interest distribution in 10 forums.
- **Wiki** contains 2,405 long text documents classified into 19 classes, and 17,981 links between them. Each document is described by a weighted vector of length 4,973 indicating the presence of words.

4.2. Results and Discussion

Normalized Mutual Information (NMI) [16] and Adjust Rand Index (ARI) [17] are used as evaluation metrics to quantify the similarity between the detected community structure and the known ground-truth. In order to evaluate the performance of our method and the importance of taking into account both the topology and the attributes of a network, we compared the proposed method with six community detection methods, Multi-Scale Community Detection (MS-CD) [3], Louvain [15], k-means, Spectral Clustering (SC), adversarially regularized graph autoencoder (ARGA) [18], and adversarially regularized variational graph autoencoder (ARVGA) [18]. The first two only consider the edge information. MS-CD uses graph wavelets while Louvain is a modularity maximization method. k-means and SC are well-known clustering methods and both of them only consider the attributes of the nodes. And, ARGA and ARVGA are graph convolutional network based methods that use both node attributes and graph structure.

Table 1 shows the results of these experiments. For our methods and for MS-CD, the results from the scale s that gives the highest evaluation metrics are reported. As we can see, taking into account both the topology and the node attributes yields better NMI and ARI results. For all datasets, the two versions of our method perform better than the rest of the algorithms except on Sinanet where k-means performs better than the proposed method using the band-pass wavelet filtered coefficients, but still worse than using the low-pass scaling filter. Since Sinanet is a network with 10 communities and 10 attributes, methods that use only the attributes like k-means may perform well as the attributes directly correspond to the different communities. On the other hand, our method using the low-pass scaling filtered coefficients $\tilde{f}_\Phi$ outperforms the rest of the methods. The low-pass filters are designed to smoothly represent the low frequency content on the graph, which is related to the similarities of nodes in the graph. Therefore, it is expected to perform better for community detection since we want nodes in a same communities to be similar.

5. CONCLUSIONS

In this paper, we proposed a community detection method for attributed networks based on graph wavelets. The proposed method detects communities taking into account both the node attributes and the topology of the network. Two classes of graph filters based on graph wavelet transform were implemented, a band-pass wavelet filter and a low-pass scaling filter. Experiments on several real-world datasets with binary and numerical attributes were conducted. The experiments show that using the attributes in addition to the topology yields more accurate communities. Moreover, results show that the low-pass scaling filter performs better than the band-pass wavelet filter, but both of them outperform the conventional clustering algorithms. Future work will consider other kernel functions and the extension of this framework to detect multi-scale community structure.

6. REFERENCES

[1] A.-L. Barabási, "Network science," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1987, p. 20120375, 2013.

[2] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010.

[3] N. Tremblay and P. Borgnat, "Graph wavelets for multiscale community mining," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5227–5239, 2014.

[4] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," vol. 328, no. 5980, pp. 876–878, 2010.

[5] C. Jia, Y. Li, M. B. Carson, X. Wang, and J. Yu, "Node attribute-enhanced community detection in complex networks," *Scientific reports*, vol. 7, no. 1, pp. 1–15, 2017.

[6] X. Zhang, H. Liu, Q. Li, and X.-M. Wu, "Attributed graph clustering via adaptive graph convolution," *arXiv preprint arXiv:1906.01210*, 2019.

[7] E. Alinezhad, B. Teimourpour, M. M. Sepehri, and M. Kargari, "Community detection in attributed networks considering both structural and attribute similarities: two mathematical programming approaches," *Neural Computing and Applications*, vol. 32, no. 8, pp. 3203–3220, 2020.

[8] D.-D. Lu, J. Qi, J. Yan, and Z.-Y. Zhang, "Community detection combining topology and attribute information," *Knowledge and Information Systems*, pp. 1–22, 2022.

[9] D. Malhotra and A. Chug, "A modified label propagation algorithm for community detection in attributed networks," *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100030, 2021.

[10] X. Wang, J. Li, L. Yang, and H. Mi, "Unsupervised learning for community detection in attributed networks based on graph convolutional network," *Neurocomputing*, vol. 456, pp. 147–155, 2021.

[11] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[12] F. Chung, "Spectral graph theory, providence, ri: Amer," *Math. Soc*, 1997.

[13] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE signal processing magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[14] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[15] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[16] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *Journal of statistical mechanics: Theory and experiment*, vol. 2005, no. 09, p. P09008, 2005.

[17] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.

[18] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, "Adversarially regularized graph autoencoder for graph embedding." in *IJCAI*, 2018, pp. 2609–2615.