論 文 の 内 容 の 要 旨

論文題目 The Network Generation Model based on Multilayer Networks
（マルチレイヤーネットワークに基づくネットワーク生成モデル）

氏　　名　　范　超

1. Introduction

Various networks can be defined in real world, such as social network, traffic network, scientific co-authorship network, etc. Many phenomena are observed on these complex networks and they have a strong influence on society. These phenomena are correlated with network structure and the mechanism underlying network generation. This research aims to clarify the mechanism of network generation and the relationship between phenomena occurring on the network and the network structure. In order to comprehend phenomena existing on networks, there is a need to explore the methodology of modeling real networks.

Since current models focusing on single layer cannot precisely represent high-modularity networks, this research proposed a high-modularity network generation model based on a multilayer network. As people belong to many communities in society, such as family, school, hobby group, and business organizations, each example is regarded as a community in a single layer of a multilayer network. However, measuring each relationship in each community is difficult. A network on SNSs that can be observed combines all communities. That is, a social network is generated from a multilayer network. A synthesized network in the model has either a community structure or a high-modularity structure. This research applied the proposed model to generate a number of networks and compared them with real-world networks. Not only did it successfully represent real-world networks but it also predicted how real-world networks are generated from the model's parameters. Moreover, information diffusion experiments are carried out and the average influence degree (AID) is calculated. Finally, a "*C-A-R*" process is proposed to evaluate accuracy of information diffusion on multilayer network. I will also apply proposed multilayer model to various types of real-world data and try to better elaborate the relationship between network structures and social phenomena which has not been clarified so far.

2. Related Work

Many previous studies have generated synthetic networks with community structures or high modularity. Lancichinetti et al. introduced a standard model called LFR benchmark that synthesizes networks with planted community structures. Pasta et al. proposed a tunable and growing network generation with community structures. Leskovec described a Kronecker graph [4]: a model to produce a hierarchical structure in a real network.

As for information diffusion, there are many methods modeling spreading processes in multilayer networks, which can be grouped into two types: epidemic-like and decision-based models. In epidemic-like models, the infection probability of a node is decided by its neighborhood. Many epidemic models like SIR, SIS and SI1I2R has been employed for modeling diffusion process over multilayer networks. On the other hand, an agent adopt a behavior is based on the decisions of its neighbors in decision-based models. In this research, a generalized independent cascade model (ICM) is utilized to carry out information spread simulation over multilayer network.

3. High-Modularity Network Generation Model Based on the Multilayer Network

In proposed model, a social network is regarded as a multilayer network. A layer represents one kind of relationship in a social interaction (**Fig. 1**). The networks obtained from social networks are superimposed networks of all the small networks that we call communities.

Each node in the superimposed network belongs to a few communities. Nodes are connected to other nodes that belong to the shared communities. The link between two nodes in the superimposed network can only appear when there is a link between two nodes in some communities. An assumption is also made that some communities are exclusive. For example, most people only belong to one university or one company. Such communities are called layers.

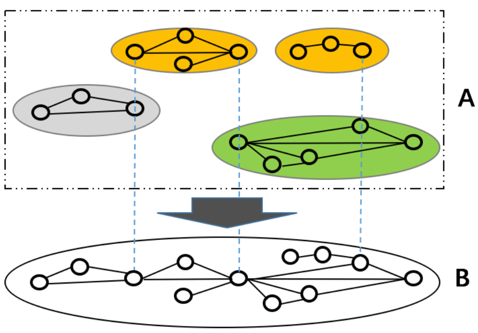

Fig. 1 Generating process of model

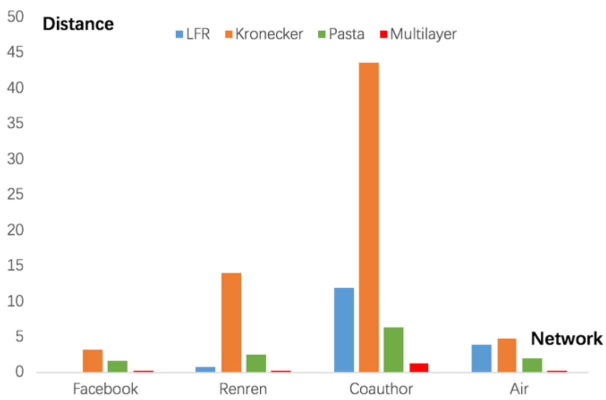

Fig. 2 Comparison of distance for different models

3.1 Model Parameters

1. Community size distribution: The distribution of the community sizes is taken from power-law distribution $p(k) \sim k^{-\beta}$ with exponent β where $1 \leq \beta \leq 2$.

2. Number of layers: In this paper, this value s ranges between 2 and 10 and is optimized by maximizing the similarity between synthesized and real networks.

3. Inner community network model: Each node in a community are connected by links as a network. Following models are used to connect nodes in a community: ER model, WS model, BA model, CNN model, a complete network and their combinations.

4. Inter layer degree correlation. In different layer, each node has different number of links. Here, whether each node has a degree correlation in each layer is considered.

