A Survey of Discovering Community Structure in Multilayer Network - Reviewing the Existing Algorithms

Er. Ridhi Bhatt[1] Dr. Kumar Gaurav[2]
[1,2]Department of Electronics & Telecommunication Engineering
[1,2]Harcourt Butler Technical University, India

Abstract— Community detection has become an evolutionary field in the network sciences that have attracted researchers. In present arena complex networks are used to represent the information system. The salient feature of the complex networks is that they naturally contain a community structure. As use of the web is increasing and a large number of social networking sites are surfacing this allows the users to interconnect on a single platform. In this paper an attempt is made to represent a multilayer network their properties and categorization of methods. It also discusses each algorithm on each type of method discussed. In a nut shell, we can say that the extraction of community in multilayer networks is significant in studying complex networks.

Keywords: Homophily, Interlayer, Intralayer, Correlation

I. INTRODUCTION

The theory of network is a dominant tool to model and interpret the complex networks in different disciplines. To extract community structure in single layer network many theories have been developed but with the development of research, the scholars have realized that the only study of communities in a single layer network doesn't provide essential results in analyzing the structures and behavior in real life applications. As the complex networks in any field such as biology, economics, social sciences etc all consists of multiple type of relations among different entities. Community in a network can be defined as the set of nodes that can be classified into groups of nodes such that each group of nodes can be internally connected. The nodes correspond to the individuals, and the interactions among the individuals are represented by the edges also the nodes are arranged into layers and edges connect nodes in the same layer or in different layers that forms a multilayer network. This interaction of nodes in different layer can be explained with the Homophily i.e. the inclination of persons with alike ideas, behaviors, and fondness to get associated with each other that results in the formation of communities. Thus, with the help of community detection we can disclose the community structure of any network. It is used for studying biological network ,transport network that include community groups on the basis of using same route, same transport vehicle, same station etc which results in different layers of the transport system. It also helps in studying hidden relations and false relations among the nodes in any network. Analyzing communities is very helpful for studying information diffusion processes in the network like rumor spreading or epidemic spreading and to take steps to reduce it. It is often noticed that some links may not get observed or some links may enter into the data falsely. These types of problems are coped by community detection algorithm.

II. LITERATURE SURVEY

Detection of community structure with accurate results in complex networks such as multilayer with more information or interactions among the layers has been an enormous point of discussion and research. As a result, a large number of papers are publishing in this area so as to obtain reasonable partitions among large scale multilayer network. Apart from this, the algorithms with high analytical difficulty dealing with temporal networks with many layers revealing their community structure are still budding.

The exploration of complex network steered from the graph theory focusing on the "Seven Bridges of Konigsberg" problem in 1736. In 1998 with the introduction of graph theory in network the representation of multilayer network becomes more realistic. In 2002 Girvan and Newman stumble on community structures that lead to upswing of pertinent research. Although there is a plethora of algorithms that have been put forward but still large multilayer networks encounter some significant issues.

In [1] Kyu-Min Lee , Byungjoon Min and K-I Goh (2015) had reviewed the literature on multilayer networks according to the layer coupling or from the statistical physics point of view. Also they discussed the two major categories of layers i.e. the interlayer and intralayer and also highlighted the difference between them.

In [2] Manlio De Domenico .et al (2016) have very tremendously analyzed the spreading processes on multilayer network and focuses on the physical phenomenon that emerged from the multilayer networks and had led to the spreading processes which has been hidden in the graph theory analysis.

In [3] Xiaoming Li et.al (2017) had proposed detection of local community model based on influence relation by joining the indirect influence relation and direct influence relation of network i.e. interlayer and intralayer relationships. It gives prominent results i.e. better time efficiency to sparse network connection and large datasets.

In [4] Fatemeh Alimadadi, Ehsan Khadangi and Alireza Bagheri (2019) proposed extraction of community structure algorithm in multilayer network which is a generalization of label propagation algorithm(LPA) that yields better results for the activity network.

In [5] Subhadeep Paul and Yuguo Chen (2020) put forward a hypothesis for selecting an appropriate null model for data which defines modularity measures, model likelihood based quality functions which is utilized to detect the community structure.

Any network includes groups of community based on common places, ideas, jobs, etc. to find the community lying in a network, is remarkable feature for the analysis. The presence of communities also enables to study different actions like spread of any disease, rumors etc on a network.

Therefore, if one want to appropriately understand such actions, extracting community structure dominates. In the nutshell, the multilayer networks are helpful to model complex network and their analysis. They provide a powerful structure to study the manifoldness requires in studying any network by organizing nodes into layers. Thus, we can conclude that the multilayer networks are flexible to define the different types of interactions which single layer networks is unable to represent.

