

Complex Networks: A Review

Kriti Sharma

Student - M.tech (CSE)
CSE Dept., Guru Nanak Dev University,
RC Gurdaspur, India

Minni Ahuja

Astt. Prof. – M.tech (CSE)
CSE Dept., Guru Nanak Dev University,
RC Gurdaspur, India

ABSTRACT

In late 1950s, two mathematicians discovered a network with complex topology by random graph theory. Complex networks have received great attention in past few decades. Many studies have been done on complex networks and many are still in progress. Complex Networks are the networks which can be seen in real as well as in technological systems. They have nodes and these nodes are connected by various links. They are called complex networks because of the underlying complex architecture and complex topology. In this paper, our goal is to study the complex networks and various basic terms related to complex networks like mutual behavior between real-networks and complex networks, average path length, clustering coefficient, degree distribution. Designing and analyzing the behavior and dynamics of a complex networked system are also discussed. Hence, A complex networked system can help us in: understanding the efficiency of new security approaches for computer networks, improving the design of computer networks to make it more robust and resilience against errors and failures occurred in system, understanding how population will respond to introduction of new nodes in system, detecting subtle vulnerabilities, and also detecting catastrophic failures in power grid.

General Terms

Complex Networks
Understanding the complex networks, Study of complex networks

Keywords

Keywords Complex Networks, Understanding the complex networks, Complex networks in real life, Designing the networks.

1 INTRODUCTION

1. INTRODUCTION
Complex Networks are structures which are composed of interacting individual nodes abstracted from natural or technological systems. These nodes are connected by links or edges. Complex Networks have received great attention from scientific communities in past decades and are currently being studied across many fields of science [1-2].

Advancement in complex network have focused on network topological characteristics and network dynamics [2]. Complex

Networks define many structures or systems in nature. There are numerous examples of complex networks in real life or in scientific or computerized world; The Internet is a network of machines and routers, which are connected by physical links such as optical fibers. Machines and routers in Internet acts as nodes and transmission media acts as links or edges as in complex networks. Different websites in www constitutes a network. Neurons in brain constitute a network. People in an organization constitute a network. Food webs, metabolic pathways, diseases and even viruses can also be represented by complex networks [1].

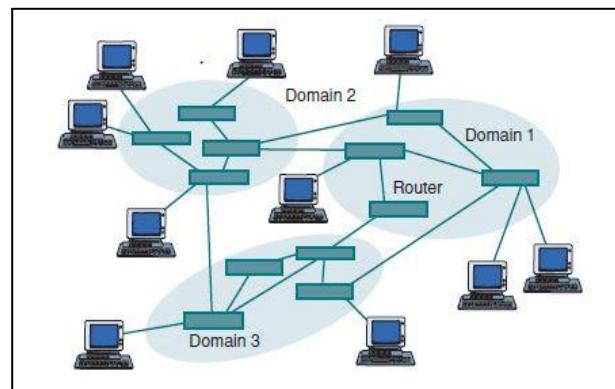


Fig. 1 – Network Structure of the Internet

However, the goal of studying complex networks is not only to examine the underlying structure of complex systems, but also to learn how we can control them more efficiently. A complex system is difficult due to the unknown architecture of the system [2]. According to control theory, a dynamic network is controllable if, for a selected choice of inputs, it can be derived from any initial state to any desired final state within a finite time [3].

In late 1950s, two mathematicians, Erdos and Renyl described a network with complex topology by random graph [1]. Complex network is the practical application of random graph. In a random

