Methods of community detection in hybrid network applied to the academic network analysis

Métodos de detección de comunidad en redes híbridas aplicado al análisis de redes académicas

Lu Jiang [1,2] https://orcid.org/0000-0002-5514-9860

Jorge Gulín-González[3] https://orcid.org/0000-0001-7912-2665

Edisel Navas Conyedo[3] https://orcid.org/0000-0002-4648-7622

Yunwei Chen [1,2*] https://orcid.org/0000-0002-6597-7416

[1]Scientometrics & Evaluation Research Center (SERC), Chengdu Library and Information Center of Chinese Academy of Sciences. No.16, South Section 2, Yihuan Road, Chengdu 610041, China. jianglu@mail.las.ac.cn.

[2]Department of Library Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences. G111, Beijing 101400, China.

[3]Centro de Estudios de Matemática Computacional (CEMC), Universidad de Las Ciencias Informáticas (UCI). Carretera a San A. de los Baños, Km 2½. Torrens, La Lisa, La Habana, Cuba. gulinj@uci.cu

*Autor para la correspondencia. (chenyw@clas.ac.cn)

Editorial "Ediciones Futuro"
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

381

ABSTRACT

In the real-world, multiple-typed objects are interconnected, making heterogeneous information networks. Hybrid network is a heterogeneous network which contains a variety of node types and a variety of relationship types. The concept of heterogeneous network emphasizes the complexity at the level of network structure, while the network formed by the integration of multiple nodes and multiple relationships emphasizes the richness of studied functions. The characteristics of the multi-node and multi-relationship hybrid network are mainly reflected in the following two aspects: first, the diversity of nodes, including a variety of node types and, second, the richness of relationships. Community structure detection is a method used to identify clusters of nodes in a network. Community structure detection is the most widely studied structural features of complex networks. In this paper, we present a review about the method of community detection in hybrid network, and the application to the network analysis. Here, we present a revision of the main techniques, methods, data sets and algorithms used in the literature and, a short description of the general features of the community detection topics. Main authors and their contributions to this fields are introduced. The study is focused on the analysis of the multi-node and multi-relationship hybrid networks with the idea to apply these results to the academic network analysis. Considering the revised literature, we propose a hybrid network community detection algorithm based on meta-path, seed nodes and extend modularity method to study the academic networks.

Keywords: Hybrid network; Heterogeneous network; Community detection; Multi-node and Multi-relationship; Network Analysis.

RESUMEN

En el mundo real, múltiples tipos de objetos están interconectados, creando redes de información heterogéneas. Una red híbrida es una red heterogénea con una variedad de tipos de nodos y una variedad de tipos de relaciones. El concepto de red heterogénea enfatiza la complejidad a nivel estructural de la red y la riqueza de funciones estudiadas. Las características de la red híbrida de múltiples nodos y relaciones se

Editorial "Ediciones Futuro"
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

382

reflejan principalmente en los dos aspectos siguientes: primero, la diversidad de nodos, incluida la variedad de tipos de nodos y, segundo, la riqueza de las relaciones entre ellos. La detección de estructuras comunitarias es un método que se utiliza para identificar grupos de nodos en una red. La detección de estructuras comunitarias es la característica estructural más estudiada de las redes complejas. Aquí presentamos una revisión sobre el método de detección de comunidades en redes híbridas y la aplicación al análisis de redes. Presentamos una revisión de las principales técnicas, métodos, conjuntos de datos y algoritmos utilizados en la literatura y, una breve descripción de las características generales de los temas de detección de comunidades. Se presentan los principales autores y sus contribuciones a estos campos. El estudio se centra en el análisis de las redes híbridas multinodo y multirelación con la idea de aplicar estos resultados al análisis de redes académicas. Teniendo en cuenta la literatura, proponemos un algoritmo de detección de comunidad de red híbrida basado en meta-ruta, nodos semilla y método de modularidad extendida para estudiar las redes académicas.

Palabras Clave: Red Híbrida, Red Heterogénea, Detección de Comunidad, Relaciones multimodo y Multirrelación, Análisis de Red.

Introduction

Social network represents a set of individuals with relationships among them. In network terms, individuals are represented as nodes and relationships represented with ties (edges) between these individuals. An edge can represent a friendship, or any type of relationship between the individuals linked (Wassernman and Faust.1994). Particularly important in information sciences and scientometrics are the relationship based on authors' collaborations in a research paper (the collaboration networks) or based in the relationship between author collaboration networks and citation networks. Social network analysis is a method used in bibliometrics and scientometrics for the analysis of collaboration in scientific publications (Otte and Rousseau, 2002).

Editorial "Ediciones Futuro"   383
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Hybrid network is a heterogeneous network which contains a variety of node types or a variety of relationship types. The concept of heterogeneous network emphasizes the complexity at the level of network structure, while the network formed by the integration of multiple nodes and multiple relationships emphasizes the richness of functions. The characteristics of the multi-node and multi-relationship hybrid network are mainly reflected in the following two aspects: first, the diversity of nodes, including a variety of node types. For example, in the academic network, nodes can be authors, literature, keywords, journals, etc., in the medical network, nodes can be doctors, drugs, patients, etc. Second, the richness of relationships. For example, in an academic network, relationships can include author collaboration, author citation, literature citation and author-literature relationship, etc. In a medical network, relationships can include relationships between doctors prescribing drugs, patients taking drugs and so on.

Mathematically, a social network is a graph (a set of nodes and edges). One common type of graph is a k-partite graph. If k = 2 it is a bipartite graph, in which the vertices are partitioned into k-disjoint subsets, and each edge connects vertices in distinct partitions. Hence a bipartite graph is one in which the nodes can be divided into two sets U and V so that every edge in the graph connects a node in V with a node in U. Studying scientific collaborations networks as well as how papers cite others (citation networks) has become increasingly important. Understanding these networks help us better understand how scientific discoveries and innovations are communicated within the scientific community (Divakarmurthy and Menezes, 2013). On the other hand, clustering (or mapping) of scientific articles is an important area in scientometric research. Clustering analysis is useful for detecting research topics and revealing scientific structure and dynamics, which can aid in systematic understanding of research fields (Wang et al., 2013; Zhang et al, 2016).

Community structure detection is a method used to identify clusters of nodes in a network. Community structure detection is the most widely studied structural features of complex networks. At present, the research on community detection methods of multi-node and multi-relationship hybrid network mainly focuses on the following two kinds: one is to extend the existing algorithm to deal with the hybrid network directly, the other is to reduce the dimension of hybrid network to homogeneous network and then conduct community detection (Shah and Zaman, 2010). Based on the above two ideas, the community detection methods of multi-node and multi-relationship hybrid network mainly fall into the following five categories:

Editorial "Ediciones Futuro"   384
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

probabilistic model-based method, meta-path based method (Sun et al., 2011), seed nodes based method, extend modularity method, and homogeneous method of heterogeneous network (Shi et al, 2017). Probabilistic model-based method usually needs to specify the number of communities according to prior knowledge, which may lead to unstable results. Although the meta-path based method is relatively simple, the similarity obtained is usually a deviation measure. The extend modularity method is still unable to avoid the limitation of the maximization of modularity, which is the resolution limitation, so it cannot detect small communities in large-scale networks. For the method based on seed nodes, there are still some problems about how to select effective seed nodes efficiently. Although the homogeneous method of heterogeneous network is easy to understand, the derivation process of the model is complicated and the implementation complexity is too high. It can be seen that there is still a lot of room for further research on community detection methods of hybrid network.

Here, we present a review about the method of community detection in hybrid network, and the application to the academic network analysis. Our objective is to introduce to the reader in the most important concepts on the community detection topics, the most common techniques used and the future development of this emergent field. This paper is organized as follow: In Computational Methodology we present an overview about the construction of a hybrid academic network with several relationships. We also introduce the concept of modularity and its definition for different types of networks. In the Results and Discussion section we present some of the most relevant papers in the domain of Network Community Detection. Besides of the most popular techniques to perform network community detection are discussed. Finally, we briefly introduce examples of the literature in this field.

