Small world, Clustering, Structured Holes and redundancy

01.02.2006

Markose Thomas
01CS3019
markysays@gmail.com

I. Introduction

Complex networks are networks characterised by non-trivial topological structure. Unlike simple/compound networks they exhibit certain structural properties like

- Heavy tail in the degree distribution
- High clustering coefficient
- Preferential attachment
- Community structure

The structure evolves and self-organises itself in the process of evolution. In the next few sections we shall study some such properties of complex networks.

II. Small World Effect

Many networks exhibit an interesting behaviour known as the *small world effect*, discovered by Milgram [**?**] in 1969. In the original experiment, Milgram sent 60 letters to various people, asking them to forward the letter to one of two unacquainted targets. The senders only knew the recipient's name, occupation and general location. They were instructed to send the card to a person whom they knew on a first-name basis who they thought was most likely to know the target personally. That person would do the same, and so on, until it was delivered to the target himself.

Of all the letters that were sent, only about a fourth of them reached the target. Of these, 80% were delivered in four or fewer steps. Almost all the chains were less than 6 steps long. Thus it lead to the general hypothesis, that any two people in the US were connected by a chain of less than 6 people. This kind of behaviour was also observed in communities involving mathematicians and actors. The small world effect observed in Milgram's experiment has not gone unchallenged. The chief contention being that the small world effect did not take into account the large percentage of undelivered messages. Nevertheless, the experiment did help in establishing the small world effect as an important characteristic of many complex networks. More formally, it can be said that if a network shows the small world effect, then the diameter of the network grows logarithmically with the number of nodes in the network.

A. Finding whether a network shows small world effect

For an undirected network of n nodes the mean *geodesic*[1] path between vertices can be expressed as

$$l = 2\sum_{i \geq j} \frac{d_{ij}}{n(n+1)},$$

where d_{ij} is the geodesic distance from vertex i to vertex j. Most of the networks observed in real life (*e.g film actors, company directors, emails, internet, electronic circuits*) have $l \leq 6$. The shortest between vertices can be found using standard BFS for unweighted graphs. Since the time required for performing BFS from a single node is $O(m)$, the total time required to find l would be $O(mn)$. For networks, with more than one component, we can find l for each connected component, and take the harmonic mean of these l's to get the overall mean geodesic path length.

III. Clustering

The clustering coefficient for a vertex v in a network is defined as the ratio between the total number of connections among the neighbours of v to the total number of possible connections between the neighbours. A node in a social network with high clustering coefficient, would indicate that its neighbours know each other pretty well.

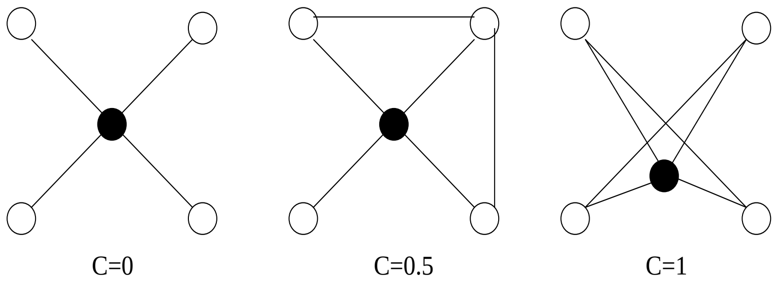

Fig. 1. Clustering Coefficient of Center Node for various networks

For a vertex i, the clustering index C_i is given by,

$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i-1)} : v_i, v_k \in N_i, e_{jk} \in E,$$

where N_i is the neighbourhood set of vertex i, $k_i = |N_i|$ and E is the set of edges in the network.

The clustering index of the whole network is then,

$$C = \frac{1}{N}\sum C_i$$

IV. Structural Holes and Redundancy

Structural holes are nodes (mainly in a social network) that seperates non-redundant sources of information, sources that are additive than overlapping.

[1]Geodesic path between two vertices is the shortest path between them.

Redundancy - There are two kinds of redundancies that are observed in social networks:

- **Cohesion** - Contacts strongly connected to each other are likely to have similar information and therefore provide redundant information benefits
- **Equivalence** - Contacts that link a manager to the same third parties have same sources of information and therefore provide redundant information benefits.