How Significant Attributes are in the Community Detection of Attributed Multiplex Networks

Junwei Cheng
School of Computer Science, South China Normal University
Guangzhou, China
jung@m.scnu.edu.cn

Chaobo He
School of Computer Science, South China Normal University
Pazhou Lab
Guangzhou, China
hechaobo@m.scnu.edu.cn

Kunlin Han
Computer Science Department, University of Southern California
Los Angeles, United States
kunlinha@usc.edu

Wenjie Ma
China Mobile Group Zhejiang Company Limited
Ningbo, China
jay_m_k_ma@163.com

Yong Tang*
School of Computer Science, South China Normal University
Pazhou Lab
Guangzhou, China
ytang@m.scnu.edu.cn

ABSTRACT

Existing community detection methods for attributed multiplex networks focus on exploiting the complementary information from different topologies, while they are paying little attention to the role of attributes. However, we observe that real attributed multiplex networks exhibit two unique features, namely, consistency and homogeneity of node attributes. Therefore, in this paper, we propose a novel method, called ACDM, which is based on these two characteristics of attributes, to detect communities on attributed multiplex networks. Specifically, we extract commonality representation of nodes through the consistency of attributes. The collaboration between the homogeneity of attributes and topology information reveals the particularity representation of nodes. The comprehensive experimental results on real attributed multiplex networks well validate that our method outperforms state-of-the-art methods in most networks.

CCS CONCEPTS

• **Computing methodologies** → Cluster analysis; Neural networks; Latent variable models.

KEYWORDS

Attributed multiplex networks, Community detection, Graph neural networks, Graph autoencoder, Unsupervised learning

ACM Reference Format:

Junwei Cheng, Chaobo He, Kunlin Han, Wenjie Ma, and Yong Tang. 2023. How Significant Attributes are in the Community Detection of Attributed Multiplex Networks. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23), July 23–27, 2023, Taipei, Taiwan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3539618.3591998

*Corresponding author.

SIGIR '23, July 23–27, 2023, Taipei, Taiwan

ACM ISBN 978-1-4503-9408-6/23/07…$15.00
https://doi.org/10.1145/3539618.3591998

1 INTRODUCTION

Attributed multiplex networks (multiplex networks for short) have been widely used to represent the multiple relations among objects in the real world. They can be intuitively represented as a multilayer structure, where nodes and their attributes are shared across all layers, but each layer exhibits a distinct topology reflecting different types of relationships between the nodes. For example, in the social network shown in Figure 1, each node represents a user with personality attributes (e.g., hobbies), and the edges in each layer are endowed with different actual relations (i.e., classmates, friends and teammates from the first layer to the third layer, respectively). Apparently, the analysis of multiplex networks is a valuable and challenging work.

Community detection is one of the most popular topics in the analysis of multiplex networks. It attempts to find optimal cluster structures with dense intra-edge and sparse inter-edge (e.g., olive and lilac areas in the Figure 1). Recently, Ma et al.[1] attempted to extract relations among various layers based on a shared bias matrix with Nonnegative Matrix Factorization (NMF). Fan et al.[2] designed a graph autoencoder method based on Graph Neural Networks (GNNs) to pull out structure representations from different layers. Lin et al.[4], Park et al.[3] and Lin et al.[5] all made an effort to design methods for community detection in multiplex networks. However, these methods focus on exploiting the complementary information from different topologies to extract commonality representation of nodes, while they are paying little attention to the role of attributes in multiplex networks. A natural question, therefore, is how significant attributes are in the community detection of multiplex networks.

