ISSN 2348 – 9928
IJAICT Volume 5, Issue 5, May 2018 Doi:01.0401/ijaict.2015.11.10 Published on 05 (5) 2018

STRUCTURE AND TOPOLOGY OF MASSIVE GRAPHS

Ms. C. K. Suja
Assistant Professor,
Department of Computer Science,
De Paul Institute of Science &Technology
Angamaly, Kochi, Kerala.

Dr. C. V. R Hari Narayanan
Assistant Professor,
Department of Mathematics,
Govt.Arts College,
Paramakudi, Madurai, Tamilnadu.

Abstract— Structure of massive graphs are highly Complex in nature. A complex network with non-trivial topological features often occurs in modeling real systems. Networks like World-Wide web, internet, Science collaboration graph, Cellular networks etc which are evolutionary and dynamic in nature. A common property of such large networks is that the vertex connections follow a scale free power law distribution. The scale-free nature of the link distributions indicates that collective phenomena play a vital role in the development of such networks. Here we illustrate different modeling approaches of such random graphs and its properties to analyze networks.

Keywords— Scale Free Network, Random Graph, Small World Property, Preferential Attachment, Clustering Coefficient.

I. INTRODUCTION

Complex structures describe a wide variety of systems which are of high technological and intellectual importance. *For example, Internet* is a complex network of routers and computers linked by various physical or wireless links; communication and information diffusion on the social network, whose nodes are human beings and whose edges represent various social relationships; the World Wide Web is an enormous virtual network of Web pages connected by hyperlinks. These systems represent just a few of the many examples that have recently motivated the scientific community to investigate the mechanisms that determine the topology of complex networks. Traditionally the study of complex networks has been in the area of graph theory.

While graph theory initially focused on regular graphs, since the 1950s large scale networks with no apparent design principles have been described as random graphs, proposed as the simplest and most straightforward realization of a complex network. Random graphs were first studied by the Hungarian mathematicians Paul Erdo˝s and Alfre´d Re´nyi,[1].This model has guided researchers to think more about complex networks for decades since its introduction. The growing interest in complex systems has prompted many scientists to reconsider this modeling paradigm because there are some organizing principles embedded in the topology. But it has been found that these networks indeed deviate from a random graph and so developed tools and measures to analyze quantitatively the underlying principles.

II. NETWORKS CHARACTERISTICS

2.1 Degree Distribution

Not all nodes in a network have the same number of edges (same *node degree*). The spread in the node degrees is characterized by a distribution function $P(k)$, which gives the probability that a randomly selected node has exactly k edges. Since in a random graph the edges are placed randomly, the majority of nodes have approximately the same degree, close to the average degree $\langle k \rangle$ of the network. The degree distribution of a random graph is a Poisson distribution with a peak at $P(\langle k \rangle)$. One of the most interesting developments in understanding of complex networks was the discovery that for most large networks the degree distribution significantly deviates from a Poisson distribution. In particular, for a large number of networks, including the World Wide Web (Albert, Jeong,and Baraba´ si, 1999), the Internet (Faloutsos *et al.*, 1999),or metabolic networks (Jeong *et al.*, 2000), the degree distribution has a power-law tail ie $P(k) \sim k^{-\square}$, such networks are known as Scale free networks.

2.2 Clustering Coefficient

A common property of social networks is that cliques form, representing circles of friends or acquaintances in which every member knows every other member. This inherent tendency to cluster is quantified by the clustering coefficient C (Watts and Strogatz, 1998),a concept that has its roots in sociology, appearing under the name "fraction of transitive triples'' (Wassermann and Faust, 1994).Using the ratio between existing and possible relations, a clustering coefficient may be computed. If a node has z nearest neighbors, a maximum of z(z-1)/2 edges is possible between them. Watts and strogatz defined the clustering coefficient for node v as the ratio of the number of edges to the possible number of edges between the direct