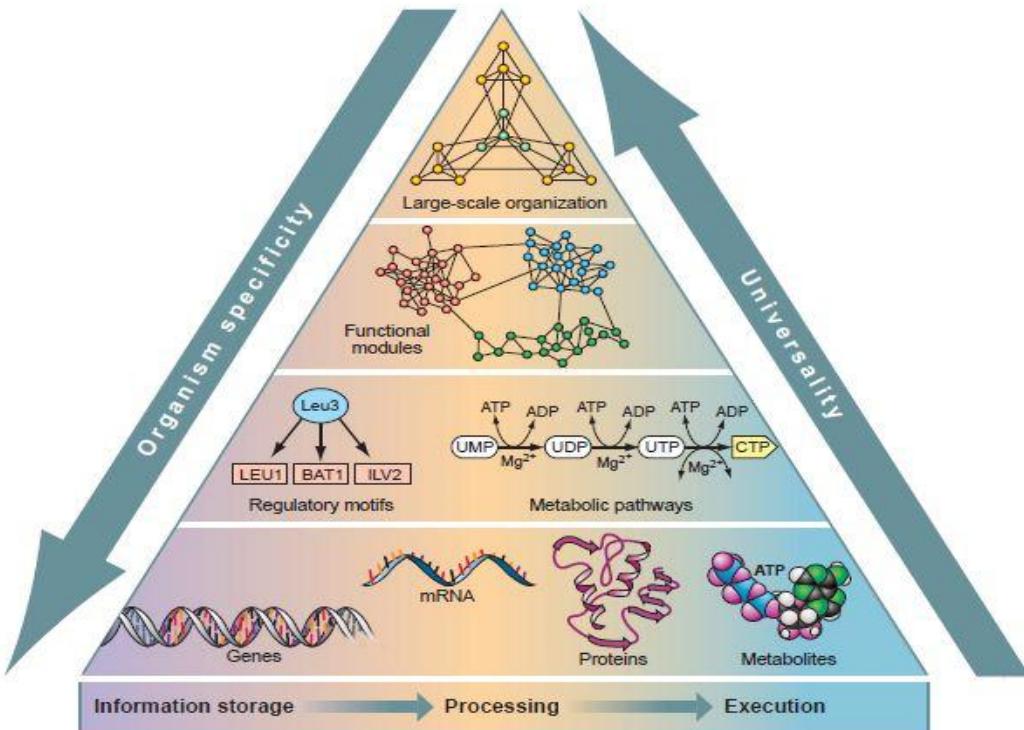


Fig. 2 - Representation of Cell as a Real-Life Complex Network

graph, each vertex has random no. of vertices connected to it [4]. Their work had made 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 thinking about complex networks for about half a century.

Small-world effect and scale-free feature are the recent discoveries in complex networks [1].

2. COMPLEX NETWORKS IN REAL LIFE

Cell is a network of genes and proteins which offers a feasible strategy for showing the complexity of living systems. Cells and microorganisms have an impressive feature, which helps them to live in an environment. Even if any kind of change occur in environment, they can recover themselves according to that change. In addition to this, they have an amazing ability to correct internal errors [5].

According to basic principle of molecular biology, DNA is the ultimate repository of biological complexity. It is generally seen and accepted that information storage, information processing and information execution of various cellular organizations lies at different levels. These different levels in a cell are known as genes, mRNA, proteins and metabolites.

However, the unconnected behavior of these organizational levels has recently known. Usually, long-term information is stored in genes and proteins are used for short-term information storage [5].

The bottom of the cell shows various molecular components of the cell which also acts as the functional organization in the cell. These molecular components are: genes, mRNA, proteins, metabolites. All these molecular components together form the level 1 of pyramid. At level 2, these molecular components form regulatory motifs or metabolic pathways, which in collaboration form functional modules at level 3. At level 4, a scale-free hierarchical architecture is generated by combining these functional modules [5]. This proves universality of complex networks from a cell to the World Wide Web. Hence, complex networks can be extracted from real as well as technological systems [1].

3. SOME BASIC CONCEPTS ABOUT COMPLEX NETWORK

Although many quantities and measures have been proposed about complex networks in the last decades, but three basic concepts defines the basic characteristics of complex networks. These basic concepts are: the average path length, clustering coefficient, and degree distribution. These concepts are used to satisfy the small-world effect and scale-free feature in complex networks. This section gives a brief review about these concepts:

Table 1– Table showing several networks; their size, clustering coefficient, average path length and degree exponent

Network	Size	Clustering Coefficient	Average Path Length	Degree Exponent
Internet, domain level	32711	0.24	3.56	2.1
Internet, router level	228298	0.03	9.51	2.1
WWW	153127	0.11	3.1	$\gamma_{in} = 2.1 \quad \gamma_{out} = 2.45$
E-mail	56969	0.03	4.95	1.81
Software	1376	0.06	6.39	2.5
Electronic Circuits	329	0.34	3.17	2.5
Food Web	154	0.15	3.40	1.13

3.1 Average Path Length

In a network, if two nodes labeled as i and j, then the distance d_{ij} between these nodes is defined as the number of edges along the shortest path connecting these two nodes. The diameter D of a network, is hence defined as the maximum distance among all the available distances between any pair of nodes in the network [6]. Therefore, the average path length L of a network is defined as mean distance between two nodes, averaged over all available paths between those nodes. Here, L defines the “Effective size of the network”. It has been discovered that average path length L of most complex networks whether they are natural or technological is small. This smallness defines small-world effect in complex networks [1].