Computational Methodology

The source of the revised articles was de Web of Science database during the period 1980-2021. In our revision, we have prioritized the conceptual and theoretical papers on the field of network community detection, particularly the application to academic networks. Based on the Web of Science database, the retrieval formula was TS= ((hybrid network OR heterogeneous network) AND (community detection OR

Editorial "Ediciones Futuro"   385
Universidad de las Ciencias Informáticas. La Habana, Cuba
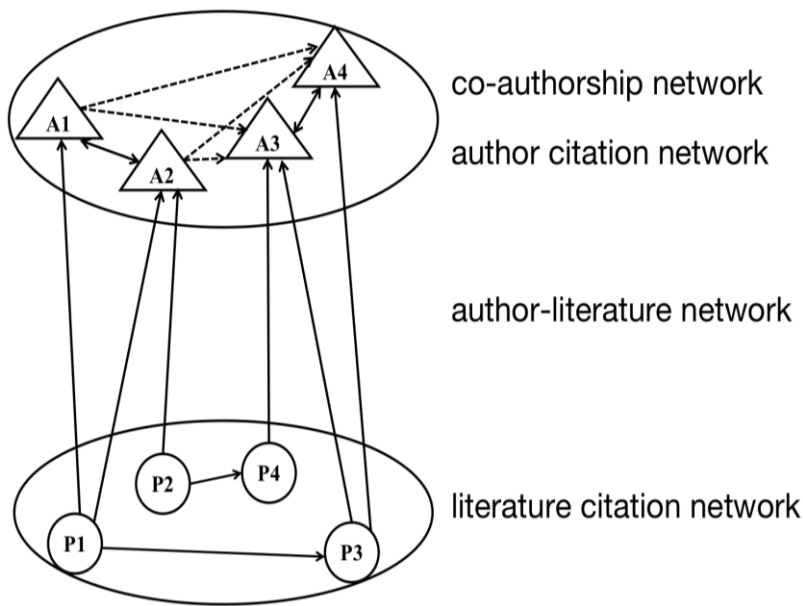rcci@uci.cu

community discovery)). Up to April 29, 2021, 625 literatures were retrieved. We focus on the literatures related to community detection in multi-node and multi-relation hybrid networks. 228 of the recovered articles are proceedings paper and the rest are articles (original researches in periodical journals or books chapter).

Definition of hybrid networks

The definition of network has been emerging endlessly. Complex networks (Newman, 2003) refer to a network with some or all of the properties like self-organization (Watanabe et al., 2015), self-similarity (Song et al., 2005; Gallos et al., 2013), attractor, small world phenomenon (Gallos et al., 2012), and scale-free degree. Since general network diagram cannot reflect the network structure truly and effectively, Estrada et al. (2006) introduced super networks, which can more accurately describe complex network structures. A heterogeneous network is a special type of network (Sun et al., 2009), which contains a variety of node types and relationship types. The corresponding homogeneous networks refer to the network containing only one node type and relationship type (Wolfe et al., 1993). Based on the above definitions, the concept of a hybrid network is adopted in this study, that is, a network containing multiple node types or multiple relationship types. Both complex networks and super networks point to a wide range of networks, which refer to a network that exhibits a high degree of complexity. Hybrid networks essentially belong to the category of heterogeneous network, but heterogeneous networks emphasize the complexity of network topology. It pays more attention to heterogeneous relations and pays less attention to isomorphic relations between nodes of the same type (Ran et al., 2013), Hybrid networks emphasize the mixing of multiple nodes and multiple relationships, fully consider heterogeneous relationships and isomorphic relationships, and reflect the richness of functions. Therefore, the definition of "hybrid networks" is adopted in this study; researchers can focus on the improvement of function instead of just focusing on the construction of network topology, in the case that it is more consistent with the real network.

Editorial "Ediciones Futuro"   386
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Constructing a hybrid network

First, we present an example of how two nodes of literature and author are selected to construct a hybrid network that includes four relationships among literature citation, author collaboration, author citation, and literature-author (as shown in Figure 1). Construct a co-authorship network based on the co-authors of literature, construct a literature citation network and author citation network based on the citation relationship between literature, and construct an author-literature bipartite network based on the affiliation between authors and literature (Figure 1). This network can not only reveal the vertical information in the time axis, but also present the "horizontal" correlation in the wide space, which shows the basis and internal dynamic mechanism of the author relationship construction.

Fig. 1 - A pictorial representation of a hybrid network.

Editorial "Ediciones Futuro"
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

387

Results and Discussion

Relevant Community Detection Studies

Community detection has involved an enormous consideration due to the growing availability of the data sets of the large-scale networks, particularly social and academic networks. To provide insightful information about community detection, much research has been conducted in the form of surveys, systematic literature reviews, and visual studies. But, only a few of them shows how the field advanced over time. To demonstrate the sense of details, information about existing literature is listed in Table 1 (Khan and Niazi, 2017).

Table 1 - Some representative papers on the domain of "Network Community Detection". Structure and some of the papers were selected from Ref. (Khan and Niazi, 2017).

Reference	Article type	Study Area
(Khan and Niazi, 2017)	Review and Visual Survey	Community structure
(Cai et al. 2016)	Survey	Evolutionary techniques for the identification communities in networks
(Fortunato and Hric, 2016)	User Guide	Identification of communities in networks
(Bedi and Sharma, 2016)	Advanced Review	Identifying communities in social networks
(Dhumal and Kamde, 2015)	Survey and empirical evaluation	Community identification in large-scale networks
(Drif and Boukerram, 2014)	Survey	Techniques for uncovering communities in social networks
(Ma et al., 2013)	Visual Analysis	Community discovery of multi-context mobile social networks
(Plantié and Crampes, 2013)	Survey	Social community identification

Editorial "Ediciones Futuro"   388
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Community Detection Techniques

Modularity

For the mathematical formulation of this subsection, we have followed (Khan and Niazi, 2017).

Usually, the techniques to find optimal communities in reasonably fast time are based on the optimization of objective functions. Modularity optimization is one of the most widely used techniques among them.

For unweighted networks, modularity Q can be defined as:

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(C_i, C_j) \qquad \text{Eq. 1}$$

In Eq. 1:

m is the number of links, k is the degree of vertex i, k is the degree of vertex j, C_i is the community to vertex i, C_j is the community to vert $\delta(C_i, C_j) = 1$ nd j belong to the same community, otherwise it equals to 0.

For weighted networks, modularity Q can be defined as:

$$Q = \frac{1}{2W}\sum_{i,j}\left[A_{ij} - \frac{s_i s_j}{2W}\right]\delta(C_i, C_j) \qquad \text{Eq. 2}$$

In Eq. 2, W is total weight of all of the links in the network, W_{ij} is the weight of the links between vertices i and j, s_i is the strength of vertex i, s is the strength of vertex j. $\delta(C_i, C_j) = 1$ j belong to the same community, otherwise it equals to 0 (Khan and Niazi, 2017).

