Special Proceeding of the 21st Annual Conference of SSCA held at SV Agricultural College (ANGRAU), Tirupati, during January 29-31, 2019; pp 67-71

A Treatise on Networks

Swati Goswami
Advanced Computing and Microelectronics Unit, Indian Statistical Institute, Kolkata 700108, India

Received: 22 June 2019; Revised: 25 June 2019; Accepted:

Abstract

Network based modelling of various physical phenomena has attracted the attention of researchers of various scientific disciplines of today. This has led to a tremendous development of the area, especially during the last two decades, in both theory and applications. This article, in a short span, introduces the fundamental concepts of network based modelling. It highlights the prominent properties of real life interaction networks and points out their differences with random networks. Most real life large networks are sparse; the article discusses the property of sparsity and elaborates on measuring sparsity of network graphs. A newly introduced measure of sparsity of the degrees of nodes of a network, called sparsity index, has been explained along with some of its uses.

Key words: Interaction networks, Network sparsity, Sparsity index, Network properties.

1 Introduction

A wide variety of physical phenomenon can be modelled using network graphs. Interactions within a group of entities are frequently studied using network graphs. Examples of such interactions are varied and many. A few of them are: voice calls or SMS messages exchanged within a certain community of people, protein-protein interactions in biological networks, co-authorship relationship within a community of researchers, interactions among brain regions and so on. The entities themselves constitute the nodes of the graph, the interactions among the entities are represented by the edges of the graph and intensity or strength of interactions is associated with edge-weights. If the direction of the relationship or interaction is of any significance for the model, then a directed graph is constructed. For very large real life graphs, the largest connected component is often used for analysis, ignoring the small components, for tractability. Real life interaction graph data can be multi-relational with a large amount of associated auxiliary data and the graph itself can be dynamic (i.e., the graph may evolve on a temporal scale). Therefore, the analysis of such networks may turn out to be quite complex.

Interaction networks can be explored from one or more of the following viewpoints: i) network characterization – structural and statistical properties of networks are studied to formulate the fundamental principles which govern and account for the characteristics of the network ii) community detection – detecting the natural grouping of nodes within the network iii) dynamic behaviour analysis – discovery of principles for dynamic behaviour of networks or network communities over time iv) network based prediction (*e.g.*, predicting the missing links in protein-protein interaction networks [Yu 2006], predicting the spread of diseases in complex networks [Chen 2014]).

In subsequent sections, this article provides a glimpse of the prominent network properties which characterize a good majority of real life interaction networks. It addresses

Corresponding Author: Swati Goswami
Email:swati.goswami2000@gmail.com

the particular property of network sparsity and its measurement. It concludes with mentioning a few directions in which future research may proceed.

2 Interaction Networks: Distinguishing Properties

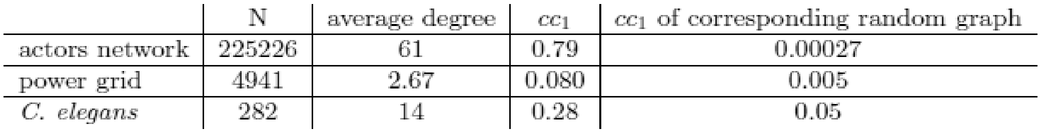
Real life interaction networks are typically large, in terms of the number of nodes, and highly sparse. They tend to display a common set of statistical properties [Newman 2003] which distinguish them from regular networks (*e.g.*, lattices) and random networks. We review three of these important properties in this section.

- **Small World effect**: Watts and Strogatz introduced this property in 1998 as they observed: "we find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. We call them 'small-world' networks". The small world networks lie somewhere in between regular and random networks on the scale of increasing randomness and are characterized by small average-shortest-path-length (or mean geodesic) over the network. The concept of "six degrees of separation" stems out of the famous small world experiment by Milgram in the 1960's. It is an experiment of reaching a letter to a target individual, unknown to the person from whom the letter originates. Most of the letters were lost, but about a quarter reached the target and, in the process, changed hands only about six times on an average. The mean geodesic distance between pairs of nodes in a network is given by $\boldsymbol{\frac{1}{\frac{1}{2}n(n+1)}\sum_{i\geq j} d_{ij}}$, where d_{ij} is the shortest path length between nodes i and j.

