

Problem Domains in Complex Networks

Mini Singh Ahuja

Dept of Computer Science & Engineering /GNDU Regional Campus, Punjab, India

Abstract: Many real world Networks can be represented with the help of complex networks. Complex networks are special graphs with non trivial topological properties—features that do not occur in simple networks such as lattices or random graphs. Complex Network is a very Young and active area of scientific research which has its roots in graph theory. In this paper most important research domains related to complex networks are reviewed such as Community detection, Influence Maximization, network sampling etc. This paper will help in selecting the research problem in field of complex networks for the researchers who are new to this field.

Keywords: Complex Networks, Network models, link prediction

I. Introduction

These day real systems have grown tremendously in size . They contain million of actors and have different relationships. Complex networks are the powerful modelling tools which represent most real world systems. Complex network paradigm is one of the modelling tools which have spread through several application fields such as sociology, communication, computer science, biology, and physics and so on during last decades. Complex networks can be represented in the form of large graphs which have large no of nodes and different types of relationships with non trival properties. These nodes can be anything: a person, an organization, a computer or a biological cell. Nodes can have different size or attributes which represent a property of real system objects. These graphs can be directed, undirected or weighted. A complex network has its roots in graph theory. One of the main reasons why complex networks are so popular is their flexibility and generality for representing virtually any natural structure, including those which are undergoing dynamic changes of topology. Few examples of complex networks are Internet maps (IP, Routers [1], web graphs (hyperlinks between pages) [3], data exchange (emails) [2], social networks (facebook, twitter, scientist collaboration networks) biological networks (protein interaction, epidemic networks) etc. Complex networks have non trival properties so they cannot be explained by uniform, random, regular or complete models. This has led to the definition of set of statistics which have become fundamental properties of complex networks. These properties are now being used by many researchers for studying various phenomena's like spreading of information, protocol performance etc.

II. Network Models

Complex Network has been modeled by various types of graph models. Each graph model shows its specific features. Three most commonly used graph models are:

2.1 Random Graph of Erdős-Rényi

Two mathematicians, Erdos and Renyi in 1950 described a network with complex topology by random graph [4]. In a random graph, each vertex has random no. of vertices connected to it [5]. Their work had laid a foundation of the random network theory, and intense studies had been done in past 50-60 years and even today. Although, studies clearly shows that many real-life networks are neither regular nor completely random, but ER random graph model was the only sensible approach that made scientists think about complex networks for about half a century ago.

There are two approaches for generating random graph models. In first approach, random graph model is represented by $G_r(N,L)$ with N nodes and L links. Initially, N nodes are disconnected and there are no links between them. Then, links are added between nodes at random. Links which are to be added to the network are chosen from total of $N(N-1)/2$ potential links.

In second approach, random graph model is represented by $G_p(N,L)$ with N nodes and L links. The links are added to the network with probability p . The number of links in the graph is not definite and probably equal to $E[L] = pN(N-1)/2$. The average node degree of the graph is $(N-1)p$.

In a random network, graph may be either connected or disconnected. A random graph is connected with probability p . If we start decreasing this probability p , a stage comes after which the random graph becomes disconnected. The stage at which the graph becomes disconnected at a certain probability is called critical probability (p_c). In a disconnected random graph, the link density p is very low to form a connected

graph and it contains isolated clusters of nodes. Random networks have shortcomings also. Real networks cannot be demonstrated by Random Graphs.

2.2 Scale-free graph of Barabási- Albert

The scale-free algorithm by Barabási- Albert can represent many real networks like Internet, World Wide Web, citation network and some social networks as they exhibit all the properties of real networks. Scale-free graph [6] uses preferential attachment method for network growth. According to preferential attachment; better the node is known, more is the probability of a new node to attach to that node. Scale-free network is based on “Richer gets richer” theory. In scale-free networks, no. of nodes increases with the span of time. Hence, in scale-free networks hubs are formed as the network increases. Therefore, the average hop count of scale-free graphs is lower than other two graph models. The Barabási- Albert model is the famous model which generates scale-free networks by using preferential attachment and growth mechanisms. Scale-free network starts initially with m -nodes and each time a new node is added is to the present graph. The network becomes bigger and grows with the span of time. Scale-free networks always follow power-law degree distribution. Power-law describes that vast majority of nodes are not highly connected but a very few nodes are highly connected. It decays very slowly so we can find a hub in the system. Because of the presence of hubs in the system, scale-free networks are robust and fragile to attacks.

2.3 Small-world model of Watts and Strogatz

Small-world model was discovered by Duncan Watts and Steven Strogatz. In small-world model [7], only few links are used to reach from one node to other node. Social network and telecommunication networks follow the small-world property. In small-world network, most of the nodes are not neighbors of each other. In small-world phenomena two nodes are linked by a mutual node. Watts and Strogatz model is a very popular model which implements ring random graph with small-world property. Small-world model uses shortcuts to reach from one node to other node. Hence, average path length is relatively less in case of small-world model rather than in random graph model and scale-free model.