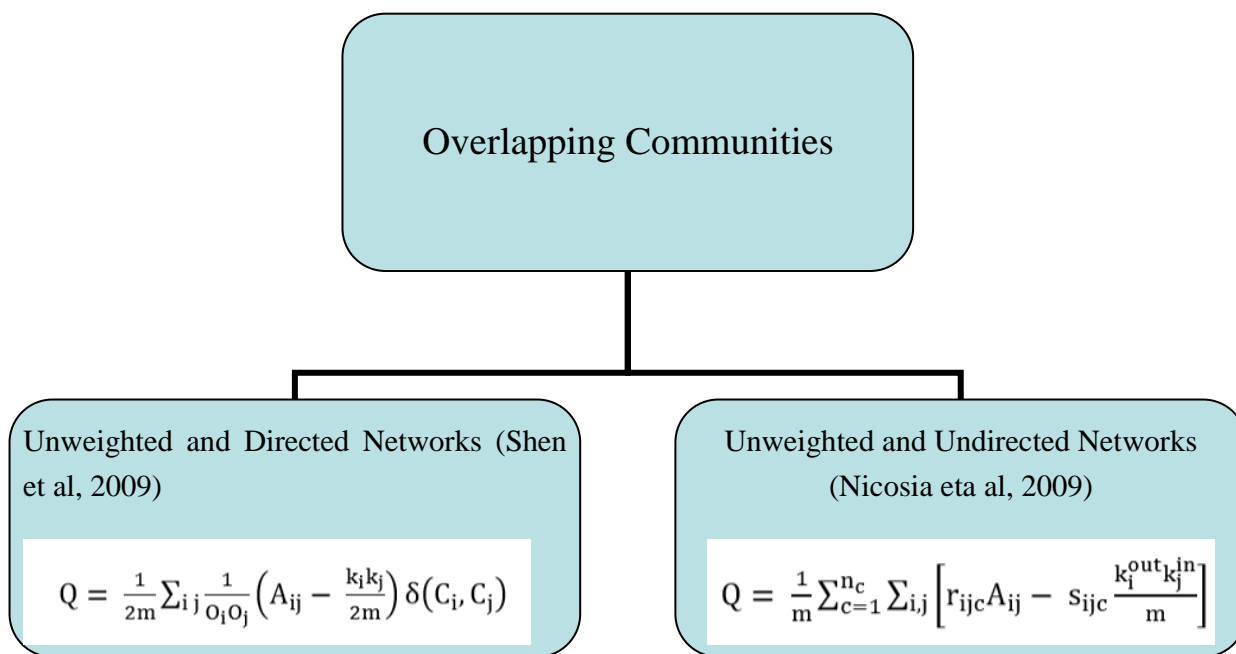If we consider now a weighted and direct network, the modularity is defined by (Arenas et al. 2007) as:

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i^{out} k_j^{in}}{m}\right]\delta(C_i, C_j) \qquad \text{Eq. 3}$$

In overlapping communities (unweighted and directed networks and unweighted and undirected networks) Shen et al. introduce the formulas showed in Figure 2:

Editorial "Ediciones Futuro"
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

389

A new quantitative measure named as modularity density (D), which is based on the density of subgraphs has been defined to solve community structure of networks (Li et al., 2008). The modularity density is defined as:

$$D_\lambda = \sum_{i=1}^{m} \frac{2\lambda L(V_i, V_i) - 2(1-\lambda)L(V_i - \overline{V_1})}{|V_i|}$$

Eq. 4

Where, V_i is the subset of V i=1, …, m, such that and $L(V_i, V_i) = \sum_{i \in V_i, j \in V_i} A_{ij}$ $\qquad L(V_i, \overline{V_1}) = \sum_{i \in V_i, j \in \overline{V_1}} A_{ij}$

Overlapping Communities

Unweighted and Directed Networks (Shen et al, 2009)

$$Q = \frac{1}{2m} \sum_{i\,j} \frac{1}{O_i O_j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j)$$

Unweighted and Undirected Networks (Nicosia eta al, 2009)

$$Q = \frac{1}{m} \sum_{c=1}^{n_c} \sum_{i,j} \left[r_{ijc} A_{ij} - s_{ijc} \frac{k_i^{out} k_j^{in}}{m} \right]$$
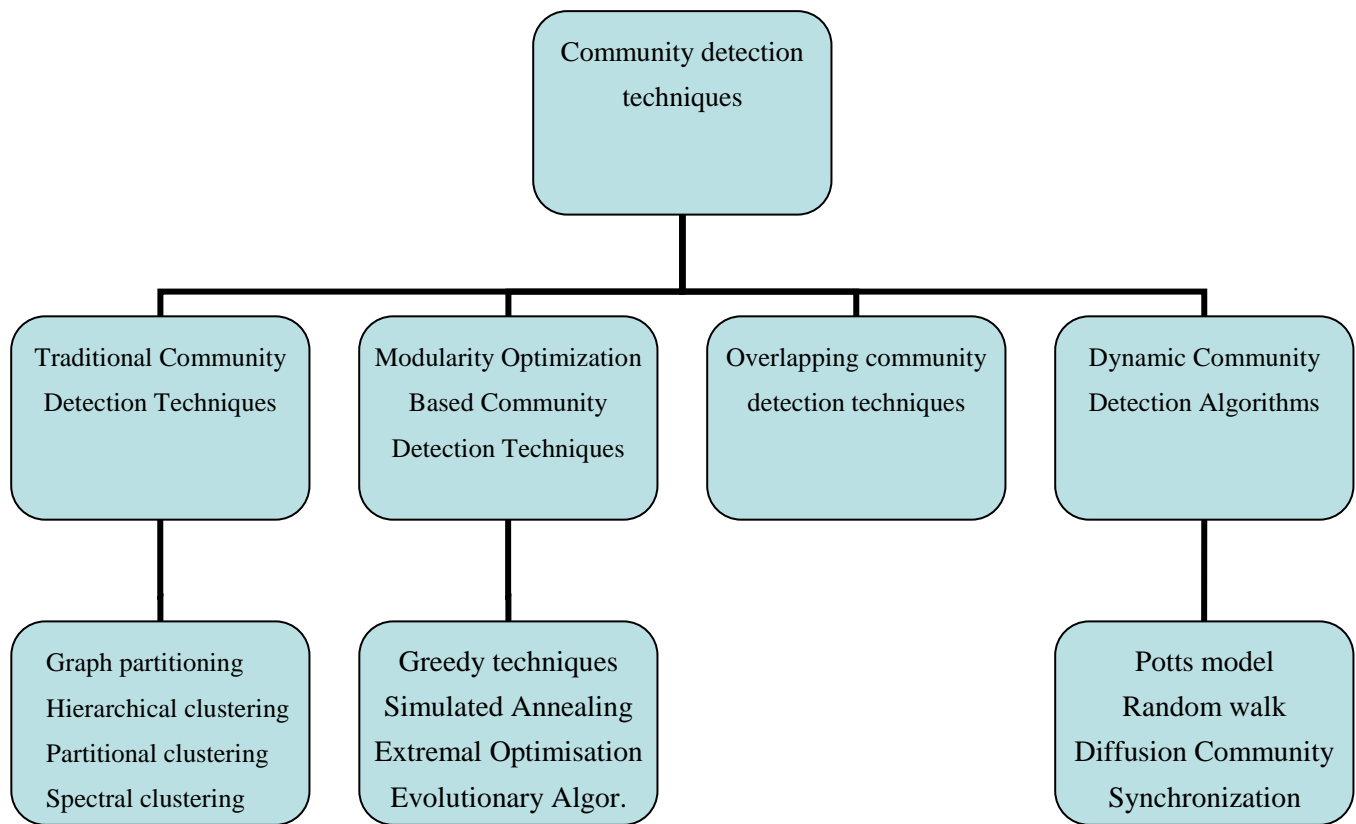
Fig. 2 - Definition of modularity (Q) for unweighted and directed networks and unweighted and undirected networks.

Other important issue is to know the most common used data set for identification of communities. A representative sample of these data set was reported in (Khan and Niazi, 2017). Thus, we only mention some examples: Southern women dataset (18 nodes), JAZZ musician network (198 nodes and 2742 links), Condense Matter collaboration network (Cond-Mat) (23133 nodes and 93497 links), ego-Facebook (4039 nodes and 88234 links), Lancichinetti–Fortunato–Radicchi, and DBLP (26950 nodes and 88742 links).