- **Degree Distribution:** The number of edges incident on a node is referred to as its *degree*. Degree provides a measure of how connected the particular node is to the rest of the graph. If $p_k$denotes the fraction of nodes in a network with degree $k,\ k = 0,1,2, \dots maximum\_degree$, then p_k also denotes the probability that a node chosen uniformly at random has degree k. If we draw a histogram of the degrees of nodes of the network, we get its degree distribution. Real world networks show a marked difference in their degree distributions from random networks. In a random network (as studied by Erdos-Renyi 1959), each edge is present or not with equal probability, and hence the degree distribution is binomial or Poisson in the limit of large graph size. Far from being Poisson distribution, the degree distributions of most real life large networks are highly right-skewed, i.e., with a long right tail. Moreover, many of them tend to display power-law, with $p_k \sim k^{-\beta}, \beta > 1$. The networks with power law degree distributions are often referred to as the scale-free networks.
- **Clustering Coefficient**: For the present context, "clustering" means transitivity. In social terms, transitivity indicates the likelihood that in the network, friends of a friend are friends themselves, i.e., if A and B are friends of C then A and B are also friends. In topological terms, transitivity denotes the tendency of the network to form triangles among sets of connected triplets of nodes. The clustering coefficient, a measure of transitivity, defined at the local and the global level for the network:

  $C_i = \frac{number\ of\ triangles\ connected\ to\ vertex\ i}{number\ of\ triplets\ centred\ on\ vertex\ i}$ and $C = \frac{1}{n}\sum_i C_i$, where C_i is the local clustering coefficient of the node i and $C$is the clustering coefficient of the network. By "triplets centred on a vertex", it means a single vertex with edges running to an unordered pair of other vertices. C_i is taken as zero for nodes for which degrees are zero or one (i.e., for which both the numerator and the denominator are zeroes in the formula). Clustering coefficient lies between 0 and 1. There is an alternative definition of clustering coefficient in the literature which we keep outside of the present discussion. Random networks and real life interaction networks tend to differ in terms of this property also; the real life interaction networks display a much heightened probability of forming closed triangles.

Let us reproduce an empirical example of small world networks that Watts and Strogatz have used in their paper [Watts 1998], which lists the clustering coefficient C ($cc1$ in Table 1) for three real networks compared to random networks with the same number of nodes (N) and same average number of edges per node.



	N	average degree	cc_1	cc_1 of corresponding random graph
actors network	225226	61	0.79	0.00027
power grid	4941	2.67	0.080	0.005
C. elegans	282	14	0.28	0.05

Table 1: Comparing observed networks against "corresponding" random graphs.

3 Sparsity of Interaction Networks

Sparsity is a fundamental property of a network. For a network graph, *sparsity* is an indication of the extent of the graph's deviation from a fully connected graph. Sparsity lies at the opposite end of the density spectrum of a graph, i.e., the less dense the graph is, the more is its sparsity. It is commonly measured by *edge density*, which is the ratio of the cardinality of the edge-set to the cardinality of the edge-set of the corresponding fully connected graph. It is computed as: $|E|/\binom{|V|}{2}$. Considering the adjacency matrix representation of the graph, the l-zero norm, or, even the proportion of zero elements compared to non-zero elements of the adjacency matrix may serve as sparsity measures. Edge density being just a simple ratio, it has got certain limitations as a measure of sparsity. For example, for weighted graphs, it completely ignores the edge weights. A new measure of sparsity of network graphs has been introduced recently [Goswami 2018], named *sparsity index*, which is based on the number of nodes of the graph, its degree sequence and a constant factor at least as large as the total degree of all nodes of the graph. The measure has been formulated by using the definition of Gini index [Gini 1912]. Sparsity index indicates the amount of disparity in distribution of degrees of the nodes of the network. In comparison with edge density, sparsity index is a more fine-grained measure, as it takes into account a lot of other information about the graph. Moreover, Goswami et al. have shown that the two measures display the same trend, i.e., if edge density of a graph increases by adding more edges to the graph, its sparsity index decreases. However, if there are two graphs with the same number of nodes, one with a higher edge density than the other, it is not necessarily true that the graph with the higher edge density will have a lower sparsity index than the other.

