Reciprocity, Citation Networks Degree Distribution, Centrality

08.02.2006

Markose Thomas

01CS3019

markysays@gmail.com

I. Reciprocity

Reciprocity is the probability that two vertices in a directed network point to each other. It measurs the fraction of edfes between vertices that point both ways.

II. Citation Networks

A citation is a reference by a paper to a technical work done in another paper in a related field. If we build a network, where each node is a paper, and a directed edge exists between paper A and paper B if A cites B, then we will get what is known as a citation network. These networks, due to the strict time ordering of the publishing of papers, are acyclic.

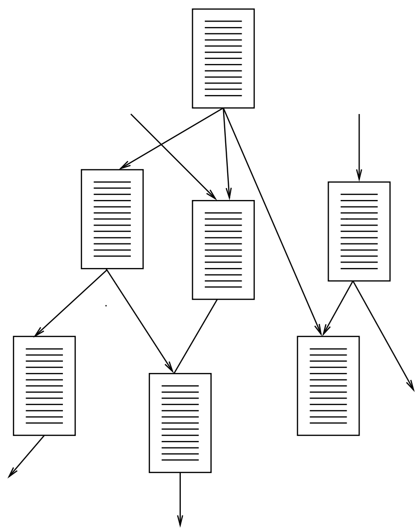

Fig. 1. A citation network

By studying such networks, Alfred Lotka in 1926 observed that the number of scientists who have written k papers falls off as $k^{-\alpha}$ for some constant α. This corresponds to the power law distribution.

III. Degree Distribution

The degree distribution of a network tells us the probability distribution of the degree of a random node in a network.

Let p_k be the fraction of vertices in the network that has a degree k. Hence, p_k is the probability that a vertex chosen uniformly at random has a degree k. Or in other words, the k versus p_k plot will give us the degree distribution of the network. For most of the real world networks, these distributions are right skewed with a long tail showing up values far above the mean - p_k varies as $k^{-\alpha}$ (power law).

Due to noisy and insufficient data, sometimes the data is slightly modified. Instead of looking at the degree distribution, we look at the cumulative degree distribution, defined by

$$P_k = \sum_{k'=k}^{\infty} p_{k'}$$

That is, P_k gives us the probability that the degree of the randomly chosen node is greater than or equal to k.

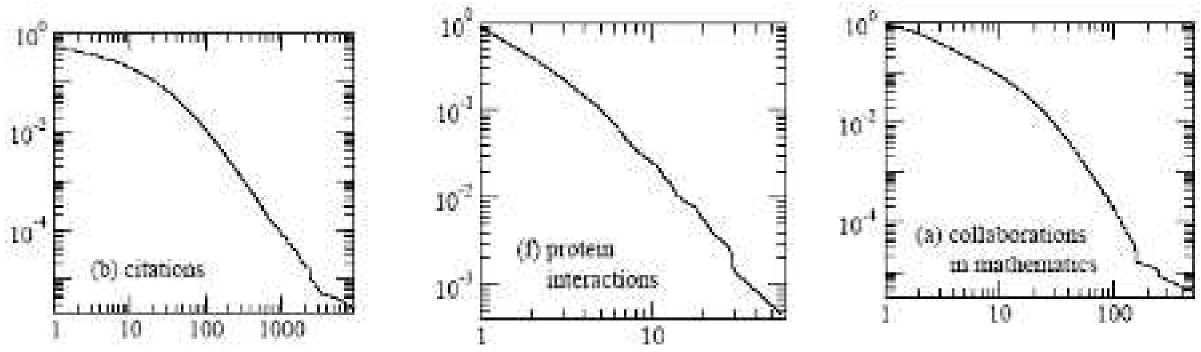

Fig. 2. Cumulative degree distributions of some complex networks

IV. Centrality

Centrality is an important measrure that is used within complex network analysis to determine the relative importance of a vertex within the graph (for example, how important a person is within a social network, or how well-used a road is within an urban network). In other words, the centrality of a node in a network is a measure of the structural importance of the node. These measures attempt to quantify the prominence of an individual actor embedded in a network. A central actor, presumably, has a stronger influence on other network members.

There are different measures of centrality for vertices, which accordingly give us different insights.

1. **Degree Centrality** - This is a measure of the immediate neighbours of a vertex. A star network would have very high degree centrality, whereas a line network would be much lesser degree centrality.
2. **Betweeness Centrality** - This centrality tries to measure the importance of a node in the network, in terms of the number of paths in the network that depend on it. Here, unlike degree centrality, a vertex with very low degree could also be deemed very important (*e.g if it is the sole node which connects two components in the network*).

   Let σ_{st} denote the number of shortest paths between vertices s and t. Let $\sigma_{st}(v)$ denote the number of shortest paths between vertices s and t that pass through vertex v. Then, the betweenness centrality measure of vertex v, $g(v)$ is given as

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

   Betweenness centrality for all nodes in a network can be computed in $O(nm)$ time [1].
3. **Closeness Centrality** - This is a measure which gives high weight to nodes which are closer to the rest of the nodes in the network. That is, those nodes which have short geodesic distances to other vertices in the graph. The closeness $C_c(v)$ for a vertex v is the reciprocal of the sum of geodesic distances to all other vertices in the graph.

$$C_c(v) = \frac{1}{\sum_{t \in V} d_G(v,t)},$$

   where $d_G(s,t)$ is the weight of the shortest path between s and t.
4. **Eigenvector Centrality** - (*covered later*).

References

[1] U. Brandes, A faster algorithm for betweenness centrality, Journal of Mathematical Sociology, 2001