Editorial "Ediciones Futuro"   390
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Overview about the Community Detection Techniques

Knowledge of the community structure is critical to comprehend network functionality. Nevertheless, community detection is computationally intractable in large-scale networks. Thus, there are numerical algorithms and techniques to optimize this process. In this section we introduce some of the most relevant techniques. A resume of the most used techniques is presented in Figure 3.

Fig. 3 - A resume of the most used community detection techniques.

Editorial "Ediciones Futuro"   391
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Now, we present an overview of some of these techniques:

Traditional Community Detection Techniques

1. **Graph partitioning**: This technique divides the graph into g clusters of predefined size, such that the number of links in a cluster is denser than the number of edges between the clusters (Fortunato 2010).

2. **Hierarchical clustering**: These techniques are based on the vertex similarity measure. They do not need a predefined size and number of communities. They can be better represented by dendrograms (Papadopoulos, 2010).

3. **Partitional clustering**: It partitions a dataset into a predefined number of k non-overlapping clusters. The aim of this technique is to divide the data points into k clusters in order to optimize the cost function based on dissimilarity measure between nodes (Jin and Han, 2011; Fortunato, 2007, 2010; Dhumal and Kamde, 2015)

4. **Spectral clustering**: It includes all techniques which use eigenvectors of matrices to divide the set of data points based on the pairwise similarity between them (Fortunato 2010; Dhumal and Kamde 2015).

5. **Divisive algorithms**: It removes inter-cluster edges in a network based on low-similarity to separate communities from each other (Murata, 2010).

Modularity Optimization Based Community Detection Techniques

1. **Greedy techniques:** It is an agglomerative technique, where initially, each node belongs to a distinct module, then they are merged iteratively based on the modularity gain. It has a time complexity of $O(n)^3$ on sparse networks.

2. **Simulated Annealing:** It has used simulated annealing-based modularity optimization approach and it is a discrete stochastic approach used for the global optimization of the given objective function (Guimera and Amaral, 2005).

Editorial "Ediciones Futuro"   392
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

3. **Extremal Optimization**: This technique is focused on the optimisation of local variables. Duch et al. in (Duch and Arenas 2005) have used it for modularity optimization.

4. **Spectral Optimization**: It refers to the use of eigenvectors and eigenvalues of the modularity matrix for

5. modularity optimization (Fortunato 2010).

Overlapping community detection techniques

One of the most intricate problem of the community detection occurs when an overlapping system is present (it is when most of the nodes simultaneously belong to multiple communities). At present, it is a very common situation. The most known technique used for the identification of overlapping communities in the network is the clique percolation. The motif of this technique is that cliques are more probable to be formed from internal edges which are densely connected than from external edges which are sparsely connected. The communities are made up of k-cliques which refer to the complete subgraphs with k vertices. Communities are formed of k-cliques (which states the complete subgraphs with k vertices). The he k-clique community is the giant component formed of all the adjacent k-cliques which are connected as a k-clique series (Macropol and Singh 2010).

Dynamic Community Detection Algorithms

1. **Potts model:** It is based in a generalization of the Ising model in statistical physics. Here, the Potts spin variables can be mapped to the nodes of the graph having community structure. From interactions between neighboring spins, it is plausible that community structure may be identified from like-valued spin clusters of the system, as there will be more interactions in the community and fewer interactions outside the community.

2. **Random walk**: The walker starts to walk inside a community from a node and at each time step it moves to the neighboring node selected randomly and uniformly (Hughes, 1996).

3. **Synchronization**: It is an emerging phenomenon which has received interest from different fields. It occurs in interacting units and is persuasive in nature, technology and society. In a synchronized

Editorial "Ediciones Futuro"
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

393

state, the system units remain in same or alike states over time. Synchronization is also used in community detection in networks.

4. In large-scale complex networks, Leader-driven community detection algorithms (LdCD hereafter) is a new trend in devising algorithms. The main idea of this is to identify some particular nodes in the target network, called leader nodes, around which local communities can be computed. As it is based on local computations, they are particularly attractive to handle large-scale networks (Yakoubi and Kanawati, 2014).

Multi-node and multi-relationship hybrid networks community detection methods

At present, the research on community detection methods of multi-node and multi-relationship hybrid networks mainly focuses on the following two kinds: one is to extend the existing algorithm to deal with the hybrid networks directly, the other is to reduce the dimension of hybrid network to homogeneous network and then conduct community detection (Berlingerio et al., 2011; Suthers et al., 2013; Tang et al., 2010). Based on the above two ideas, the community detection methods of multi-node and multi-relationship hybrid networks mainly fall into the following five categories: probabilistic model-based methods, meta-path based methods, seed nodes based methods, extend modularity methods, and homogeneous methods of heterogeneous networks methods.

Probabilistic model-based methods

In the method based on the probabilistic model, some algorithms combine the ranking problem with the community detection problem, which is complementary to each other. RankClus (Sun et al., 2009) algorithm is the earliest proposed ranking clustering algorithm based on hybrid networks, but it is only applicable to two types of nodes. Sun and Han (2009) proposed a new algorithm, NetClus, based on RankClus algorithm, which uses Links between multiple nodes to generate high-quality network clusters. This algorithm has better clustering effects, but it is only applicable to the star network structure, and the representative objects in the data set need to be known in advance. In this regard, the RankClass (Ji et al., 2011) algorithm is improved to be suitable for hybrid networks of any network mode, and can make full use

Editorial "Ediciones Futuro"   394
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

of the label information of any data object. In order to reveal the evolution process of each type of node, Gupta et al. (2011) proposed the EnetClus algorithm, which performs an evolutionary clustering and uses a time smoothing method to display clusters that change over time. Qiu et al. (2015) proposed the OcdRank algorithm, which has low time complexity and supports incremental data updating.

Since the ranking-based method needs to set the number of communities in advance, it is instability. To this end, Sengupta and Chen (2015) proposed a hybrid network spectral clustering method for random block models, and applied the variational EM algorithm for a posteriori reasoning suitable for large networks, allowing different types of nodes to have multiple membership relationships, but this algorithm does not solve the problem of overlapping communities.

Therefore, it can be seen that the methods based on probabilistic models include two methods: ranking-based method and probabilistic statistical model. Although it is developed comprehensively, it is difficult to be applied to large-scale networks due to its high complexity in time and space. In addition, most communities need to be specified according to prior knowledge. When the network is large, it is difficult to predict accurately. And the effectiveness of community detection depends on the estimated number of communities in advance, which leads to instability of the results.

Meta-path based methods

Multiple types of nodes are connected by multiple links. The links linking different nodes contain different semantics. Such links form meta-paths. A meta-path is a series of object types that represent A semantic relation between two nodes (Li et al., 2018), it is an effective semantic capture tool that can capture rich semantic information in hybrid networks (Sun et al., 2011; Shi et al., 2017).

PathSim (Shi et al., 2017) is the earliest proposed algorithm based on meta-path. This algorithm is proposed for homogeneous networks and performs well in measuring the similarity between nodes of the same type. Li et al. (2018) pointed out that most of the meta-path-based hybrid networks community detection methods have two problems. One is that the similarity directly obtained from the meta-path is usually a deviation

Editorial "Ediciones Futuro"   395
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

measure, and the other is how to fuse the similarity of different meta-paths. Therefore, based on the standardization of PathSim to eliminate similarity deviations, they designed a flexible fusion mechanism to dynamically optimize the results, so that the results of community detection are better. Shi et al. (2012) proposed a similarity algorithm (HeteSim), that can measure the same or different types of nodes based on meta-paths. This algorithm calculates similarity through a two-way random walk, and performs better in query and clustering tasks than traditional algorithms, however, HeteSim is only applicable to a single meta-path environment, cannot capture multiple semantic information in a heterogeneous information network, and due to its high complexity, it's not suitable for large-scale networks. Subsequently, Meng et al. (2014) proposed a double random walk process based on a given meta-path and a reverse meta-path to calculate the similarity of two objects (AvgSim), which can be applied in large-scale networks, and has better clustering effect.