III. Representation of Multi-Layer Network

Multilayer network[1] represents multiple types of relations where different types of relation are represented by different network layer. It is a class of linked nodes that are shared in different layers. They help in modeling the complex networks that are encountered in the field of social sciences, biology, economics and machine learning etc. For example relationships among people include different types of interconnections such as relationships between members of the family, companions, etc that forms the different layers of the social system. The significance of multilayer network lies in its flexibility to disclose the thorough details than the monolayer network. The interaction among the nodes may be directed or undirected is represented by the edges. They are classified into 2 categories:

A. Intralayer edges

In the intralayer edges the edges connect the nodes in the same layer. This can be explained in the case of social network an interlayer edge represents the friendship between two individuals on a face book.

B. Interlayer edges

In the interlayer edges the edges connect the nodes in the different layers. - This can be explained in the case of social network an intralayer edge represents an individual switching from face book to use Instagram. The group of nodes and edges together forms the layer in multilayer network. They are classified into 2 categories:

C. Ordinal Layers[6]

The layers are arranged in a unique pattern and are connected to the correlated nodes in the adjoining layers by the interlayer edges. For example- the network that changes with time, the different layers represents the different activities and the time sequence represents the arrangement of layers.

D. Categorical layers[7]

The layers are classified into different sets where each set represents a specific interaction. For example- diffusion of disease first layer represents how the disease is transmitted, the second represents interactions of prevention to be taken and the third layer shows whether the person is aware of or not of the disease.

As we know that the representation of multilayer network is quite complicated. Two kinds of models are defined for representing the multilayer networks. A graph (Bollobas 2013) is a tuple (finite ordered list) G= (N, D) where N is a group of nodes; D belongs to N*N represents set of edges connecting pair of nodes.

Kivela et.al (2014) is more prominent for real life problems. It is explained as M= (N_M,D_M,N,L) where each notation is described as :N_M- nodes in multilayer network, D_M- edges in multilayer network, N- nodes in monolayer network, L- collection of layers. N_M □ D*L_1*L_2*……..*L_d, D_M□N_M*N_M; L= $\{L_\alpha\}^d_{\alpha=1}$ α is aspect ;L_α = L_1*L_2…..*L_d is an elementary layer. As it focuses on real life problems it is more prominent.

Boccaletti Model[3] is quite easy to realize. It can be represented with the help of supra adjacency matrix which is quite similar as we studied in monolayer network.

$$M=\begin{bmatrix} B_1 & B_{12} & \cdots & B_{1N} \\ B_{21} & B_2 & \cdots & B_{2N} \\ \vdots & \vdots & & \vdots \\ B_{N1} & B_{N2} & \cdots & B_N \end{bmatrix}$$

A1, A2,…AL represents how the layer 1, 2,….L are arranged in the form of matrix ; N represents total nodes. $I_{\alpha\beta}$ is the non diagonal block that shows the edges of layer a and layer b.

IV. Properties

Once we extract the structure of the community a variety of properties of the network topology can be computed easily from it. There are many statistical properties of network systems that help in the analysis of network data from a variety of fields and reveal some exceptional attributes and patterns that play prime role in the study of community. The statistical properties used in monolayer are reformulated in multilayer networks as the multilayer networks are complex as compared to monolayer networks.

Centrality[3]- It plays an important role as it calculates node's degree, as the node with a large number of links is considered as the prominent node. Node's centrality "c" is defined by a vector k_c= (k_c^1, k_c^2,……., kcL), where k_c^L is node's degree 'c' in the L^{th} layer. It's an important feature but poses the computational complexity when the nodes are different in different layers.

Correlations[3]- As we know, we can extract more information from Multilayer networks as compared to monolayer networks since there is correlation among the nodes in different layers. This may include interlayer degree correlations, layers overlapping etc. Overlapping is a salient feature to consider as each layers in the communities gets influence in several ways including assortative, disassortative, directed structure. Local overlap (Cellai et al.2016) is defined as $O^{ab}{}_y$= ∑υ $(w^a{}_{yz})\theta(w^b{}_{yz})$ **;**$O^{ab}{}_y$= overlapping edges that connects node 'c' in a as well as b layers; $w^a{}_{yz}$,$w^b{}_{yz}$ =weights of edges (r,s) in a as well as b layers respectively. Thus, correlation of nodes is an important factor in studying the multilayer networks.