To answer this question, we observe real multiplex networks and find that, although each layer reflects a different topology relation, the attributes are shared across all layers, i.e., attributes exhibit consistency across different layers. Moreover, the attributes

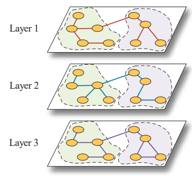

Figure 1: A three-layer toy social network with two communities

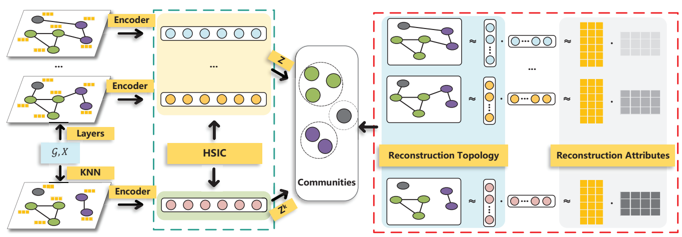

Figure 2: The framework of ACDM

exhibit homogeneity within communities. For example, in social networks, individuals with similar hobbies tend to cluster together normally[6].

Based on these observations, in this paper, we emphasize the significance of attributes in the community detection of multiplex networks. Our proposed method, namely ACDM, extracts commonality representation of nodes among various layers by using consistency of attributes without relying on the complementary information from different topologies. Additionally, the collaboration between topology information and attribute homogeneity reveals the particularity representation of each layer. To achieve this, we use two specific objectives to guide the learning process of ACDM, because they cover the two salient characteristics of the community. Also, we construct a K-Nearest Neighbor(KNN) layer, based on attributes, to extract commonality representations of nodes. Without loss of generality, we make a deduction to show that the new layer fully depicts the complementary information between nodes in different layers. Finally, we highlight the difference between commonality and particularity representations using the Hilbert-Schmidt Independence Criterion(HSIC) [7].

The contributions of this paper are:(1) By observing real multiplex networks, we discover two distinctive characteristics of attributes, i.e., consistency and homogeneity. To this end, we design a novel method, namely ACDM, to illustrate the importance of attributes in the community detection of multiplex networks. (2) Our method aims to extract the commonality representations of nodes in the multiplex network by using consistency of attributes. Notably, the advantage is to alleviate the reliance on topology information. Furthermore, the particularity representation of nodes will be revealed through the collaboration between topology information and homogeneity of attributes. (3) The extensive experiments on real-world multiplex networks show that our method achieves SOTA or on par performance with existing methods.

2 THE PROPOSED METHOD

2.1 Problem Definition

(1) Multiplex Network. A multiplex network can be defined as a non-directed graph $\mathcal{G} = \{\mathbf{A}^1, \mathbf{A}^2, ..., \mathbf{A}^{\hbar}, \mathbf{X}\}$ with $\hbar$ layers, where $\mathbf{A}^i \in \mathbb{R}^{N\times N}$ denotes the adjacency matrix of i-th layer with N nodes, and if there is a relation between u and v, $\mathbf{A}^i_{uv} = 1$, otherwise, 0. $\mathbf{X} \in \mathbb{R}^{N\times M}$ indicates the attribute matrix with M attributes.
(2) Community Detection in the Multiplex Network. Given N nodes, community detection in the multiplex network aims to divide these nodes into K disjoint communities $\{C_1, C_2, ..., C_K\}$ under unsupervised learning, so that these communities have the following features (1) dense intra-edge and sparse inter-edge. (2) similar intra-attributes.

2.2 ACDM method

(1) Overview. The main idea of ACDM is to extract the commonality representations without relying on the complementary information from different topologies, and to ensure that the particularity representation of each layer contains two salient characteristics of the community, i.e., dense edges and highly similar attributes within the communities.

Figure 2 shows the overall framework of ACDM. The multiplex network can be divided into multiple layers with different topology structure. For particularity representation $\mathbf{Z}^i$ of each layer, we use an inference method to encode layer information and use two specific objectives (i.e., reconstructing topologies and attributes) to guide the learning process of ACDM. Meanwhile, we construct a KNN layer, based on attributes, to extract commonality representation of nodes. In addition, we make a deduction to show that the new layer contains complementary information between nodes in different layers. Finally, we employ HSIC to highlight the difference between these two representations (i.e., particularity representation and commonality representation), since GNNs aggregate the same attribute information while encoding.