Different meta-paths contain different information, and the selection of different meta-paths will lead to different community detection results. How to determine the number of selected meta-paths or the optimal meta-paths among multiple meta-paths is a difficult problem. Sun et al. (2013) proposed the Pathselclus algorithm, which can assign different weights to different meta paths in hybrid networks. Shi et al. (2017) introduced the random walk method HRank, which is based on meta-paths to evaluate the importance of nodes and meta-paths, and the experimental results showed the unique advantages of meta-paths.

From the above analysis, it can be seen that the meta-path-based community detection methods in hybrid networks are mostly improved by the homogeneous network PathSim method. The meta-path-based method is relatively simple and easy to understand, but the similarity obtained is usually a deviation measure (Li et al., 2018). Moreover, different meta-paths contain different information, so it is still a difficult problem how to accurately calculate the similarity between nodes to show rich semantic correlation, and how to select the optimal meta-path among multiple meta-paths to obtain the optimal partition effect.

Editorial "Ediciones Futuro"   396
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Seed nodes based methods

Seed nodes-based methods have become an emerging trend in community detection methods (Hmimida and Kanawati, 2015). The basic idea of the method based on seed nodes is to identify certain specific nodes in the network, called seed nodes, and then build communities around these nodes (Kanawati, 2011; Papadopoulos et al., 2010; Shah and Zamn, 2010).

Yakoubi first proposed the seed node driven community detection algorithm (Licod). The basic idea is to select nodes with higher centrality than most direct neighbors as seed nodes, conduct local community calculation around these nodes, and then conduct community detection from local community sets (Yakoubi et al., 2014). However, this algorithm is only applicable to homogeneous networks. Hmimida and Kanawati (2015) extended the Licod algorithm to hybrid networks, called mux-Licod. This method considers the different types of relationships between nodes in different layers of hybrid networks, and the experimental results show that this method had good practicability.

Seed nodes-based method is a partial calculation method, which is easy to understand and suitable for processing large-scale networks and dynamic networks (Yakoubi et al., 2014). However, there is still no consensus on how to choose effective seed nodes efficiently. In addition, when merging non-seed node communities, there will be be problems of excessive merger of large communities and excessive number of small communities.

Extend modularity methods

Modularity was first used to evaluate the results of community detection. With the deepening of research, community detection algorithms based on modularity appeared (Newman and Girvan, 2004; Tang et al., 2009; Nicosia et al., 2009). Newman et al. first proposed the modularity optimization algorithm FN, which regards each node as a community and calculates the modularity value after the combination of two communities. The community combination methods of increasing the largest or decreasing the smallest in the modularity value is adopted, and the community detection is completed by iteration until the modularity

Editorial "Ediciones Futuro"   397
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

does not increase any more (Newman and Girvan, 2004). However, this algorithm is only applicable to single-node networks.

Guimerà et al. (2007) proposed an extended modularity algorithm suitable for bipartite networks, which can independently identify nodes with similar output connections and nodes with similar input connections, but it is not universal. Murata proposed a modularity algorithm suitable for k-core networks (Murata and Ikeya, 2010). This algorithm has the resolution limitation problem that general modularity algorithms have, and it is not suitable for hybrid networks with general morphology. Liu et al. (2014) proposed a composite modularity method. Its core idea is to decompose a heterogeneous network into multiple sub-networks, integrate the modularity in each sub-network, and optimize the composite modularity based on Louvain algorithm to achieve community detection. The algorithm does not require prior knowledge, and is suitable for large-scale networks and general morphology networks.

Obviously, the extended modularity methods evolves from the algorithm in the homogeneous network, and has high stability. However, the method has high time complexity and cannot avoid the limitation of maximum modularity resolution, which makes it impossible to detect the small communities in the large-scale network (Fortunato, 2009; Lancichinetti et al., 2011; Fortunato and Barthélemy, 2007).

Homogeneous methods of heterogeneous networks methods

Since the community detection method of homogeneous networks is relatively mature, the dimensionality of heterogeneous networks can be reduced into homogeneous networks, and then homogeneous network community detection method can be used for it. The dimension-reduction methods of heterogeneous networks mainly include non-negative matrix factorization (NMF) (Lee et al., 1999), topic model (Blei et al., 2003), principal component analysis (PCA) (Jolliffe, 2002), linear discriminant analysis (LDA) (Scholkopft and Mullert, 1999), among other.

Editorial "Ediciones Futuro"
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

398

Non-negative matrix factorization methods

Non-negative matrix decomposition methods can decompose any given non-negative matrix into two non-negative matrices (Liu et al., 2016), which are the basis matrix and the coefficient matrix respectively. The coefficient matrix is used to replace the original matrix to achieve dimensionality reduction. Tafavogh proposed a heterogeneous network community detection method based on matrix decomposition and semantic paths (Tafavogh, 2014). Zhang proposed a non-negative matrix three-factor decomposition method, HMFClus, which calculates similarity and integrates the information between objects of the same type into HMFClus. This method can simultaneously cluster all types of objects in a hybrid network (Zhang et al., 2016). Liu et al. proposed a penalized alternating factorization (PAF) algorithm to solve the corresponding optimization problems from the perspective of matrix decomposition for multi-layer attribute network. PAF algorithm not only has good effect on community detection, but also has strong applicability to network morphology (Liu et al., 2020).

Topic model methods

The introduction of topic model can excavate the hidden topic information in text information, so as to improve the effect of community detection. Mei et al. (2008) combined topic model with social network analysis, made full use of the advantages of statistical topic model and discrete regularization, improved topic model through regularization, and realized community detection.

Principal component analysis (PCA) methods and linear discriminant analysis (LDA) methods

Both of these two methods are linear dimensionality reduction methods, and use linear projection methods to map high-dimensional data to low-dimensional space. The difference between the two methods lies in that the former ensures that the data after dimensionality reduction retains more original information, while the latter makes the data after dimensionality reduction easier to distinguish. Existing studies only apply these two methods in single-node type networks (Lin et al., 2014; Li et al., 2016; Yuan et al., 2016) or bipartite networks (Liu and Chen, 2013).

It can be seen that although homogeneous methods of heterogeneous networks methods are easy to understand, the process of dimension reduction of heterogeneous networks into homogeneous networks is complicated, which is easy to cause information distortion, and the implementation complexity is too high. Because some methods need to estimate the number of communities, this method cannot meet the community detection requirements of large-scale networks.

To sum up, the existing community detection methods of hybrid networks with multi-node and multi-relationship are mostly based on probabilistic model and meta-paths, while the methods based on seed nodes, extended modularity and homogeneous methods of heterogeneous networks are still in the exploratory stage. In addition, there are still many problems to be solved in various methods, which indicates that there is still a lot of room for further research on community detection methods of multi-node and multi-relationship hybrid networks.

Commonly used evaluation indicators of community detection effect

There are many kinds of evaluation indicators for the effect of community detection, and different evaluation indicators are used for different community detection methods and different experimental needs. Three commonly used indicators are mainly introduced here: Normalized Mutual Information (NMI) (Lancichinetti et al., 2009), Adjusted Rand Index(ARI) (Santos and Embrechts, 2009), and Modularity Q (Newman and Girvan, 2004). Among them, NMI and ARI are evaluation indicators for known real community detection results, while Modularity Q is for unknown.

Normalized Mutual Information (NMI) is a similarity measurement method based on information theory and probability theory to evaluate community detection results. It is usually used to detect the difference between real detection results and actual detection results, which can intuitively show the quality of community detection results. The NMI calculation formula is as follows:

$$NMI = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} N_{ij} \times \log(\frac{N_{ij} \times N}{N_i \times N_j})}{\sum_{i=1}^{C_A} N_i \times \log(\frac{N_i}{N}) + \sum_{j=1}^{C_B} N_j \times \log(\frac{N_j}{N})}$$

Eq. 5

Editorial "Ediciones Futuro"   400
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

A and B are the result sets divided from the network, N is the number of all nodes, C_A and C_B represent the number of communities in A and B respectively, C_{ij} represents the number of nodes shared by the two communities, and C_i (C_j) is the sum of the elements in row i(j) in C. The value range of NMI is [0,1], and the larger the value is, the more accurate the result is.