III. Challenges In Complex Networks

There are many issues in the field of Complex networks where still a lot of work needs to be done:

- Collection of data and network sampling
- Dynamics based scalability of complex networks
- Algorithmic Challenge
- Dealing with massive data
- Community Detection
- Link Prediction
- Influence Maximization

There are many important real-world applications that involve complex networks, but then also very little is known about how to address these issues.

3.1 Collection of data and network sampling

A major challenge in the study of complex networks is how to collect data for analysis. But before collecting network data, one must first decide what kinds of networks and what kinds of relations they will be studying. They have to decide between one-mode vs. two-mode networks and complete vs. ego networks. After deciding what kinds of networks and what kinds of relations we are going to study, researchers should decide about how to collect network data. We cannot directly collect data from these real world complex networks to study them. Collecting data of a network requires the use of a measurement procedure which is mostly partial or gives a biased view. So researches have to make an assumption that initially data is not fit to find the real properties but as the size of the data grows the properties become more and more stable. The research is going on this side of complex network too [2]. One can study how complex network properties evolve when the sample grows during the measurement. Researchers are trying to find the impact of the measured procedures on the obtained data to study the induced bias [9].

3.2 Dynamics based scalability of complex networks

Scalability is an important issue in almost all branches of Science and Engineering. In relation to complex networks, we can study about whether a dynamical phenomenon which occurs in networks of size say N , can the same phenomenon can occur in a networks which are much larger in size than N . Does the scalability so defined depend on the network topology? Finding solution to above questions can provide insights into fundamental issues such as the ubiquity of certain types of networks in nature with respect to specific

dynamical functions. A good understanding of the scalabilities of networks of different topologies can also be important for practical design of various technological networks. The issue of network scalability is an extensive research area which has not been considered yet.

3.3 Algorithmic challenge

Complex networks have both static and dynamic properties. So in order to understand the behavior of complex networks, we need to understand their static and dynamic properties. Till date many algorithms have come up to understand their properties. But the fundamental issue is the identification of the class of algorithms that can be used for the analysis of the network properties. Most of the algorithms so far designed have polynomial execution time. So approximation or heuristic algorithms are used for studying the properties. Heuristic algorithms work well with different input combinations but they don't guarantee a good solution. On the other hand approximation algorithms guarantee performance but with a factor of optimum value.

3.4 Dealing with massive data

Complex networks are massive in size. They contain million of nodes which are dynamically generated. So the modeling and analysis of such huge data is also one of the challenging problems. In order to study these massive graphs, we need to represent and process large scale graphs. It is very difficult to do exact in core computations on this massive data on a single work station due to limited physical memory. Parallel computing techniques can be applied to obtain the exact solution. The main characteristic of this massive data is its volume, size and high dimensionality which leads to many problems such as problems concerned with storing the data, difficulty in understanding the network structure and network dynamics.

3.5 Community Detection

Large real-world networks are generally characterized by heterogeneous structures which have some particular properties. The heterogeneous distribution of the links has led to community structure [10–12]. A community is a set of entities which are linked to all the other entities in the network. The entities in one community perform same function and share some common properties. A community structure reveals the internal organization of the nodes. Different communities combine to form a complex network. In other words, a community can be described as a collection of vertices within graph which are densely connected among themselves but are loosely connected to the rest of the graph [13]. Communities[8] can also be called as clusters or modules which share common properties. These communities have many features. They can have hierarchal or overlapping structure inside them. Moreover, these communities can be dynamic which change with time or can be multirelational (multiple relations). Many real networks such as social networks, biological networks exhibit community structure. Detecting clusters or communities in real-world network is a problem of considerable practical interest. The community detection problem has plenty of challenges as it is highly related to the problem of clustering large heterogeneous datasets. Till date many researchers have proposed number of algorithms, but all the community detection algorithms are different from each other and are not clearly defined [14,15]. So heterogeneity of different algorithms poses a challenge to community detection. Different networks (biological, social, etc.) have their own properties. This difference in properties as led to the unsolved question: which algorithm is suitable for which type of network? Moreover, these algorithms do not detect the same communities. So the problem is how to compare the performance of these algorithms. Actually, the researchers are interested in following information

- What type of information is used by the algorithm? A network can have different type of data: link attributes (weights, directions) node attributes, different types of links
- What type of community produced (partition, overlapped). The nature of communities the algorithm identifies.

Community evolution, overlapping communities, communities in directed networks, community characterization and interpretation are also some areas of research related to complex network communities.