ISSN 2348 – 9928
Doi:01.0401/ijaict.2015.11.10 Published on 05 (5) 2018

neighbors of node v(Watts and Strogatz,1998),

$$C^* = \frac{2i}{z(z-1)}$$

2.3 Average Path Length

Path length between two nodes of a network is defined as the number of edges between them. The minimal path length is the shortest path between two nodes. The average path length is the average of all minimal path length between all pairs of nodes in a network. This concept is known as small world, the fact that despite their often large size, in most networks there is a relatively short path between any two nodes. The distance between two nodes is defined as the number of edges along the shortest path connecting them. The most popular manifestation of small worlds is the ‘‘six degrees of separation’’ concept, uncovered by the social psychologist Stanley Mailgram (1967).The average distance between vertices scales logarithmically with the total number of nodes.

III. DIFFERENT MODELING PARADIGMS

The discovery of characteristics has initiated researchers to model such networks to study the theory of evolution and network dynamics. First, random graphs, which are variants of the Erdo˝ s-Re´nyi model, which serve as a benchmark for many modeling and empirical studies. Second, because of clustering property, a class of models, collectively called small-world models, has been proposed. These models interpolate between the highly clustered regular lattices and random graphs. Finally, the discovery of the power-law degree distribution has led to the construction of various scale-free models that, focusing on the network dynamics, aim to offer a universal theory of network evolution.

3.1 ER Models

The theory of random graphs was introduced by Paul Erdo˝s and Alfre´d Re´nyi (1959, 1960, 1961) after Erdo˝s discovered that probabilistic methods were often useful in tackling problems in graph theory. Mathematically a random graph is a pair of sets $G=\{P,E\}$, where P is a set of N nodes (or vertices or points) $P1$,$P2$,...,PN and E is a set of edges (or links or lines) that connect two elements of P. Networks with a complex topology and unknown organizing principles often appear random; thus random-graph theory is regularly used in the study of complex networks. Erdo˝s and Re´nyi in his classical paper defined a random graph as N labelled nodes connected by n edges, which are chosen randomly from the $N(N$-1)/2 possible edges (Erdo˝s and Re´nyi, 1959).In total there are n$CN(N$ -1)/2 graphs with N nodes and n edges, forming a probability space in which every realization is equiprobable. According to the Erdo˝s-Re´nyi model, we start with N nodes and connect every pair of nodes with probability p, creating a graph with approximately $pN(N$-1)/2 edges distributed randomly.ie the process of creating an ER-network depends on probability p.

3.2 Small World Networks

In most real world networks, there exist a relatively shortest path between pair of nodes. Such networks are called Small World Networks. They are similar to random graphs but having high cluster coefficient. The first successful attempt to generate graphs with high clustering coefficients and small l is that of Watts and Strogatz Model(1998).

3.3 Watts and Strogatz Model
Algorithm

1) Start with order: Start with a ring lattice with N nodes in which every node is connected to its first K neighbours (K/2 on either side). In order to have a sparse but connected network at all times, consider $N \gg K \gg \ln(N) \gg 1$.

2) Randomize: Randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded. This process introduces pNK/2 long-range edges which connect nodes that otherwise would be part of different neighbourhoods. By varying p one can closely monitor the transition between order and randomness .For small p, average path length l scales linearly with the system size, while for large p the scaling is logarithmic. In a regular lattice (p=0) the clustering coefficient does not depend on the size of the lattice but only on its topology.

As the edges of the network are randomized, the clustering coefficient remains close to $C(0)$ up to relatively large values of p. For p=0, each node has same degree and when p>0,changes occur in the network while maintaining the average degree equal to K. Since only a single end of every edge is rewired (pNK/2 edges in total), each node has atleast K/2 edges after the rewiring process. Consequently for K>2 there are no isolated nodes and the network is usually connected.