We consider a graph $G = (V, E)$, with $|V| = n$ and $|E| = m$, as a representation of interactions among n individuals. Let A denote the adjacency matrix of the graph, such that $A = [a_{ij}]_{n \times n}$with a_{ii} equals to zero $\forall i = 1,2 \dots, n$ and $a_{ij} \in \{0,1\}$, $i, j = 1,2 \dots, n$. Clearly, the graph $G$is unweighted (binary a_{ij} values indicate only the presence of an edge between the corresponding nodes). Let the degree of node i be given by a_i, where $a_i = \sum_{j=1}^{n} a_{ij}$ $\forall i = 1,2 \dots, n$ and the total degrees of all nodes in the graph be denoted by T, where $T = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$ .Let the degree sequence of G be represented by $\underline{a} = [a_1, a_2, \dots, a_n]$.Let the elements of $\underline{a}$ be arranged in an ascending order and let b_i denote the i-th ordered statistic in the sequence of values $a_1, a_2, \dots, a_n$, i.e., $\underline{b} = [b_1, b_2, \dots, b_n]$ such that $b_1 \leq b_2 \leq \cdots \leq b_n$. The vector $\underline{b}$ is called the ordered degree sequence of the graph G. It follows that $\sum_{i=1}^{n} b_i = T$.Let T_1be a quantity at least as large as T.The sparsity index of the graph G is given by:$\boldsymbol{SI(G) = 1 - 2}[\sum_{i=1}^{n} \frac{b_i}{T_1} \left(\frac{n-i+\frac{1}{2}}{n} \right)]$.The quantity T_1 is chosen depending on the quantity with respect to which we are interested in measuring the sparsity. For $T_1 = T$, $SI(G)$ becomes exactly equal to the Gini Index of the degree distribution of the network. For a simple graph, undirected and unweighted, the most plausible choice of T_1 is $n(n-1)$, to calculate sparsity with respect to the potential total degrees in the graph rather than the actual total degrees.

It may be noted that sparsity index is a summary measure and lies between 0 and 1. The sparsity index of a regular cycle, for example, with n nodes is $\frac{n-3}{n-1}$, taking $T_1 = n(n-1)$, whereas, Gini index of a regular cycle is 0.For weighted networks, if the edge-weights are integers then the network can be expressed as a multigraph [Newman 2004], i.e., a pair of nodes would have those many edges as the weight of the edge between them. In other words, the weights add to the degrees of the nodes and hence a sparsity index can be calculated for a weighted graph.

Gini index and sparsity index together reveal characteristics of a network without even doing further analysis. For example, for networks with a Gini index value close to zero, it is highly unlikely to find good clusters, as the nodes of the network would have more or less the same degrees. On the other hand, a high value of Gini index may indicate the presence of a few very well connected individuals (or, influential individuals in social networks) in the network.

4 Conclusion

This article has been an attempt to sketch interaction networks on a small canvas. The importance of interaction networks to the research community is immense; almost in every scientific discipline of today there is some application of such networks. The properties, because of which the interaction networks are different from random networks, or other special graphs, have been highlighted. However, we have stopped short of discussing centrality measures and their various applications. Network community detection is a whole area in itself, which we have not discussed here. Newer methods of network community detection are coming up even today from different fields, addressing different types of networks. More than static networks, dynamic networks or temporal networks are able to represent the evolving nature of the relationships among entities in a much better way and hence their analysis is gaining momentum. It may be of interest to study how the various network measures change over a period of time in a dynamic network. Also, inter-relationships among different network measures is another area which merits further scrutiny.

References

Chen, D.B., Xiao, R. and Zeng, A. (2014). Predicting the evolution of spreading on complex networks. *Scientific Reports*, **4,** p.6108.

Erdös, P. and Rényi, A. (1959). On random graphs. I. *Publicationes Mathematicae (Debrecen)*, **6**, 290-297.

Gini, C. (1912). Italian: Variabilità e mutabilità. *Variability and Mutability'. C. Cuppini, Bologna.*

Goswami, S., Murthy, C.A. and Das, A.K. (2018). Sparsity measure of a network graph: Gini index. *Information Sciences*, **462**, 16-39.

Milgram, S. (1967). The small world problem. *Psychology Today*, **2**(**1**), 60-67.

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

Newman, M.E. (2004). Analysis of weighted networks. *Physical Review E*, **70**(**5**), p.056131.

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature*, **393**(**6684**), 440-442.

Yu, H., Paccanaro, A., Trifonov, V., Gerstein, M.(2006). Predicting interactions in protein networks by completing defective cliques. *Bioinformatics*, **22**(**7**), 823-829.