3.2 Clustering Coefficient

In a complex network, it is quite possible that neighbor of a node's neighbor is also direct neighbor of a node, or in another way two neighbors of a node are also neighbors of each other. This property is known as the “clustering of a network”. Clustering Coefficient C is defined as the average pairs of neighbors of a node that are also neighbors of each other [1]. Suppose that a node i in a network has q_i edges, and these q_i edges connect this node to q_i other nodes. All these nodes are neighbors of node i. If every neighbor of node i is connected to every other neighbor of node i, then atmost $q_i(q_i-1)/2$ edges can exist among them. Thus, the clustering coefficient C_i of node i is defined as, $C_i = 2E_i / (q_i(q_i-1))$, where C_i is the ratio between E_i of edges that exist among these q_i nodes and the total number of $q_i(q_i-1)/2$ edges. Hence, the clustering coefficient C of whole network is the average of C_i over all i. Thus, $C <= 1$; and $C=1$ iff the network is globally connected or coupled [7,20].

3.3 Degree Distribution

The degree of each and every single node is the most important characteristic in a network. The degree q_i of a node i is defined as the total number of its connections to other nodes. Thus, more the degree of a node, more important is the node in a network. The average of k_i over all number of nodes is the average degree of the network. The average degree of a network is denoted by $\langle k \rangle$.

A distribution function $P(k)$ is defined as the probability that a node which is selected randomly has exactly k edges. A regular lattice has a simple degree, because each node is connected to equal no. of other nodes. So, a single sharp spike

is seen in the degree distribution graph. In a completely random graph, the degree distribution sequence obeys the familiar Poisson distribution; and the shape of the Poisson distribution fall off rapidly, after reaching the peak value $\langle k \rangle$. Because of this rapid fall, the probability of finding a node with k edges becomes negligibly small. In the past many years, many researchers have proved that the degree distribution of the large-scale real networks, deviates exceptionally from the Poisson distribution. The degree distribution of a number of networks can be defined as the power-law of the form $P(k) \sim k^{-\gamma}$, where γ is about 2 [11,12,16,18].

If the power-law distribution has the value of degree exponent under 3, then a scale-free network continue to remain connected indefinitely [11]. The power-law distribution falls off more slowly than Poisson distribution allowing for a few nodes of very high degree to exist. Because these power-laws are free of any scale; thus, such a network is called scale-free network [1].

Thus, many real-world or technological complex networks possess small-world and scale-free features in common.

4. DESIGNING OF COMPLEX NETWORKS

Many real-world systems such as food-web, metabolic systems and infrastructure systems such as road maps, circuit systems can be analyzed and designed as complex networks. Optimization flows and designing a network are the two classes of problems which are focused by network designers and optimization researchers. The problem of optimization flows in a network include finding the shortest path, minimum cost and maximum flow [8]. Optimization flow in a network means there should be minimum number of nodes included while transferring the data from sender to receiver in a network [10]. A network can be designed by using two approaches. These approaches are described below:

4.1 Exogenous Network

In an exogenous network, a network designer has direct control over all the nodes and how these nodes are connected to other nodes. In such kind of network, an external network designer controls the various activities of a network; Hence, it is called the exogenously designed network [8].

In an exogenously designed network, a node does not have any information about its neighbors. All the control information about a network, is kept by a network designer. A

network designer uses such kind of information while transferring data from one node to another.

Though an exogenous designed network is efficient and optimal while performing a task; but they are slow, which causes congestion in a network. Hence, Such kind of approach is now getting obsolete [8].

4.2 Endogenous Network

In an endogenous network, a network designer does not control the nodes. However, decisions are made at local-level by self-directed entities called nodes [8]. Local-level behavior of these nodes affect the global structure of a network and its properties which in turn affects the overall performance of the network. This approach of designing networks have gained interest in various research communities ranging from social to economical to technical networks in last few years.

Consider an example of the Internet as an endogenous network where routers acts as nodes and physical transmission media acts as communication links within these routers. These routers make optimal decisions about connecting to other routers while routing the data from one router to another. These local-level decisions made by routers affect the overall structure of the Internet and therefore affecting the performance of the Internet, which in turn providing more robustness and more resilience in the Internet.

This approach is a model of collaborative designing of a network where each node is capable of taking local-level decisions and hence contributing to the performance of the network [10].

