MATEC Web of Conferences **232**, 01021 (2018)
EITCE 2018
https://doi.org/10.1051/matecconf/201823201021

Comparative Study Of Complex Network Community Structure Algorithms In network Pharmacology Analysis

Wangping Xiong[1], Xian Zhou[1], Bin Nie[1], Jianqiang Du[1,a]

[1]School of Computer, Jiang Xi University of Traditional Chinese Medicine, NanChang, JiangXi, China

Abstract. Community structure is an extremely important characteristic of complex networks composed of network pharmacology.The mining of network community structure is of great importance in many fields such as biology, computer science and sociology.In recent years, for different types of large-scale complex networks,researchers had proposed many algorithms for finding community structures.This paper reviewed some of the most representative algorithms in the field of network pharmacology, and focused on the analysis of the improved algorithms based on the modularity index and the new algorithms that could reflect the level and overlap of the community.Finally,a benchmark was established to measure the quality of the community classification algorithm.

1 Introduction

Network pharmacology is based on systems biology and network biology.It is possible to better understand the effects of cell and organ behavior on biological function at the molecular level of the system. This method can speed up the confirmation of drug targets and discover new drugs and targets.Thus,a combination of multi target drugs and drugs with efficacy and safety can be designed.

The content of network pharmacology covers a variety of aspects,including a variety of omics,system biology, multidirectional pharmacology, network biology,computational biology and biological analysis.Based on the drug-target -disease network,by analyzing the network topology,key protein function,drug disease network library and other existing information,Using professional network analysis software and algorithms, A systematic and holistic approach to reveal the mystery of disease-disease,disease phenotype-target protein, target protein-drug and drug-drug.Observe the intervention and influence of drugs on diseases from the network level.Projecting drug targets into complex disease networks (Figure 1)，revealing the mystery of complex drugs acting on the human body.

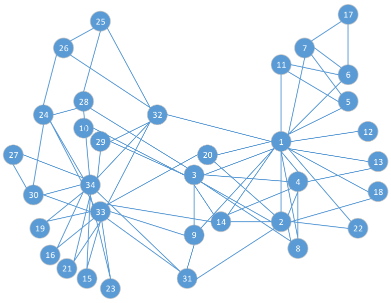

Figure 1. Component-Target network diagram.

Researchers have built a variety of network analysis models, and a certain network pharmacology analysis was made.These models include drug-target network and protein interaction network. The community structure in a network refers to a group of nodes with large similarities and different parts from other parts of the network.The feature is that the links between nodes are very close,and the links between communities are relatively sparse.

Finding community structure and analyzing it is a very important way to understand the structure of various network organizations in real life(Figure 2).It has been widely applied in biology, computer science and sociology，For example, the community structure in the social network enables people to clearly understand some characteristics or beliefs that they are different from other societies.In the biomolecular reaction network, Nodes aggregated together to form functional modules often assume specific roles or have specific functions.There are many algorithms to find the network community structure.Some classical algorithms such as Graph Segmentation classical problems and clustering analysis in sociology can be used for reference.

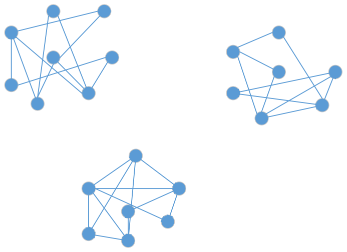

Figure 2. Community structure diagram.

[a] Corresponding author: xiaoxiongxwp@126.com

2 The application of network pharmacology

Yildirim et al.Used network analysis technology to integrate all kinds of data,and established the drug target interaction network for the 1178 drugs approved by FDA and the 394 target protein in Drugbank.For the generated network,the structural therapeutic chemistry classifier was used to highly connect most of the drugs according to the strong local clustering trend.Through the topology analysis of drug target interaction network,the relationship between the center of the drug action network and the nodes was revealed. There were not only a variety of drugs that can be used on a certain node, but a single drug could also be used in multiple targets.Analysis showed that each drug could be used on 1.8 targets on average.

The effectiveness and toxicity of drugs can be understood by analyzing the role of drugs in the "hub" and "specific" nodes of the network.Using network analysis,we can distinguish the direct and indirect effects of drugs on genes.On this drug target network,if some target proteins are highly connected, these proteins are likely to be basic proteins,and once interfered it will affect the safety of the network and produce toxicity.Understanding and understanding this drug target network can improve our research and development of complex diseases.