V. Categorization of Algorithms

The approaches that were introduced initially: - a) either applies the monolayer algorithms to different layers or then by using clustering merge them .b) or integrate the multilayer

network into a monolayer network. These approaches results in low accuracy as the interlayer connections have been overlooked by them. Scholars have directed their research in the direction to propose the algorithm that can be applied to extract the community structure in the multilayer network directly. This has been categorized into three (Tagarelli et al. 2017)

METHODS	DESCRIPTION
Flattening method	It merges all the information provided in the different layers into a monolayer and then single layer algorithms are applied to extract the community structure. It can be applied to multiplex networks.
Aggregation method	Each layer community is extracted and is integrated by using certain aggregation method. In this process some repeated layers also gets computed (Kao and Porter 2018) put forward layer communities grouping method to avoid the repeating layers.
Direct method	In this method community can be directly extracted in the multilayer network with the help of some criteria for ex- multilayer modularity index is collaborated with the improved GN and Louvain algorithms.

VI. PROPOSED ALGORITHM ON EACH METHOD

A. Particle competition for multilayer network-

It consists of the following steps:

- Adjacent matrix is obtained for a multilayer network.
- As we are unaware of communities present so let there be T communities present.
- Now the probabilities of preferential and random walk are calculated using $p^k_r(y,z)=w_{i,j}/\sum^V_{u=0} w_{y,z}$; $p^k_p(y,z,t)=w_{y,z}N^k_j(t)/\sum^V_{u=0} w_{y,z}N^k u(t)$
- The particle competition system is computed for each T until it converges.

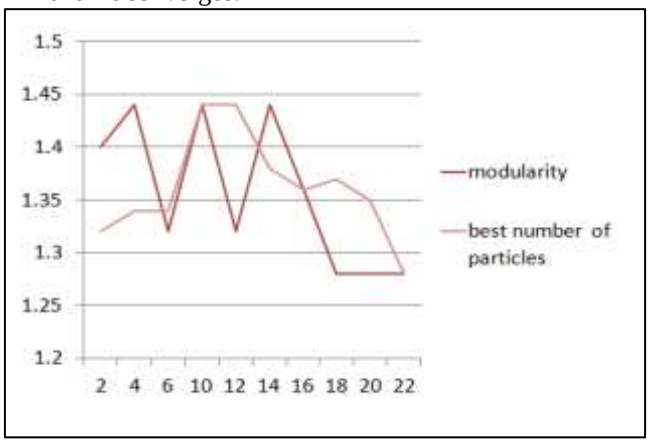

Fig. 1: horizontal line shows number of particles and vertical line shows modularity

B. Multilayer community detection Ensemble based

It consists of the following steps:

- A multilayer network is given in which co-association matrix T is defined as common community between two nodes divided by the group of layers and node's clustering[10] is done.
- The significant co-association values are kept while lower values are neglected.
- T is cut according to minimum co-association [0, 1].
- The community structure gets extracted.

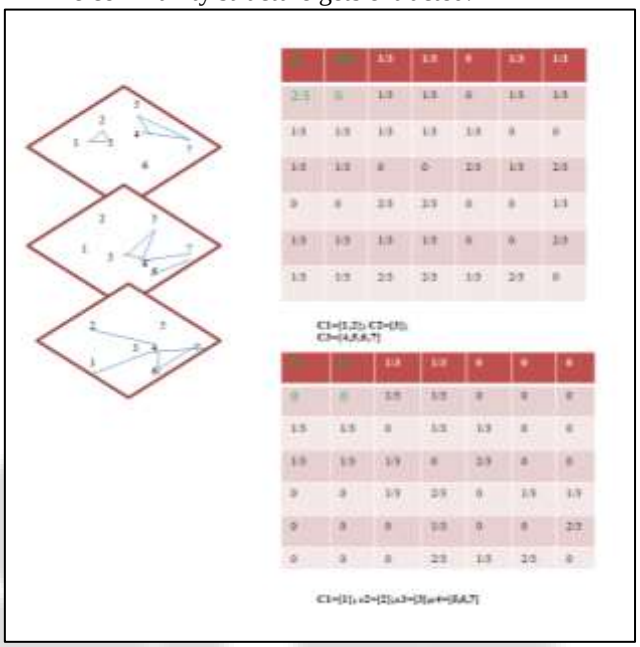

Fig. 2: multilayer network and its co-association matrix and community structure with(up) and without(down)node linkage

C. Multilayer network label propagation algorithm (MNLPA)

- It consists of the following steps:
- Nodes are randomly ordered and a unique label i.e. the information is used to initiate each node.
- Neighbors of same labels in all layers are obtained.
- Similarity[8] between the nodes is calculated.
- The steps are repeated till stop criterion is satisfied.
- The communities are divided on the basis of nodes of the same label.