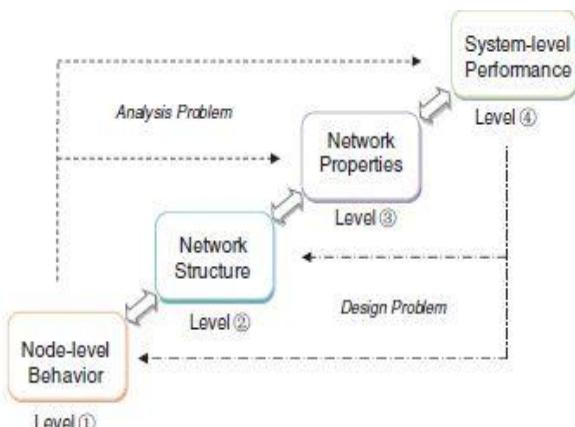


Fig. 3 – The relationship between node-level behavior, network structure, network properties and performance of a network

Choosing a design for a network is a bit difficult process from a set of available designs. We can also modify the structural features of a complex network in such a way that it exhibits the desired dynamics [9]. A design should be chosen on the basis of utilities provided by that particular design. So, we define a utility function for each design which represents the utilities for every point in a particular design. This utility function is compared from design to design; Then, a design is chosen on the basis of maximum optimal value of utility function [10].

Thus, the goal of design process is to try finding a design with the maximum optimal utility value, to obtain a ‘good enough’ design [10,14].

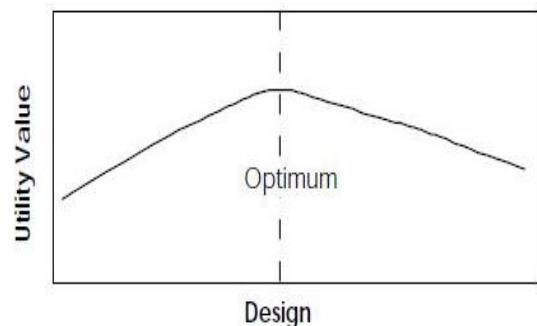


Fig. 4 – A simple utility function

5. ANALYSIS OF COMPLEX NETWORKED SYSTEMS

The complex networked systems should be analyzed to develop capabilities for understanding, designing and to control these networked systems. For analyzing these systems, we should need to examine four major elements: understanding network structure; understanding network dynamics; mathematical modeling and simulation; situational awareness, design and control of complex networks. All these four elements are described below:

5.1 Understanding Network Structure

The underlying topology of a network should be known beforehand to understand the network’s structure and its behavior. However, the network systems that are of interest have very large scale and are complex too. The network structure of these systems cannot be concluded directly. So, statistical data and machine learning methods are used to conclude the network structure of these systems.

The major problem is to conclude the network structure of these systems from the limited set of noisy observation. These observations are not directly connected, though they may be indirectly connected. In many cases, these systems are of very large scale and may be spatially distributed [13].

5.2 Understanding Network Dynamics

A networked system serves some dynamic processes that are the principle objects of interest. Network’s Structure is the underlying layer of Network’s dynamics and also provides some abstraction to the user. The objective of understanding the network dynamics is to understand the network behavior – under a steady state or some varying state. In the steady state, behavior possessed by a network remains constant. But, in varying state, network’s behavior also varies [13].

5.3 Mathematical Modeling and Simulation

If we have given mathematical or statistical model of a complex network, we can study the model’s structure and produce dynamic network’s behavior. Multiple levels of exactness and accuracy are needed while modeling and simulating a complex network. At the highest resolution level, simulations reproduce bits and instructions of the real complex dynamic networked system. At the lowest resolution level, sophisticated mathematical models are required to abstract the crucial information about a network. Mathematical modeling of a complex network and simulating its model’s structure helps in understanding the network’s behavior and its dynamics.

Performance optimization is done in large-scale simulations. Analysis of stability, errors and various convergence properties are also done in complex network simulations [13,17].

5.4 Situational Awareness, Design and Control of Complex Network

The fundamental goal of modeling and simulating a complex network is to develop methods for understanding network's behavior, controlling various events and remove anomalies from a complex networked system. The progress of a complex networked system depends on the study of these research areas. The final step in analysis procedure of a complex system focuses on awareness or activeness, designing and controllability of a complex networked system.

This step of analysis focuses on requirement of online learning in complex networks and the control methods are used to modify behavior of a network. Awareness in a complex network allows the network to change its dynamics according to the situation, various design and control methods are used to remove errors and optimization flow in a network [13,15,19].