The development of bioinformatics technology provides the possibility for drug target prediction. This not only reduces the time and cost of drug research and development,It also reduces the difficulty of locating the target.Another important reason for drug development is that there is still a large number of potential drug target interactions that have not been discovered so far.Although a variety of biological experiments can be used to identify potential drug-target interactions,it is challenging to test these potential interactions with expensive costs.

At present,all possible small molecule compounds are estimated to contain more than 1060.This makes it difficult to understand the interaction between compound space and biological system.A large amount of evidence has proved that the drug target interaction mode is far more complex than the one to one interaction mode. The reasons are:(1) the drugs with different structures may exhibit similar biological activity and bind to the same target protein;(2) a drug may interact with multiple target proteins.Therefore,it is necessary to develop appropriate theoretical calculation methods to detect complex drug target interactions.

3 Hierarchical and overlapping algorithms

Hierarchy:Nodes in the network may have different levels of organizational structure,For example,Large associations may contain smaller associations.In this case,the network has a hierarchical community structure. Scholars have proposed a hierarchical clustering algorithm earlier.This clustering technology can reflect the multi-level structure of the graph. It has been applied in the fields of social network analysis, biological network and so on. Hierarchical clustering method may also get a hierarchical result when a network does not have a hierarchy.

In addition,Nodes in the community may not be properly divided.Moreover,some nodes or edges that have a key role in the module may also be lost. This situation is more obvious when dealing with large network data.To this end,they propose a top-down split algorithm that distinguishes networks that do not have community characteristics,or networks with community characteristics but not hierarchical,or networks with hierarchical societies.But because of its large amount of computation,the algorithm is not applicable to large medical data processing.

Reduplication: an important feature of community structure is its overlap.It means there are some "fence nodes" in the network.They are included in multiple societies at the same time.It belongs to the cross section of these societies. In real networks, the overlapping of community structure is very obvious.Most of the existing algorithms can only be standardized,That is, a node belongs only to a community. The nodes, however, usually belong to a number of societies. The real network is also composed of many overlapping and interrelated societies.Therefore, overlapping societies have become a hot topic in recent years.

4 Kernighan-Lin algorithm

The Kernighan-Lin algorithm is an exploratory optimization method. Based on the greedy algorithm,it divides the network into two dichotomy of known communities.The basic idea is to introduce a gain function Q for network partition.It is defined as the number of edges within two communities minus the number of edges between two communities.Then look for a division that makes the maximum Q value.

However, Kernighan-Lin algorithm must know the size of each community in advance. Otherwise,the correct results may not be obtained.The defect of Kernighan-Lin algorithm makes it difficult to apply in actual network analysis.

The whole algorithm can be described as follows: First,the nodes in the network are randomly divided into two communities with known size.On this basis,all possible node pairs are considered,The node pairs are derived from two communities. For each node,calculating the gain value ΔQ of the two nodes that may be obtained by exchanging the Q. At the same time,the Q value of the exchange after the exchange is recorded,It is stipulated that each node can only be exchanged once.Repeat the exchange process, Until all nodes in a society are exchanged once.

It should be noted that Q value does not necessarily increase monotonously during the process of node to swap. However,even if the exchange of a certain step decreases the Q value,It is still possible that a larger Q will appear in subsequent steps. When the exchange is finished,The maximum Q value recorded in the above exchange process is found.At this time the corresponding community structure is thought to be correct.

5 GN algorithm

In 2001, Girvan and Newnan proposed a community discovery algorithm based on edge betweenness.The algorithm is a splitting method.According to the betweenness,the edges of any society should not be deleted.The betweenness is defined as the number of shortest paths through each side in the network.That is,the sum of the times of all the shortest paths passing through the edge is the betweenness of the edge.

It provides an effective measure for distinguishing the inner side of a community and the side between the associations. According to the definition of a community in a complex network, The internal nodes of a society are closely related, but the connections between associations are rather loose. Therefore, the connection between the edges of the community has greater edge betweenness than the inner edge of the community.By gradually removing these higher numbers of edges, we can separate the communities they connect.

The basic steps of the GN algorithm (Figure 3) are as follows:

(1)The number of all sides in a network node is calculated.

(2)Find the highest number of edges and delete it from the network.

(3)Repeat execution steps(1),(2)until each node is a degraded community.