Corresponding Author: Ms. C. K. Suja, De Paul Insitute of Science and Technology, Kochi, Kerala, India.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2015.11.10 Published on 05 (5) 2018

3.4 Scale-Free Networks

The term scale free refers to the distribution principle of how many links per node. Empirical results showed that many real networks follow power law distribution ie $P(k) \sim k^{-\square}$.Also for such networks if degree distribution $P(k)$ has an exponential tail, the degree distribution significantly deviates from a Poisson distribution . The modeling of scale-free networks will put the emphasis on capturing the network dynamics. To understand the origin of this discrepancy, it has beene argued that there are two generic aspects of real networks that are not incorporated in these models [1]. First, both the previous models assume to start with a fixed number (N) of vertices, that are then randomly connected (ER model), or reconnected (WS model), without modifying N. In contrast, most real-world networks are open, i.e. they form by the continuous addition of new vertices to the system, thus the number of vertices, N, increases throughout the lifetime of the network. For example, the actor network grows by the addition of new actors to the system, the www grows exponentially in time by the addition of new web pages,the research literature constantly grows by the publication of new papers. Second, the random network models assume that the probability that two vertices are connected is random and uniform. In contrast, most real networks exhibit preferential attachment. For example, a new actor is casted most likely in a supporting role, with more established, well-known actors. These examples indicate that the probability with which a new vertex connects to the existing vertices is not uniform, but there is a higher probability to be linked to a vertex that already has a large number of connections. The scale-free model introduced by Barabasi-Albert,incorporating only these two ingredients, naturally leads to the observed scale invariant distribution [3][4].

3.5 The Baraba´si-Albert Model

In the previous models we start with a fixed number of nodes and edges are connected randomly. But in this model, start with a set of nodes and the network ***growth*** by addition of nodes and edges are connected by a mechanism called ***Preferential Attachment***.ie the likelihood of connecting to a node depends on the node's degree.

Algorithm:

(1) Growth: Starting with a small number (m_0) of nodes, at every time step, we add a new node with $m(<m_0)$ edges that link the new node to m different nodes already present in the system.

(2) Preferential attachment: When choosing the nodes to which the new node connects, we assume that the probability $\prod$ that a new node will be connected to node i depends on the degree ***ki*** of node i, such that

$$\prod(k_i) = \frac{k_i}{\sum_j k_j}$$

IV. APPLICATION IN REAL NETWORKS

4.1 World-Wide Web

Due to the increasing communication, the unregulated growth of the network leads to the generation complex web which can be modeled as a directed network in which, nodes are the documents (web pages) and the edges are the hyperlinks (URL's) that point from one document to another. The topology of this graph determines the web's connectivity and consequently how effectively we can locate information on it .But due to huge size and dynamics, it is impossible to find all the number of edges and vertices. But it has been found that the local connectivity can enabled us to explore and characterize its large-scale properties. Because of the power law tail the probability of finding documents with a large number of links is significant, as the network connectivity is dominated by highly connected web pages[2].Also for incoming links, it is non negligible. The results shows that the network shows the small world property and the average diameter of only 19 links. The scale-free nature of the link distributions indicates that collective phenomena play an unsuspected role in the development of the web[5][6].