(2) Particularity Representation. Because different layers reflect different relations among the same set of nodes, one of the core issue is how to reveal the unique information of nodes in each layer, i.e., particularity representation $\mathbf{Z} \in \mathbb{R}^{N\times d}$.

To this end, a simple yet effective inference method, based on a two-layer GCN, can encode each layer of the multiplex network. The inference method can be described as:

$$q(\mathbf{Z}^i \mid \mathbf{X}, \mathbf{A}^i) = \prod_{n=1}^{N} \mathcal{N}(z^i_n \mid \mu^i_n, diag[(\sigma^i_n)^2]) \tag{1}$$

Here, we feed adjacency matrix $\mathbf{A}^i$ and attributes matrix $\mathbf{X}$ to GCN to fit mean vector μ^i and standard deviation vector $\log \sigma^i$, as follows:

$$\mu^i = \widetilde{\mathbf{A}}^i ReLU(\widetilde{\mathbf{A}}^i \mathbf{X}\mathbf{W}^i_0)\mathbf{W}^i_{\mu,1} \tag{2}$$

$$\log \sigma^i = \widetilde{\mathbf{A}}^i ReLU(\widetilde{\mathbf{A}}^i \mathbf{X}\mathbf{W}^i_0)\mathbf{W}^i_{\sigma,1} \tag{3}$$

where $\mathbf{W}^i_{\mu,1}$, $\mathbf{W}^i_{\sigma,1}$ and $\mathbf{W}^i_0$ are trainable weight matrices in i–th layer of the multiplex network. $\widetilde{\mathbf{A}}^i = (\mathbf{D}^i)^{-\frac{1}{2}} A^i (\mathbf{D}^i)^{-\frac{1}{2}}$ is the symmetric normalized adjacency matrix in i-layer where $\mathbf{D}^i_{nn} = \sum_j A^i_{nj}$.

Although the aforementioned information of layers has been encoded into compact particularity representations, it is difficult to cover the two characteristics of the community. Therefore, two specific objectives need to be introduced to guide the learning process of ACDM. Inspired by VGAE[8], our first objective is to optimize the variational lower bound such that the communities have the features of dense intra-edge and sparse inter-edge, as follows:

$$\mathcal{L}_v^i = \mathbb{E}_{q(\mathbf{Z}^i|\mathbf{X},\mathbf{A}^i)}[\log p(\mathbf{A}^i \mid \mathbf{Z}^i)] - KL[q(\mathbf{Z}^i \mid \mathbf{X},\mathbf{A}^i)\|p(\mathbf{Z}^i)] \quad (4)$$

where $p(\mathbf{Z}^i) = \prod_n \mathcal{N}(z_n^i \mid 0, \mathbf{I})$ is Gaussian prior. The first item is to reconstruct the topology. The second term is a Kullback-Leibler divergence that measures the similarity of prior $p(\cdot)$ distribution and posterior $q(\cdot)$ distribution.

Homogeneity of attributes plays an important role in the formation of communities, as nodes with similar attributes tend to cluster together to form communities generally. Therefore, another objective is to reconstruct attribute information so that they are highly similar within communities. The reconstructing function is:

$$\mathcal{L}_h^i = \sum_{n=1}^{N}\sum_{j=1}^{d}(z_{nj}^i - \sum_{u=1}^{M} x_{nu} w_{h,uj}^i) \quad (5)$$

where $\mathbf{w}_h^i$ is a trainable membership matrix in i-th layer and need to satisfy membership constraint as follows:

$$\mathcal{L}_r^i = \sum_{n=1}^{d}[w_{h,n}^i (w_{h,n}^i)^T - 1] + \sum_{n=1}^{d}\sum_{j\neq n}^{d} w_{h,n}^i (w_{h,j}^i)^T \quad (6)$$

(3) Commonality Representation. Observing multiplex networks, we find that all layers share the same attributes, which shows the consistency of attributes across different layers. Inspired by this discovery, a novel idea of extracting commonality representation is to construct a new layer via attributes. Putting it differently, the new layer will contain complementary information between nodes from the original layers, as shown in the following corollary: 1) **condition**: node u and v have similar attributes at layer 1. 2) **condition**: node u and s have similar attributes at layer 2. 3) **conclusion**: the complementary information is that node v and s have similar attributes. This inference process is analogous to inferring complementary information between topologies, so it is a reasonable idea.