In a broad sense, ARI measures the coincidence degree of two data distributions, that is, the similarity between real detection results and actual detection results is compared through whether each point pair is consistent under different community detection, which is defined as follows:

$$ARI = a_{11} - \frac{\frac{(a_{11} + a_{01})(a_{11} + a_{10})}{a_{00}}}{\frac{(a_{11} + a_{01}) + (a_{11} + a_{10})}{2} - \frac{(a_{11} + a_{01})(a_{11} + a_{10})}{a_{00}}}$$

Eq. 6

Among them, a_{11} represents the number of point pairs that belong to the same community in the real community and the actual community, a_{00} represents the number of point pairs that do not belong to the same community in the real community and the actual community, a_{10} represents the number of point pairs that belong to the same community in the real community but not belong to the same in the actual community, a_{01} represents the number of point pairs that do not belong to the same community in the real community but belong to the same in the actual community. The value range is [-1,1]. The larger the value is, the more consistent the actual result is with the real result. Compared with NMI, ARI has a higher degree of discrimination.

The modularity function Q was proposed by Newman and Girvan (2004). By optimizing the modularity Q, a better community division result can be obtained. The modularity Q can make the nodes within the community more closely connected, so it is an indicator to measure the strength of the community, which is defined as follows:

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \delta(C_i, C_j)$$

Eq. 7

Editorial "Ediciones Futuro"   401
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Among them, i and j are any two nodes, k_i and k_j are the degrees of nodes i and j respectively, and m is the total number of edges in the network. When two nodes are directly connected, $A_{ij}=1$, otherwise it is 0; C_i and C_j are the communities to which nodes i and j belong respectively. If the two nodes belong to the same community, it is =1, otherwise it is 0. The value range is [0,1]. The larger the Q value, the more stable the community detection structure is and the better the effect is.

Examples of papers on the community detection in hybrid network

In this section we present relevant studies about the community detection, particularly in academic networks. One of the relevant a complete publication we have found is the book edited by Wasserman S, Galaskiewicz J. (Wasserman and Galaskiewicz, 1994). The book shows how social network analysis has been used to advance substantive research in the social and behavioral sciences. The editor introduced researches in several topics from the social psychology and diffusion research, to anthropology and communications, and to politics and organization studies, span the range of social network analysis applications. From the theoretical and basic point of view we highlight the delightful paper of Fortunato (Fortunato, 2010). Here, the author presents a thorough exposition of the topic, from the definition of the main elements of the problem, to the presentation of most methods developed, with a special focus on techniques designed by statistical physicists (as the Pott technique), from the discussion of crucial issues like the significance of clustering and how methods should be tested and compared against each other, to the description of applications to real network. Other basic research was the paper of Newman (Newman et al, 2004). They proposed a set of algorithms for discovering community structure in networks—natural divisions of network nodes into densely connected subgroups. In (Van den Besselaar and Heimeriks, 2006) introduced a method to a method in which title words are used as indicators for the content of a research topic, and cited references are used as the context in which words get their meaning. In that paper, research topics were represented by sets of papers that are similar in terms of these word-reference combinations.

Editorial "Ediciones Futuro"   402
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Community detection in the more specific field of the professional network was studied by Suthers et al (Suthers et al., 2013). Also, interesting approximations on heterogeneous networks using a combination of techniques can be found in (Sun et al, 2009; Tang and Liu, 2010; Ji et al, 2011). Application of evolutionary clustering for analysis of bibliographic networks was reported by (Gupta et al., 2011). A relevant study using spectral clustering for heterogeneous networks was reported by Sengupta and Chen (Sengupta and Chen, 2015). On the other hand, Sun et al. (Sun et al, 2013) integrated meta-path selection with user-guided clustering to cluster objects in networks to generate clusters under the learned weights of meta-paths. Using Principal Component Analysis (Jolliffe, 2002) has been used to perform community detection in disease-gene network (Liu and Chen, 2013). A very interesting contribution in multiplex networks using a seed-centric approach was reported by (Hmimida, Kanawati, 2015)

More recently, the book "Heterogeneous Information Network Analysis and Applications" provides a comprehensive survey of current developments of heterogeneous information network. It also presents the newest research in applications of heterogeneous information networks to similarity search, ranking, clustering and recommendation (Shi and Yu, 2017). A very complete survey about this topic, including concept, techniques, algorithms and data was published by (Khan and Niazi, 2017). Also, the survey of heterogeneous information network analysis by (Chuang et al., 2017) introduce basic concepts of heterogeneous information network analysis, examine its developments on different data mining tasks, discuss some advanced topics, and point out some future research directions.

Recent developments

In the community detection field, new topics and techniques are emerging every day. Based on the extracted (Wang and Han, 2021) developed a novel attractive research community ranking (ARTRank) algorithm to rank the research communities. They are interested to answer the question: which is the most attractive research community in the academic social network? The main idea of the proposed algorithm lies in two measurements for each community: a positiveness score and a negativeness score, which measure the attractiveness of a community from the in-attention aspect and the out-attention aspect, respectively. New

Editorial "Ediciones Futuro"
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

403

metrics as the collaboration score, to quantify the collaboration between collaborators, are introduced (Dilmaghani et al., 2019). Extensions of the meta-path approach with has new similarity measures been recently used to study heterogeneous information networks (Wang et al, 2020). Also, a model based on the community detection method allows to detect topic-oriented densely-connected communities in which community members have active interactions among each other (Das and Anwar, 2019). On the other hand, (Liu et al., 2020) studied the integrative community detection problem of a multi-layer attributed network from the perspective of matrix factorization. Authors proposed a penalized alternative factorization (PAF) algorithm to resolve the corresponding optimization problem, followed by the convergence analysis of the PAF algorithm. A new multiplex network embedding model with high-order node dependence has been recently developed estimate the local structural complementarity of nodes as an embedding constraint of interlayer dependence (Ning et al, 2021). At present, meta-graph is one of the most powerful tools for similarity search on heterogeneous information networks (meta-graph is a composition of meta-paths that captures the complex structural information). In this sense, some interesting developments including embedded meta-paths information have recently reported (Sun, L. et al., 2018). The use of deep learning techniques has been in the focus in recent years. Particularly, unsupervised deep learning models (e.g. stack autoencoders) have been successfully proposed for the problem of community detection, which can extract network features and use them in splitting the network into communities (Pinaya et al., 2020; Al- Andoli et al., 2020)). A survey about this topic has been published recently (Al- Andoli, 2021)

Finally, a novel community detection method based on the similarity of communication behavior between IP nodes allowed the accurate detection and analysis of the community structure in large-scale IP networks could be highly beneficial for their optimization and security management (Zhang, et al, 2020).

Conclusions

We present a revision about the methods of community detection in hybrid network, and its application to the network analysis. Both the main techniques and the methods used for community detection were

Editorial "Ediciones Futuro"   404
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

presented. Also, a short description of the general features of this field was introduced. Traditional Community Detection Techniques, Modularity Optimization Based Community Detection Techniques, overlapping community detection techniques and Dynamic Community Detection Algorithms were described. Examples of application of these techniques were presented. Finally, relevant examples of the community detection in hybrid network applied to network analysis and recent developments in this field were discussed. Considering the revised literature, we propose a hybrid network community detection algorithm based on meta-path, seed nodes and extend modularity method to study the academic networks. Our idea is to construct a hybrid network community detection method with strong network applicability, low complexity, and the ability to combine network topology and text information simultaneously.