3.6 Link prediction

In a given network, a major challenge is to infer the links which can occur in near future. This fundamental problem of predicting new links is called as the link prediction problem [16, 17]. Their occurrence is based on the edge, node and topological properties of the networks. Link prediction focuses on links between objects rather than objects themselves and due to this Link Prediction is interesting and different from traditional data mining area which focus on objects. Link prediction has many interesting applications in various domains such as in Recommender Systems, detecting selfish nodes in networks, completing or reconstructing of networks, molecular biology, criminal investigations, information retrieval, record linkage, marketing and bioinformatics. etc. The link prediction can be divided into three groups:

- Similarity measure based strategies
- Classification model based strategies which focus on exploiting either local or global patterns
- Probabilistic methods.

Many techniques have come up till date for link prediction. The link prediction in directed networks is attracting more and more researchers. Compared to predicting the existence of a link, determining its direction is more challenging. Predicting links in heterogeneous networks is also a new challenging problem related to link prediction.

3.7 Influence Maximization

Spreading process is one of the basic processes taking place in complex networks [18, 19]. It has been applied in many fields, such as information diffusion [20], disease propagation [21], and cascade failure. A whole network consisting of nodes and edges relies on a specific set of structural nodes. These structural nodes are very less in number as compared to the total number of nodes. These structural nodes if activated can spread the information to the whole network or if immunized can prevent a large scale epidemic. These structural nodes are also called influencers. The problem of identifying the minimal set of influential nodes in complex networks is crucial for developing efficient strategies to control the spreading process in complex networks. Many heuristic strategies have been used to identify these influential spreaders.

IV. Conclusion

Complex networks have changed our way of living and working. Complex networks have brought many opportunities and challenges such as large volume information and various security concerns. To meet such challenges, we need new mathematical tools and frameworks to model, analyze, understand, and predict the structures and functions of various complex networks.

References

- [1]. C. Gkantsidis, M. Mihail, and E. Zegura, "Spectral analysis of internet topologies," in *IEEE INFOCOM*, 2003.
- [2]. S. Voulgaris, A.-M. Kermarrec, L. Massoulié, and M. van Steen, "Exploiting semantic proximity in peer-to-peer content searching," in *IEEE FTDCS*, 2004
- [3]. P. Boldi and S. Vigna, "The webgraph framework i: compression techniques," in *WWW*, 2004
- [4]. Albert – Laszlo Barabasi and Reka Albert, "Emergence of Scaling in Random Networks", *Science*, vol. 286 (1999).
- [5]. Albert – Laszlo Barabasi and Reka Albert, "Topology of Evolving networks: Local Events and Universality", *The American Physical Society*, vol. 85, no. 24, 11 December (2000).
- [6]. Albert – Laszlo Barabasi et al. , "Deterministic scale-free networks", *Physica A, Science Direct*, pg no. 559-564 (2001).
- [7]. Xiao Fan Wang and Guanrong Chen, "Complex Networks: small-world, scale-free and beyond", *IEEE Circuits and Systems Magazine* (2003).
- [8]. Mini Singh ahuja and Jatinder singh , "Future prospects in community detection". Vol. 4, Issue 5, Oct 2014, 37-48.
- [9]. Z. Li and P. Mohapatra, "Impact of topology on overlay routing service," in *IEEE INFOCOM*, 2004.
- [10]. S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75–174, 2010.
- [11]. Richardt J and Bornholdt S, 2006, Statistical mechanics of community detection, *Physical Review E*74, 016110.
- [12]. Gulbahce, N. and S. Lehmann. The art of community detection [Journal Article]. *Bioessays*, 2008. 30(10): p. 934–938.
- [13]. M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, 2004.
- [14]. Andrea Lancichinetti, Santo Fortunato, 2009, Community detection algorithms: a comparative analysis, arXiv: 0908.1062v1 physics soc-ph.
- [15]. Hubert L and Arabie P 1985 Comparing partitions *Journal of Classification* 2 193–218.
- [16]. Lü L, Zhou T. Link prediction in complex networks: A survey [J]. *Physica A: Statistical Mechanics and its Applications*, 2011, 390(6): 115
- [17]. Ryan N. Lichtenwalter, et al. New Perspectives and Methods in Link Prediction, *KDD 2010*.0-1170
- [18]. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data[J]. *Nature*, 2009, 457(7232):
- [19]. Wang P, González MC, Hidalgo CA, Barabási A L. Understanding the spreading patterns of mobile phone viruses[J]. *Science*, 2009, 324(5930): 1071–1076
- [20]. Goldenberg J, Libai B, Muller E. Talk of the network: A complex systems look at the underlying process of word-of-mouth[J]. *Marketing Letters*, 2001, 12(3): 211–223. doi: 10.1023/A:1011122126881
- [21]. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals[M]. Princeton University Press, 2008.