To this end, we construct an attribute-based KNN layer $\mathcal{G} = (\mathbf{A}^k, \mathbf{X})$ to fully depict the complementary information between nodes in different layers, as follows:

$$\mathbf{A}_{ij}^k = \frac{\sum_{m=1}^{M} x_{im} \times x_{jm}}{\sqrt{\sum_{m=1}^{M}(x_{im}^2)} \times \sqrt{\sum_{m=1}^{M}(x_{jm}^2)}} \quad (7)$$

Obviously, the adjacency matrix $\mathbf{A}^k$ already contains complementary information between layers without relying on the original topology information. Furthermore, the encoding in part (2) can guarantee that the commonality representation $\mathbf{Z}^k$ still covers both characteristics of the community.

Finally, we find that encoding the same attributes easily narrows the gap between particularity representation $\mathbf{Z}$ and commonality representation $\mathbf{Z}^k$, even though these representations come from different layers. Therefore, we employ HSIC to emphasize the difference between them. Formally, the slack HSIC is defined as:

$$HSIC(\mathbf{Z},\mathbf{Z}^k) = (\mathbf{I} - \frac{1}{N}\mathbf{E})\sum_{n=1}^{N}\sum_{j=1}^{d} z_{ij}z_{ji}(\mathbf{I} - \frac{1}{N}\mathbf{E})\sum_{n=1}^{N}\sum_{j=1}^{d} z_{ij}^k z_{ji}^k \quad (8)$$

Table 1: Community detection results %. (Bold: best; "*": runner-up; "-": out-of-memory)

Dataset	IMDB				ACM				DBLP			
Metric	NMI	F1	ACC	ARI	NMI	F1	ACC	ARI	NMI	F1	ACC	ARI
GCN	0.52	35.02	42.47	0.32	43.45	53.18	62.05	41.02	4.15	23.67	23.84	3.68
VGAE	2.75	39.72	42.97	2.00	41.09	56.36	54.61	34.27	71.20	88.20	89.25	76.06
MNE	0.17	33.16	39.58	0.08	29.99	64.79	63.70	24.86	-	-	-	-
RMSC	0.54	37.75	27.02	0.18	39.73	57.46	63.15	33.12	71.11	82.48	89.94	76.47
PwMC	0.23	31.64	24.53	0.17	3.32	37.83	41.62	3.95	1.90	28.08	32.53	1.59
SwMC	0.56	37.14	26.71	0.04	8.38	47.09	38.31	1.87	37.60	56.02	65.38	38.00
HAN	9.86	41.52	55.47	8.56	58.81	88.44	88.23	59.33	**78.59**	90.78	91.14	81.24
DMGI	13.17	42.53	**58.27**	14.57	69.74	89.85	89.73	72.96	69.31	86.91	87.22	70.34
PMNE(n)	3.59	39.06	49.58	3.66	46.48	69.55	69.36	43.02	59.14	79.66	79.25	52.65
PMNE(r)	0.14	31.83	46.97	1.15	40.63	66.18	64.92	34.53	8.72	36.88	38.35	6.89
PMNE(c)	2.85	38.82	47.19	2.84	47.75	70.03	69.98	44.31	-	-	-	-
HDMI	**15.96**	**43.50**	45.03	**18.66**	68.55	72.75	72.56	72.14	71.52	40.77	45.68	77.20
O2MAC	4.21	41.59	45.02	5.64	69.23	90.53	90.42	73.94	72.87	90.13	90.74	77.80
O2MA	5.24	42.29	46.97	7.53	65.15	88.94	88.80	69.87	72.57	89.76	90.40	77.05
ACDM	5.48	42.67*	55.02	11.44	**72.69**	**91.83**	**91.80**	**77.21**	76.70*	**91.86**	**92.38**	**81.71**