References

Al-Andoli, M., Ping Cheah, W., Chiang Tan, S. Deep learning-based community detection in complex networks with network partitioning and reduction of trainable parameters. Journal of Ambient Intelligence and Humanized Computing, 2021,12: 2527–2545.

Arenas, A, Duch, J, Fernández, A, Gómez, S. Size reduction of complex networks preserving modularity. New Journal of Physics, 2007, 9(6): 176-184.

Bedi, P., & Sharma, C. Community detection in social networks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2016, 6(3): 115-135.

Blei, D.M., NG, A.Y. and Jordan, M.I. Latent Dirichlet allocation. Journal of machine Learning research, 2003, 3(1): 993-1022.

Berlingerio, M., Coscia, M., Giannotti, F. Finding and characterizing communities in multidimensional networks. 2011 International conference on advance in social networks analysis and mining, 490-494.

Cai, Q., MA, L., Gong, M., Tian, D. A survey on network community detection based on evolutionary computation. International Journal of Bio-Inspired Computation, 2016, 8(2): 84-98.

Das S. and Anwar M. Discovering Topic-Oriented Highly Interactive Online Communities. Frontiers in Big Data. 2019, 2.

Editorial "Ediciones Futuro"   405
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Dhumal, A., & Kamde, P. Survey on Community Detection in Online Social Networks. International Journal of Computer Applications, 2015, 121(9).

Dilmaghani S, Brust Mr, Piyatumrong A, Danoy G, Bouvry P. Link Definition Ameliorating Community Detection in Collaboration Networks. Frontiers in Big Data. 2019, 2.

Divakarmurthy P, Menezes R. The effect of citations to collaboration networks, Studies in Computational Intelligence, 2013, 424: 177-185.

Drif, A., and Boukerram, A. Taxonomy and survey of community discovery methods in complex networks. International Journal of Computer Science and Engineering Survey, 2014, 5(4): 1-10.

Duch, J., & Arenas, A. Community detection in complex networks using extremal optimization. Physical Review E, 2005, 72(2): 027104.

Estrada, E. Rodriguez-Velázquez, J.A. Subgraph centrality and clustering in complex hyper-networks. Physica A: statistical mechanics and its applications, 2006, 364: 581-594.

Fortunato S, Barthélemy M. Resolution limit in community detection. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(1): 36-41.

Fortunato, S. Community detection in graphs. Physics reports, 2009, 486(3): 75-174.

Fortunato S. Community detection in graphs. Physics Reports, 2010, 486(3): 75-174.

Fortunato, S., and Hric, D. Community detection in networks: A user guide. Physics reports, 2016, 659: 1-44.

Gallos, L.K., Makse, H.A., Sigman, M. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(8): 2825-2830.

Gallos, L.K., Potiguar, F.Q., Andrade, J.S., MAKSE, H.A. IMDB network revisited: Unveiling fractal and modular properties from a typical small-world network. PLoS One, 2013, 8(6): 8.

Guimerà R, Marta S P, Amaral LA. Module identification in bipartite and directed networks. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 76(2): 036102.

Gupta M, Aggarwal C C, Han J, et al. Evolutionary Clustering and Analysis of Bibliographic Networks// International Conference on Advances in Social Networks Analysis and Mining, USA: IEEE, 2011.

Editorial "Ediciones Futuro"   406
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Hmimida M, Kanawati R. Community detection in multiplex networks: A seed-centric approach. Networks & Heterogeneous Media, 2015, 10(1): 71-85.

Hughes, B. D. Random walks and random environments, 1996.

Ji Ming, Han Jiawei, Danilevsky M. Ranking-based classification of heterogeneous information networks// Acm Sigkdd International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2011.

Jin, X., & Han, J. Partitional clustering. In Encyclopedia of Machine Learning (pp. 766-766): Springer, 2011.

Jolliffe I T. Principal component analysis. Berlin: Springer-Verlag, 2002.

Kanawati, R. Licod: leaders identification for community detection in complex networks. 2011 IEEE third international conference on privacy, security, risk and trust and 2011 ieee third international conference on social computing. (pp. 577-582). New Jersey: IEEE, 2011.

Khan, B.S., Niazi, M. A. Network Community Detection: A Review and Visual Survey. https://arxiv.org/abs/1708.00977, 2017.

Lancichinetti, A., Fortunato, S., Kertész, J. Detecting the overlapping and hierarchical community structure of complex networks. New Journal of Physics, 2009, 11(3).

Lancichinetti, A., Fortunato, S. Limits of modularity maximization in community detection. Phys Rev e Stat Nonlin Soft Matter Phys, 2011, 84(6): 066122.

Lee, D.D., Seung, H.S. Learning the parts of objects by non-negative matrix factorization. Nature, 1999, 401(6755): 788-791.

Li, Z., Zhang, S., Wang, R.-S., Zhang, X.-S., & chen, L. Quantitative function for community detection. Physical Review E, 2008, 77(3): 036109.

Li, L., Fan, K.F., Zhang, Z.Y., et al. Community detection algorithm based on local expansion K-means. Neural network world, 2016, 26(6): 589-605.

Li, J., Sun, P.Y., Mao, Q.R., et al. Path-Graph fusion-based community detection over heterogeneous information network. 2018 IEEE 20th international conference on high performance computing and communications. New Jersey: IEEE, 2018.

Editorial "Ediciones Futuro"   407
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Lin, L., Xia, Z.M., Li, S.H., et al. Detecting overlapping community structure via an improved spread algorithm based on pca. International conference on computer science and software engineering (CSSE), 115-121, 2014.

Liu W., Chen L. Community detection in disease-gene network based on principal component analysis. Tsinghua science and Technology, 2013, 18(5): 454-461.

Liu, T.L., Gong, M.M., Tao, D.C. Large-cone nonnegative matrix factorization. IEEE transactions on neural networks and learning systems, 2016, 28(9): 2129-2142.

Liu, X., LIU, W.C., Murata, T., et al. A framework for community detection in heterogeneous multi-relational networks. Advances in complex systems, 2014, 17(6).

Liu Jun, Wang J., Liu B. Community Detection of Multi-Layer Attributed Networks via Penalized Alternating Factorization. Mathematics, 2020, 8(2).

Ma, Y.-X., Xu, J.-Y., Peng, D.-C., Zhang, T., Jin, C.-Z., QU, H.-M., et al. A Visual Analysis Approach for Community Detection of Multi-Context Mobile Social Networks. Journal of Computer Science and Technology, 2013, 28(5): 797-809.

Macropol, K., and Singh, A. Scalable discovery of best clusters on large graphs. Proceedings of the VLDB Endowment, 2010, 3(1-2): 693-702.

Mei, Q.Z., Cai, D., Zhang, D., et al. Topic modeling with network regularization. In proceeding of the 17th international conference on world wide web. New York: ACM press, 2008.

Meng, X.F., Shi, C., LI, Y.T., et al. Relevance measure in large-scale heterogeneous networks. Asia-pacific web conference. Germany: Springer, 2014.

Murata T, Ikeya T. A new modularity for defecting one-to-many correspondence of communities in bipartite networks. Advances in Complex Systems, 2010, 13(1): 19-31.

Newman, M.E.J. The structure and function of complex networks. SIAM Review, 2003, 45(2): 167-256.

Newman, M.E.J, Girvan, M. Finding and Evaluating Community Structure in Networks. Physical Review E, 2004, 69(2): 026113.

Nicosia V, Mangioni G, Carchiolo V, et al. Extending the definition of modularity to directed graphs with overlapping communities. Journal of Statistical Mechanics: Theory and Experiment, 2009, 03.

Editorial "Ediciones Futuro"   408
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Ning N, Li Q, Zhao K, Wu B. Multiplex Network Embedding Model with High-Order Node Dependence, Complexity. 2021, Vol. 2021, Article ID 6644111.