In LPA[9] (label propagation model) a label is hold by every node which is updated iteratively until the algorithm converges and community structure is extracted. It is not possible to compute multilayer network on the basis of NMI, F_score. Therefore LPA and MNLPA are compared on the basis of homophily.

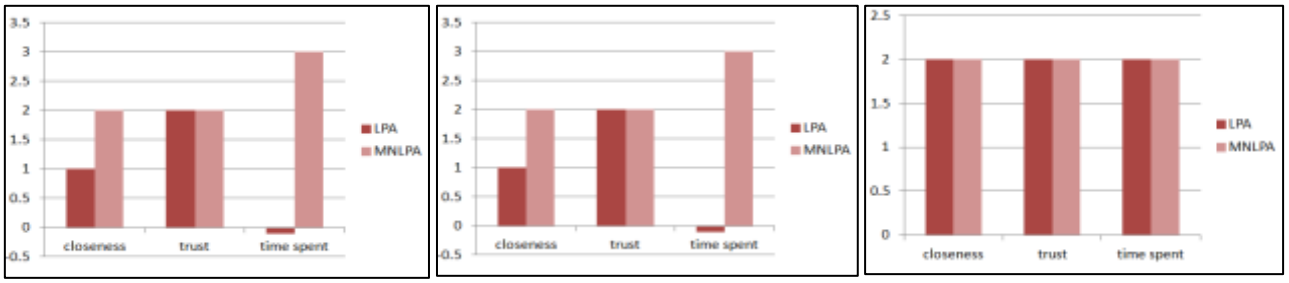

Fig. 3: LPA and MNLPA algorithms conducted on face book data are compared

D. Collation of multilayer network algorithms

Algorithm	Method	Based on	Complexity	Network
Particle competition for multilayer network	Flattening	Competition among the particles	$O(dTCl^2)$	Multilayer , directional, weight
Multilayer community detection ensemble based	Aggregation	Maximizing the modularity	O(T (e+lC))	Multilayer
Multilayer network label propagation	Direct	Propagation of the label	O(dN)	Multilayer, directional, weight

d -number of nodes; e -group of edges; l -group of layers ;N- node's average degree; T -iteration times ;C -number of communities.

VII. Conclusion

Multilayer network plays a pivotal role in representing complex networks as there are nodes that represent the individual and edges that correspond to the interaction among the nodes. There are many real life networks: transportation network, social network, biological network etc where there are communities present and scholars are dedicating their work to extract the communities directly. In this paper we discussed about the representation of the multilayer network with the mathematical approach. We focus on three community detection algorithms such as MNLPA[8], EMCD[10], ParticleGao[9] and compare them on the parameters including classification, strategy, complexity and network. There are many proposed algorithms to detect community in multilayer network but still a great deal of work has to be done in this field to find highly efficient and high accuracy algorithms.

References

[1] De Domenico, Manlio, et al. "The physics of spreading processes in multilayer networks." Nature Physics 12.10 (2016): 901-906.
[2] Lee, Kyu-Min, Byungjoon Min, and Kwang-Il Goh. "Towards real-world complexity: an introduction to multiplex networks." The European Physical Journal B 88.2 (2015): 1-20.
[3] Lin, Shuyang, et al. "Understanding community effects on information diffusion." Pacific-Asia conference on knowledge discovery and data mining. Springer, Cham, 2015.
[4] Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." Proceedings of the national academy of sciences 99.12 (2002): 7821-7826..
[5] Ovens, Katie, B. Frank Eames, and Ian McQuillan. "Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution." Frontiers in Genetics 12 (2021).
[6] Huang, Xinyu, et al. "A survey of community detection methods in multilayer networks." Data Mining and Knowledge Discovery 35.1 (2021): 1-45.
[7] Ovens, Katie, B. Frank Eames, and Ian McQuillan. "Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution." Frontiers in Genetics 12 (2021).
[8] Alimadadi, Fatemeh, Ehsan Khadangi, and Alireza Bagheri. "Community detection in facebook activity networks and presenting a new multilayer label propagation algorithm for community detection." International Journal of Modern Physics B 33.10 (2019): 1950089.
[9] Gao, Xubo, et al. "Particle competition for multilayer network community detection." Proceedings of the 2019 11th international conference on machine learning and computing. 2019.
[10] Tagarelli, Andrea, Alessia Amelio, and Francesco Gullo. "Ensemble-based community detection in multilayer networks." Data Mining and Knowledge Discovery 31.5 (2017): 1506-1543.