3.2 Evaluation Function

Evaluation function optimizes the parameter values by minimizing the distance between the real-world data and the generated data in the proposed model. Five network features are employed to quantitatively evaluate the distance D of two networks: (1) Clustering coefficient C; (2) Assortativity r; (3) Modularity Q; (4) Power index of degree distribution γ; (5) Coefficient of determination of degree distribution R^2. Distance D is a normalized Euclidean distance between produced network G_i and target network G_0.

$$D(G_i, G_0) = \left(\frac{C_i - C_0}{\sigma_C}\right)^2 + \left(\frac{r_i - r_0}{\sigma_r}\right)^2 + \left(\frac{Q_i - Q_0}{\sigma_Q}\right)^2 + \left(\frac{\gamma_i - \gamma_0}{\sigma_\gamma}\right)^2 + \left(\frac{{R^2}_i - {R^2}_0}{\sigma_{R^2}}\right)^2$$

3.3 Network Datasets

1. Facebook network: The network used in experiments contained 8,578 nodes and 405,450 links and reflects the social relationship within Yale University.
2. Renren network: Renren network is a Facebook-like SNS network in China. The network data have 2,309 nodes and 60,532 links. These users were all 2009 Peking University (PKU) graduates.
3. Collaboration network: It is a scientific co-authorship network called "ca-GrQc" (Arxiv General Relativity and Quantum Cosmology), which is an undirected network with 5,242 nodes and 14,496 links. The data cover papers from January 1993 to April 2003.
4. Air traffic control network: This network dataset is part of the Koblenz Network Collection. It was constructed from the USA's Federal Aviation Administration (FAA) National Flight Data Center (NFDC), Preferred Routes Database. 1,226 nodes and 2,408 links were identified in 2016.

3.4 Experimental Results and Analysis

In order to mimic four real networks, algorithm traverses all the possible discrete values of the parameters and minimizes the distances *D* for both the baseline and proposed models. The distances between the generated and real networks with different models are shown in **Fig. 2**.

According to **Fig. 2**, this model has the shortest distance for four datasets, thus it outperforms other models. A multilayer model is capable of reproducing two social networks (Facebook and Renren), a co-authorship network and an air traffic control network with small distance *D*.

Depending on simulation results, it can be predicted how real-world networks are generated from model's parameters. The data shown in **Table 1** give the distance from a scientific co-authorship network for different inner community network models. A combination of the CNN and CPN models as an inner community model performed the best. The reason lies in the mechanism behind the co-authorship network. All authors should be acquainted with each other if they co-author a paper. When constructing a network, complete networks fit fairly well with small communities, which connect all of the co-author nodes who collaborate on a paper. Additionally, the CNN model creates networks with a high clustering coefficient as well as scale-free properties and captures the basic features of larger communities in co-authorship networks.

Table 1 Distance from co-authorship network (ca-GrQc) with different inner model combinations

Distance D	ER	BA	WS	CNN	CPN
ER	14.1916	14.6491	13.9773	13.7575	6.9651
BA	16.6345	16.5003	16.5852	16.6140	4.7510
WS	6.1614	9.9046	9.9789	8.3799	6.2434
CNN	2.3690	4.2168	4.1124	4.1100	**1.3256**

4. Evaluation of Accuracy of Information Diffusion on Multilayer Network

In this part, simulation of information diffusion over multilayer network is conducted to evaluate the accuracy of propagation, utilizing an extended independent cascade model (ICM). The propagation process in multilayer network can be depicted in **Fig. 3**.

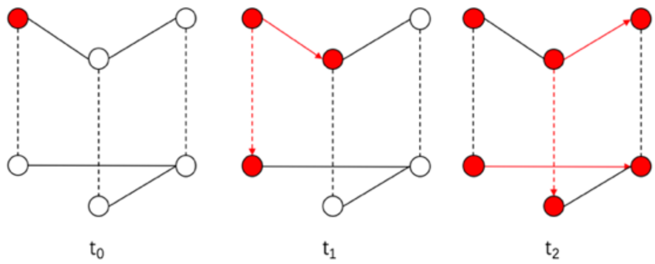

Fig. 3 Illustration of information propagation in multilayer networks

According to **Fig.3**, an across-layers propagation happens at time t_1 and t_2. Here, the cross-layers probability is equal to intra-layer probability. If an inactive node is successfully activated by its neighbors, it becomes active and will disseminate information to both intra-layer and across-layers neighbor nodes. An average influence degree (AID) σ of spreading process is calculated by:

$$\sigma = \frac{\sum_{v=0}^{N} \sigma(v)}{N}$$

where the influence degree $\sigma(v)$ of a node v is defined as the expected number of active nodes at the end of the information diffusion process that starts from a single initial activated node v. The network owns a high propagation ability if the value of AID is high.

For the purpose of evaluating the accuracy of information diffusion on multilayer network, a "*O-A-R*" process is proposed. The first "*O*" expresses the construction of an original multilayer network. Information spreading is performed on the built-up multilayer network and its AID is calculated as baseline. The second "*A*" represents the aggregation process of original multilayer network. All layers of a multilayer network are aggregated into one layer and the AID of monolayer network is computed. The final "*R*" means the reconstruction of a multilayer network. The multilayer structure is rebuilt by proposed model in previous part using the network index of aggregated single layer. By comparing the AID of rebuilt multilayer network with baseline in the original multilayer network, it can be used to evaluate the accuracy of information diffusion on multilayer network. In my simulation experiments, two real multilayer networks (CKM physicians' innovation network & London multiplex transport network) were used to build original multilayer networks in "*O*" step and comparison of information diffusion was made in three steps. Finally, experimental results led to a conclusion that the multilayer-based method can model original multilayer networks effectively.

5. Conclusion

This research made an effort to deal with the problem of network generation model. An effective high-modularity network generation model is proposed by layer aggregation based on a multilayer network. Moreover, simulation of information propagation is carried out. Finally, a "*O-A-R*" process is proposed to construct, aggregate and rebuild a multilayer network in order to evaluate the accuracy of information diffusion on multilayer network.

In future, some work will deal with datasets that have more nodes and links on a large scale. Since weight in the evaluation function was not considered, future work will effectively employ it to emphasize specific features. Other network features may also subsequently be introduced into the evaluation function. Further, more studies will be done to consider different types of propagation and more complicated information spread model in multilayer network.