Otte, E., Rousseau, R. Social Network Analysis. A powerful strategy, also for the information sciences. J. Inf. Sci, 2002, 28: 441–453.

Papadopoulos S, Kompatsiaris Y, Vakali A. A graph-based clustering scheme for identifying related tags in folksonomies. Data Warehousing and Knowledge Discovery, 2010: 65-76.

Plantié, M., and Crampes, M. Survey on social community detection. In Social media retrieval (pp. 65-85): Springer, 2013.

Pinaya Whl, Vieira S, Garcia-Dias R, Mechelli A. Autoencoders. Machine learning. Elsevier, Amsterdam, 2020.

Qiu, C.H., Chen, W., Wang, T.J., ET AL. Overlapping community detection in directed heterogeneous social network. Dong X., Yu X., Li J., Sun Y. (eds) Web-Age Information Management. WAIM 2015. Lecture Notes in Computer Science, vol 9098. Springer, 2015.

Ran, W., Shi, C., Yu, P.S., et al. Integrating clustering and ranking on hybrid heterogeneous information network. Pacific-Asia conference on knowledge discovery and data mining, AKDD 2013, Part I, LNAI 7818, Springer, Heidelberg, 2013, pp. 583–594.

Santos, J.M., Embrechts, M. On the use of the adjusted rand index as a metric for evaluating supervised classification. Springer, 2009.

Scholkopft, B., Mullert, K.R. Fisher discriminant analysis with kernels. Neural neworks for signal processing IX, Proceedings of the 1999 IEEE signal processing society workshop, 1999, 1, pp.41-48.

Sengupta S, Chen YM, Spectral clustering in heterogeneous networks. Statistica Sinica, 2015, 25(3), 1081-1106. Hindawi Complexity

Shah D, Zaman T. Community detection in networks: The leader-follower algorithm. Workshop on Networks Across Disciplines in Theory and Applications, Arxiv.2010.

Shen, H., Cheng, X., CAI, K., & HU, M.-B. Detect overlapping and hierarchical community structure in networks. Physica A: Statistical Mechanics and its Applications, 2019, 388(8): 17061712.

Shi, C., Kong, X.N., Yu, P.S., ET AL. Relevance search in heterogeneous networks. Proceedings of the 15th international conference on extending database technology. New York: ACM press, 2012.

Editorial "Ediciones Futuro"   409
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Shi Chuan, Li Yitong, Zhang Jiawei, et al. A survey of heterogeneous information network analysis. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(1): 17-37.

Shi Chuan, Yu P S. Heterogeneous Information Network Analysis and Applications. Springer International Publishing, 2017.

Song, C.M., Havlin, S., Makse, H.A. Self-similarity of complex networks. Nature, 2005, 433(7024): 392-395.

Su, X., Xue, S., Liu, F. et al. A Comprehensive Survey on Community Detection with Deep Learning. https://arxiv.org/pdf/2105.12584.pdf, 2021.

Sun, Y.Z., Han, J.W. Ranking-based clustering of heterogeneous information networks with star network schema. Acm sigkdd international conference on knowledge discovery & data mining. New York: ACM press, 2009.

Sun, Y.Z., Hany, J.W., Zhao, P.X., et al. RankClus: Integrating Clustering with Ranking for Heterogeneous Information Network Analysis, EDBT, 2009, 24-26.

Sun Y., Han Jiawei, Yan Xifeng, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. Proceedings of the VLDB Endowment, 2011, 4(11): 992-1003.

sun Y., norick B，Han Jiawei, et al. Pathselclus: integrating meta-path selection with user-guided object clustering in heterogeneous information networks. ACM transactions on knowledge discovery from data (TKDD), 2013, 7(3): 1-23.

Sun, L., He L., Huang, Z. Joint Embedding of Meta-Path and Meta-Graph for Heterogeneous Information Networks. https://arxiv.org/abs/1809.04110, 2018.

Suthers D, Fusco J, Schank P, ET AL. Discovery of community structures in a heterogeneous professional online network. 2013 46th Hawaii International Conference on System Sciences, 2013, 3262–3271.

Tafavogh, S. Community detection on heterogeneous networks by multiple semantic-path clustering. IEEE international conference on computational aspects of social networks. New Jersey: IEEE, 2014.

Tang, L., Wang, X.F., Liu, H. Uncovering groups via heterogeneous interaction analysis. Ninth ieee international conference on data mining. New Jersey: IEEE computer society, 2009.

Tang L., Liu H. Community Detection and Mining in Social Media, 2010, 1–137.

Editorial "Ediciones Futuro"   410
Universidad de las Ciencias Informáticas. La Habana, Cuba
rcci@uci.cu

Yakoubi Z, Kanawati R. Licod A Leader-driven algorithm for community detection in complex networks. Vietnam Journal of Computer Science 2014, 1(4): 241-256.

Yuan, P.Y., Wang, W., Song, M.Y. Detecting overlapping community structures with pca technology and member index. Proceedings of the 9th EAI international conference on mobile multimedia communications, ICST (institute for computer sciences, social-informatics and telecommunications engineering), 2016, 121-125.

Zhang Xianchao, Li Haixin, Liang Wenxin, et al. Multi-type Co-clustering of General Heterogeneous Information Networks via Nonnegative Matrix Tri-Factorization// 2016 IEEE 16th International Conference on Data Mining (ICDM). New Jersey: IEEE, 2016.

Zhang Z, Zhang Y, Zhou M, et al. Community detection based on similarities of communication behavior in IP networks. Ambient Intell Human Comput, 2020.

Van Den Besselaar P, Heimeriks G. Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. Scientometrics, 2006, 68(3): 377–393.

Wang Ran, Shi Chuan, Yu P S, et al. Integrating clustering and ranking on hybrid heterogeneous information network// Pacific-Asia conference on knowledge discovery and data mining. Germany: Springer, 2013: 583-594

Wang Y, Wang Z, Zhao Z, et al. Effective Similarity Search on Heterogeneous Networks: A Meta-path Free Approach. IEEE Transactions on Knowledge and Data Engineering., 2020.

Wang Y, Han X. Attractive community detection in academic social network. Journal of Computational Science, 2021, 51: 101331.

Wasserman S, Galaskiewicz J. Advances in social network analysis: Research in the social and behavioral sciences. USA: Sage, 1994.

Watanabe, A., Mizutaka, S., Yakubo, K. Fractal and Small-World Networks Formed by Self-Organized Critical Dynamics. Journal of the Physical Society of Japan, 2015, 84(11), 114003-114012.

Wolfe, W.J., Rothman, J., Macmillan, J., ET AL. Introduction to homogeneous networks. Proceedings of SPIE - The International Society for Optical Engineering, 1993.

Conflicto de interés

Los autores autorizan la distribución y uso de su artículo.

Contribuciones de los autores

1. Conceptualización: Jorge Gulín-González, Yunwei Chen

2. Curación de datos: Lu Jiang, Jorge Gulín-González

3. Análisis formal: Lu Jiang, Jorge Gulín-González, Edisel Navas Conyedo, Yunwei Chen

4. Adquisición de fondos: Lu Jiang, Yunwei Chen

5. Investigación: Lu Jiang, Jorge Gulín-González, Yunwei Chen

6. Metodología: Lu Jiang, Jorge Gulín-González, Edisel Navas Conyedo

7. Administración del proyecto: Yunwei Chen

8. Recursos: Jorge Gulín-González, Yunwei Chen

9. Software: No se desarrolló software propio

10. Supervisión: Jorge Gulín-González, Yunwei Chen

11. Validación: Jorge Gulín-González, Yunwei Chen

12. Visualización: Lu Jiang

13. Redacción-borrado original: Lu Jiang, Jorge Gulín-González

14. Redacción – revisión y edición: Lu Jiang, Jorge Gulín-González, Yunwei Chen

Financiación