6. CONCLUSION

Complex Networks have received great attention in last few decades. Complex networks are the networks which can be seen in real-life or in technological systems. Complex networks have nodes and these nodes are connected by many links. They are called complex networks because these networks have complex underlying architecture and complex topology which is difficult to understand. Complex Networks have been discovered by two mathematicians Erdos and Renyl in late 1950s. Life's complexity pyramid satisfies the mutual behavior of real-life networks with complex networks. Complex Networks share three basic concepts: average path length, clustering coefficient and degree distribution. Average path length defines the criteria for a network to have small-world property. Clustering coefficient defines the coupling or connectedness of network. It defines the average pairs of neighbors of a node that are also neighbors of each other. These pairs in a network defines the clusters in a network. Degree distribution defines the criteria for a network to have scale-free property. Power-law degree distribution sequence of a network allows few nodes to have high degree. These power-laws are free from any scale. Therefore, complex networks satisfy the scale-free property.

Design procedure of a network defines the criteria to design a network with maximum optimization achieved by a network. Network can be designed by two approaches: Exogenous and Endogenous approach. Utility function is used to calculate the maximum optimization achieved and maximum utilities provided by the system.

Analysis process studies the four main elements to develop capabilities, to understand the network's behavior and its dynamics. These four major elements are: understanding network structure, understanding network dynamics, understanding network dynamics, mathematical modeling and simulations, situational awareness, design and control of complex networks. A proper analysis should be done to control various events and remove anomalies from a complex networked system.

7. ACKNOWLEDGEMENTS

We want to thank Dr. Sandeep Sood for their guidance towards the work.

8. REFERENCES

- [1] Xiao Fan Wang and Guanrong Chen, "Complex Networks: small-world, scale-free and beyond", in IEEE Circuits and Systems Magazine, First Quarter, 2003.

- [2] Cun - Lai Pu, Wen-Jiang Pei, Andrew Michaelson, "Robustness analysis of network controllability", in Science Direct, Elsevier, pp. 4420-4425, 24 April 2012.
- [3] Yang - Yu Liu, Jean - Jacques Slotine and Albert - Laszlo Barabasi, "Controllability of Complex Networks", in Nature, vol. 473, pp. 167-173, 12 May, 2011.
- [4] http://en.wikipedia.org/wiki/Random_graph
- [5] Zoltan N. Oltvai and Albert - Laszlo Barabasi, "Life's complexity pyramid", in Science, vol. 298, pp. 763-764, 25 October, 2002.
- [6] Reka Albert, Hawoong Jeong and Albert - Laszlo Barabasi, "Diameter of the world-wide web", in Nature, vol.401, pp. 130-131, 9 September, 1999.
- [7] Matthieu Latapy and Clemence Magnien, "Measuring fundamental properties of real-world complex networks", unpublished.
- [8] Zhenghui Sha and Jitesh H. Panchal, "Towards the design of complex evolving networks with high robustness and resilience" in Science Direct, Elsevier, pp. 522-531, 2013.
- [9] Raoul - Martin Memmesheimer, Marc Timme, "Designing Complex Networks", in Science Direct, Elsevier, pp. 182-201, 3 November, 2006.
- [10] Mark Klein, Hiroki Sayama, Peyman Faratin, Yaneer Bar-Yam, "What complex systems research can teach us about collaborative system", unpublished.
- [11] Albert - Laszlo Barabasi, "Network Science", Phil. Transactions of Royal Society A, vol. 371, 18 February, 2013.
- [12] Albert - Laszlo Barabasi, "Scale-free Networks: A decade and beyond", in Science, vol. 325, pp. 412-413, 24 July, 2009.
- [13] James M. Brase, David L. Brown, "Modeling, Simulation and analysis of complex networked systems", in May 2009.
- [14] Ana Pinto, "Designing and the functioning of a productive learning network", in 9th International Conference on Productive Learning , 2014.
- [15] A. Clauset, H.G. Tanner, C.T. Abdallah, R.H. Byrne, "Controlling across complex networks: emerging links between network and control", submitted to Elsevier.
- [16] I. Farkas et. al., "Networks in life: scaling properties", in Science Direct, Elsevier, vol. 314, pp. 25-34, 2002.
- [17] Marta C. Gonzales and Albert - Laszlo Barabasi, "Complex Networks: from data to models", in Nature Physics, vol. 3, pp. 224-225, April 2007.
- [18] Reka Albert, Hawoong Jeong and Albert - Laszlo Barabasi, "Error and attack tolerance of complex networks", in nature, vol. 406, pp. 378-381, 27 July, 2000.
- [19] Tao Jia and Albert- Laszlo Barabasi, "Control capacity and a random sampling method in exploring controllability of complex networks", in Scientific Reports, 5 August, 2013.
- [20] Matthieu Latapy and Clemence Magnien, "Complex Network Measurements: Estimating the relevance of observed properties", unpublished.