where $\mathbf{E}$ is an all-one matrix, and $\mathbf{I}$ is an identity matrix. Notably, the particularity representation $\mathbf{Z}$ is not a simple mean aggregation of all values. In fact, the literature[9] confirmed that a multiplex network usually contains at least one layer that has a greater influence on the performance of methods over other layers. Because an attention mechanism can adaptively measure the importance of each particularity representation, it should be used to replace this simple way of aggregation. The attention mechanism is:

$$\mathbf{Z} = \sum_{i=1}^{\mathcal{k}} \frac{exp(\Theta^i)}{\sum_{j=1}^{\mathcal{k}} exp(\Theta^j)}\mathbf{Z}^i \quad (9)$$

where Θ^i is formally described as:

$$\Theta^i = \sum_{l=1}^{N}\sum_{u=1}^{d}[ReLU(\sum_{n=1}^{N}\sum_{j=1}^{d} z_{nj}^i w_{0,jn}^i + b_j^i)]_{lu} w_{1,u1}^i \quad (10)$$

where w_0^i, w_1^i and b^i are trainable parameters.

(4) Community Detection. Combining the aforementioned analysis, we have the following overall objective function to train ACDM:

$$\mathcal{L} = \sum_{i=1}^{\mathcal{k}+1}(\mathcal{L}_v^i + \mathcal{L}_h^i + \mathcal{L}_r^i) + \lambda HSIC(\cdot) \quad (11)$$

where λ is a coefficient that controls the degree of difference, and $\mathcal{k} + 1$ shows that the function includes the KNN layer. With the guidance of the above objective function, we can optimize the ACDM via backpropagation and detect communities via K-means.

3 EXPERIMENTS

In this section, we conduct experiments to answer the following questions: (Q1) How significant attributes are in the community detection of multiplex networks? (Q2) Why is the consistency and homogeneity of attributes important in community detection of multiplex networks? and (Q3) Is HSIC redundant?

3.1 Experiment setup

(1)Datasets. We use the same datasets as [2], i.e., IMDB, ACM and DBLP. Table 2 shows the details of datasets.

(2)Baselines. To answer Q1, we compare ACDM with many representative methods (i.e., embedding-based methods, clustering-based methods and contrastive learning-based methods). They are: GCN [10], VGAE [8], MNE [11], HAN [9], PMNE [12], DMGI [3],

Table 2: The statistics of the datasets.

Datasets	# Node	# Attributes	Relation Types	# Edges	K
IMDB	4780	1232	co-actor co-director	98,010 21,018	3
ACM	3025	1870	co-paper co-subject	29,281 2,210,761	3
DBLP	4057	334	co-author co-conference co-term	11,113 5,000,495 6,776,335	4

HDMI [13], PMSC [14], PwMC & SwMC [15] and O2MAC & O2MA [2].

(3)Parameter settings and Metrics. For ACDM, we use the Adam optimizer and set the learning rate to 0.01. The dimension of hidden layers in GCN are set to 64 and 32, respectively. And the dimension of hidden layer in the attention mechanism is 64. We set $k \in \{1, ..., 10\}$ for KNN layer to find the optimal value of k. The coefficient λ is set to $1e^{-10}$. For all benchmark datasets, our method is trained with 200 iterations, and we run K-means 3 times for each iteration. Then, we report the best result. Other baseline methods adopt the settings described in their papers. We employ four widely used metrics to measure performance of methods: Normalized Mutual Information(NMI), F1-score(F1), Accuracy(ACC) and Adjusted Rand Index(ARI).