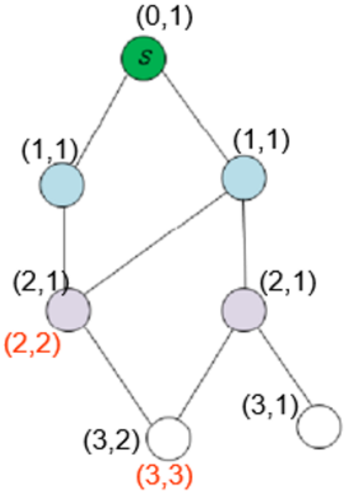

Figure 3. GN algorithm step diagram.

The GN algorithm makes up for the shortcomings of some traditional algorithms.But,without knowing the number of societies,The GN algorithm also does not know which step to decompose this decomposition.In order to solve this problem, Newman et al. Introduced a standard to measure the quality of network partition-Modularity.It is defined based on the same kind of matching. Consider some sort of division,It divides the network into k communities.A symmetric matrix of K*k dimension is defined. The element represents the ratio of the edges of the nodes that connect two different communities in all the edges.These two nodes are located in the i and j societies respectively.Notice that all the edges here are in the original network,All the edges here are in the original network,It is not necessary to consider whether to be removed from the community structure algorithm.

Therefore,the modularity measure is calculated by means of a complete network. After the improvement of Newman and others, The GN algorithm does not use redundant information to determine whether the community structure has practical significance,but directly analyzes the topology of the network. But the time consuming of the algorithm is large.So it's only used in medium-scale networks.In order to solve this problem, People have proposed many improved algorithms based on the GN algorithm.It mainly includes node set GN algorithm,self contained GN algorithm,extremum optimization algorithm and so on.

6 Conclusion

As a new discipline,network pharmacology has provided new ideas for many research fields.The study is an interdisciplinary field in pharmacology,biology, informatics and computer science.In addition to the limitations of various disciplines,it is also subject to technical limitations. For example,limited database information can not accurately reflect the state of the body,etc.To a certain extent,it has influenced its scientific nature.Now more and more researchers are paying attention to network pharmacology.The research of its related disciplines is also in depth.With the increasing number of data on diseases and drugs,The continuous improvement of computer technology and computer analysis software.It is believed that network pharmacology will be used more and more in the field of pharmaceutical research in the future.

Although there were many different community structure algorithms in the field of network pharmacology in recent years, there are still many problems. First of all, there was still a lack of a clear and consensus definition of community structure so far. Secondly, we need to define a set of reliable benchmark networks.It is used to test and compare the quality of various algorithms and the results of community partition. Now most algorithms are compared by modularity. The defect of the modularity index mentioned in this article,Therefore, whether or not to put forward a better evaluation index will be a very important research content in the future.

Acknowledgment

This research was supported by the National Natural Scie nce Foundation of China (No.61762051, No.61363042, N o.61562045)、the Jiangxi province science and technolog y plan projects (No.20151BBE50074, No.20171BBG701 08, No.20171ACE50021) 、the Project of Jiangxi Health Planning Commission (No.20165537, No.2017A284, No. 2017A313, No.20185516).

References

1. Anson,B.D.;Ma,J.Y.;He,J.Q.Identifying Cardiotoxic Compounds.Genetic Engineering & Biotechnology News 2009,29,34-35.
2. Muegge, I.Selection criteria for drug-like compounds. Medicinal Research Eeviews 2003,23,302-321.
3. Wang,J.M.;Hou,T.J.Drug and Drug Candidate Building Block Analysis.Journal of Chemical Information and Modeling 2010,50,55-67.
4. Walters,W.P.;Stahl,M.T.;Murcko,M.A.Virtual screening-an overview.Drug Discovery Today 1998,3,160-178.
5. Charifson,P.S.;Walters,W.P.Filtering database and chemical libraries.Molecular Diversity 2000,5,185-197.
6. Jia J, Zhu F, Ma XH, Cao ZW, Li YX, Chen YZ. Mechanisms of drug combinations: interaction and network perspectives[J]. Nat Rev Drug Discov, 2009, 8: 111-128.
7. J Zimmermann GR, Lehar J, Keith CT.Multi-target therapeutics:when the whole is greater than the sum of the parts[J]. Drug Discov Today, 2007, 12: 34-42.
8. The Indian Polycap Study (TIPS). Effects of a polypill (Polycap) on risk factors in middle-aged individuals without cardiovascular disease (TIPS): A phase II, double-blind, randomised trial. Lancet, 2009, 373: 1341-1351.