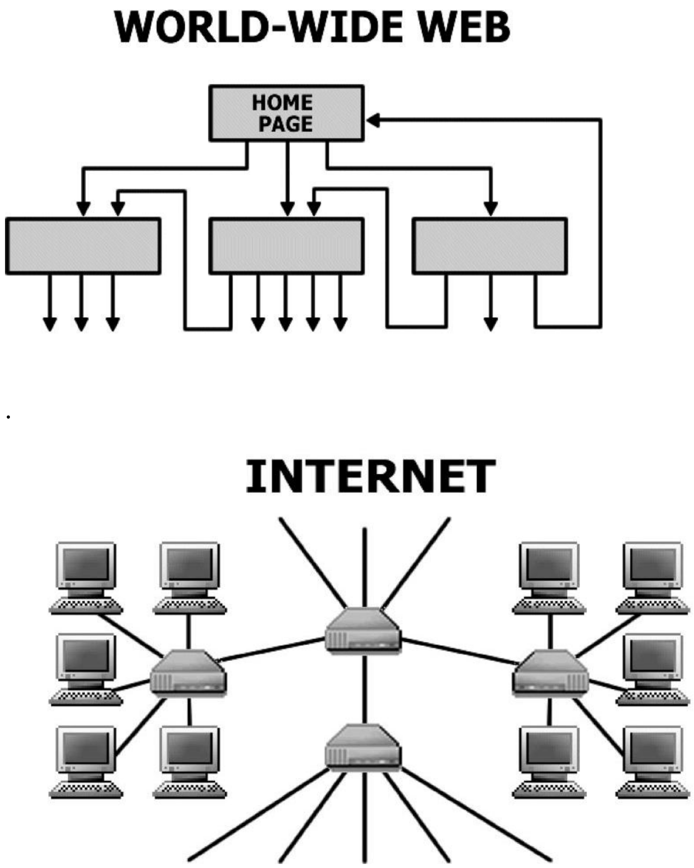
4.2 Internet

The Internet is a network of physical links between computers and other telecommunication devices domain, composed of hundreds of routers and computers, is represented by a single node, and an edge is drawn between two domains if there is at least one route that connects them. Earlier studies shows that the degree distribution follows a power law.

Also it has small path length and clustering coefficient. Its clustering coefficient ranged between 0.18 and 0.3, and the average path length of the Internet at the domain level ranged between 3.70 and 3.77 (Pastor-Satorras *et al.*, 2001; Yook *et al.* 2001a) and at the router level it was around 9 (Yook *et al.*, 2001a), indicating its small-world character[7][8][9].

Corresponding Author: Ms. C. K. Suja, De Paul Insitute of Science and Technology, Kochi, Kerala, India.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2015.11.10 Published on 05 (5) 2018



.

Fig 1 : Network structure of the World Wide Web and the Internet. Top the nodes of the World Wide Web are web documents, connected with directed hyperlinks (URL's).Bottom: on the Internet the nodes are the routers and Computers, and the edges are the wires and cables that physically connect them.

4.3 Proposed Work

The scale-free nature of the link distributions indicates that collective phenomena play an important role in the development of the web, forcing us to look beyond the traditional random graph models. A better understanding of web topology and its structural characteristics can be help to model such networks and develop search algorithms or designing strategies for making information widely accessible on the World-Wide Web.

V. CONCLUSION

Power laws are quite different from the bell-shaped distributions that characterize random networks. The two mechanisms-growth and preferential attachment helps to explain the existence of hubs when new nodes appear, they tend to connect to the more connected sites, and these popular locations thus acquire more links over time than their less connected neighbors. This "rich get richer" process will generally favor the early nodes, which are more likely to eventually become hubs. Determining whether a network is scale-free is important in understanding the system's behavior, and analyzing its significant parameters are of paramount importance.

References

[1] Re´ka Albert* and Albert-La´ szlo´ Baraba´ si, Statistical mechanics of complex networks, Reviews of modern physics, volume 74, January 2002

[2] Albert, R., H. Jeong, and A.-L. Baraba´ si, 1999, Internet:Diameter of the world-wide web, Nature (London), **401**, 130-131.

[3] Erdo˝s, P., and A. Re´nyi,1960, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17.

[4] Wasserman, S., and K. Faust, 1994, *Social Network Analysis: Methods and Applications* (Cambridge University, Cambridge

[5] Watts, D. J., and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393(1998)**,440-442.

[6] Oliver Hein.Michael Schwind,Wolfgang Kong,2006"Impact of Fat Tailed Degree distribution on Diffusion and communication Processes",267-275

[7] Albert-L_aszl_o Barab_asi_, R_eka Albert, Hawoong Jeong, "Mean-field theory for scale-free random networks" Department of Physics, University of Notre-Dame, Notre-Dame, IN 46556, USA(1999)

[8] Yook, S., H. Jeong, A.-L. Baraba´ si, and Y. Tu, 2001, Phys. Rev.Lett. **86**, 5835.

[9] Pastor-Satorras, R., and A. Vespignani, 2001a, Phys. Rev. Lett.**86**, 3200

Corresponding Author: Ms. C. K. Suja, De Paul Insitute of Science and Technology, Kochi, Kerala, India.