3.2 Community Detection Performance

The community detection results are reported in Table 1. We have the following observations:

(1) Compared with all baselines, the proposed ACDM achieves the best performance on most benchmark datasets, where ACDM achieves the best performance on all metrics of ACM and most metrics of DBLP. Notably, for NMI, F1, ACC and ARI, ACDM respectively achieves maximum relative improvements of 69.37%, 54%, 61.81% and 75.34% on ACM, and certain boosts of 2.95%, 1.3%, 1.38% and 3.27% over contemporaneous methods. Analogously, for F1, ACC and ARI, ACDM respectively achieves improvements of 1.08%, 1.24% and 0.47% on DBLP. The above experimental data demonstrates the effectiveness of ACDM.

(2) On IMDB, ACDM outperforms other deep methods, except some methods based on contrastive learning. To explain this result, we make empirical research on all benchmark datasets. Specifically, we calculate the average attribute similarity $\mathcal{S}$ of communities on datasets based on Eq. (12). Figure 3 shows the calculation results for all datasets. We find that the average similarity in IMDB is much lower than the other two datasets. This means that, to some extent, the attributes within the community are not highly correlated, which explains why ACDM does not achieve the best performance on IMDB. Compared with contrastive learning-based methods, neither ACDM nor other baseline methods considers the global receptive field and thus they are less competitive in the face of the aforementioned perturbations.

$$\mathcal{S} = \frac{1}{K}\sum_{i=0}^{K}\sum_{u,v\in C_i}\mathbf{X}_{uv} \tag{12}$$

As a result, we show that attributes are indispensable and just as significant as topology, and we answer the first question posted at the beginning with the experiments.

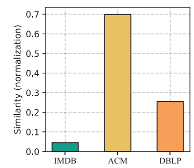

Figure 3: Empirical research.

Table 3: Ablation study results %. ($ACDM-k$: ACDM without KNN layer. $ACDM-h/r/H$: ACDM without constraints $\mathcal{L}_h$, $\mathcal{L}_r$ and $HSIC$. $ACDM-h/r$: ACDM without constraints $\mathcal{L}_h$ and $\mathcal{L}_r$.)

Dataset	IMDB				ACM				DBLP			
Metric	NMI	F1	ACC	ARI	NMI	F1	ACC	ARI	NMI	F1	ACC	ARI
$-h/r$	4.60	34.31	39.08	5.70	69.30	90.86	90.81	74.46	74.95	90.98	91.57	79.77
$-h/r/H$	4.35	36.88	40.56	5.43	67.30	90.09	90.05	72.58	74.89	90.80	91.50	79.97
$-k$	2.21	37.56	40.56	1.65	44.20	67.85	66.81	38.45	73.27	90.00	90.61	77.46
ACDM	**5.48**	**42.67**	**55.02**	**11.44**	**72.69**	**91.83**	**91.80**	**77.21**	**76.70**	**91.86**	**92.38**	**81.71**

3.3 Ablation Study

We compare ACDM with its three variants and VGAE on all benchmark datasets to answer the remaining questions posted at the beginning of this section. From the results in Table 3, we can draw the following conclusions:

(1) Comparing the result of $ACDM-k$ and ACDM, for NMI, F1, ACC and ARI, ACDM achieves maximum improvements of 28.39%, 23.98%, 24.99% and 38.76% on ACM and average boosts of 11.73%, 10.31%, 13.74% and 17.6% on all other benchmark datasets. Obviously, these results show that the KNN layer plays a vital role in ACDM, because the KNN layer, which is designed when we are inspired by the consistency of attributes, can fully describe the complementary information among nodes in different layers.

(2) ACDM always outperforms $ACDM-h/r$ on all datasets, which implies the specific objective function, based on homogeneity of attributes, helps communities to have the feature of similar intra-attributes. For example, for NMI, F1 ACC and ARI, ACDM achieves average improvements of 2%, 3.4%, 5.9% and 3.5% across all benchmark datasets.

(3) On all datasets, $ACDM-h/r$ is generally outperforms $ACDM-h/r/H$ on most metrics, indicating the need to emphasize the difference between particularity and commonality representations via HSIC.

4 CONCLUSION

In this paper, we find that most of methods focus on exploiting the complementary information from different topologies to extract commonality representations of nodes without fully exploring the role of attributes. To this end, we propose a novel method, i.e., ACDM, to make full use of two characteristics of attributes to detect communities. The extensive experimental data demonstrates the effectiveness of ACDM.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant U1811263, and Grant 62077045.

REFERENCES

[1] Xiaoke Ma, Di Dong, and Quan Wang. 2019. Community Detection in Multi-Layer Networks Using Joint Nonnegative Matrix Factorization. IEEE Trans. Knowl. Data Eng., 31(2), 273-286. https://doi.org/10.1109/TKDE.2018.2832205.

[2] Shaohua Fan, Xiao Wang, Chuan Shi, Emiao Lu, Ken Lin, and Bai Wang. 2020. One2Multi Graph Autoencoder for Multi-view Graph Clustering. In WWW '20: The Web Conference 2020, 3070-3076. https://doi.org/10.1145/3366423.3380079.

[3] Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. 2020. Unsupervised Attributed Multiplex Network Embedding. In the Thirty-Fourth AAAI Conference on Artificial Intelligence, 5371-5378.

[4] Zhiping Lin, Zhao Kang, Lizong Zhang, and Ling Tian. 2021. Multi-view Attributed Graph Clustering. IEEE Transactions on Knowledge and Data Engineering, 1-1. https://doi.org/10.1109/TKDE.2021.3101227.

[5] Zhiping Lin, and Zhao Kang. 2021. Graph Filter-based Multi-view Attributed Graph Clustering. In proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2723-2729. https://doi.org/10.24963/ijcai.2021/375.

[6] Xiao Wang, Di Jin, Xiaochun Cao, Liang Yang, and Weixiong Zhang. 2016. Semantic Community Identification in Large Attribute Networks. In proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 265-271.

[7] Arthur Gretton, Olivier Bousquet, Alexander J. Smola, and Bernhard Schölkopf. 2005. Measuring Statistical Dependence with Hilbert-Schmidt Norms. In Algorithmic Learning Theory, 16th International Conference, 3734, 63-77. https://doi.org/10.1007/11564089_7.

[8] Thomas N. Kipf, and Max Welling. 2016. Variational Graph Auto-Encoders, arXiv preprint arXiv:1611.07308 (2016).

[9] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S. Yu. 2019. Heterogeneous Graph Attention Network. In The World Wide Web Conference, 2022-2032. https://doi.org/10.1145/3308558.3313562.

[10] Thomas N. Kipf, and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In 5th International Conference on Learning Representations.

[11] Hongming Zhang, Liwei Qiu, Lingling Yi, and Yangqiu Song. 2018. Scalable Multiplex Network Embedding, In proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 3082-3088. https://doi.org/10.24963/ijcai.2018/428.

[12] Weiyi Liu, Pinyu Chen, Sailung Yeung, Toyotaro Suzumura, and Lingli Chen. 2017. Principled Multilayer Network Embedding. In 2017 IEEE International Conference on Data Mining Workshops, 134-141. https://doi.org/10.1109/ICDMW.2017.23.

[13] Baoyu Jing, Chanyoung Park, and Hanghang Tong. 2021. HDMI: High-order Deep Multiplex Infomax. In WWW '21: The Web Conference 2021, 2414-2424. https://doi.org/10.1145/3442381.3449971.

[14] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. 2014. Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition. In proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2149-2155.

[15] Feiping Nie, Jing Li, and Xuelong Li. 2017. Self-weighted Multiview Clustering with Multiple Graphs. In proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2564-2570. https://doi.org/10.24963/ijcai.2017/357.