Lecture Notes: Social Networks: Models, Algorithms, and Applications
Lecture 1: Feb 21, 2012
Scribes: Varun Nelakonda and Patrick Rhomberg

1 Configuration Model

Last time we discussed the Configuration Model and analysing the Clustering Coefficient of the model. Recall that the model takes an input of degree distribution of P_k.

Definition 1 *We are given a degree sequence* $(d_1, d_2, ..., d_n)$ *where* $d_1 \geq d_2 \geq ... \geq d_n$ *and the Configuration Model generates a random graph that realizes this provided degree sequence.*

Pick a pair of stubs uniformly at random and connect them. Repeat until no stubs left. Recall that the graph generated by the Configuration Model contains many vertices $v_1, v_2, ..., v_n$. What is the probability that a vertex v_i is connected to a vertex v_j?

$$\text{Probability[vertex } v_i \text{ and vertex } v_j \text{ are connected by an edge]} = \frac{d_i d_j}{2m}$$

In the above expression, d_i represents a stub from vertex i and similarly d_j is a stub from vertex j. m represents the number of edges. The above fraction represents the probability of a stub from vertex i being connected to vertex j. Similarly, we can also start with an arbitrary vertex and walk along on an arbitrary incident edge to the other end point. What is the degree distribution of the other end point? Specifically what is the probability that the degree of the other end point is some value k?

$$\text{Probability[degree of other end point is } k] = \frac{k}{2m-1} \cdot p_k \cdot n$$

$$\sim \frac{k}{2m} \cdot p_k \cdot n = \frac{k}{< k >} \cdot p_k$$

$< k >$ represents the average degree which is equal to $\frac{2m}{n}$. p_k is the degree distribution of the graph so a high degree correlates to $\frac{k}{<k>}$ being very large which means it's more likely to occur. A low degree correlates to less likely to occur. The **excess degree** if the other endpoint is the number of edges besides the one we used to arrive at that vertex (i.e. degree - 1). Therefore, the excess degree distribution of the other end point denoted q_k is $\frac{(k+1)}{<k>} \cdot p_k + 1$. We are just trying to analyse the probability of the other end point being $k+1$. This is extremely useful in analysing the Configuration Model.

Example 1: Calculate Clustering Coefficient in the Configuration Model

Recall : C(Clustering Coefficient) = Average probability of a pair of neighbours of a vertex being connected by an edge.

$$c = \sum_{ki,kj} \cdot \frac{k_i \cdot k_j}{2m} \cdot q_{ki} \cdot q_{kj}$$

Given the probability of i having k_i excess degree, what is the probability of edge j having k_j excess degree from i to j? We can calculate it as follows:

$$= \frac{1}{n} \cdot \left(\frac{<k^2> - <k>}{<k>^3}\right)$$

If $<k>$ and $<k^2>$ are fixed, i.e. independent of n, then c $\to 0$ as n $\to \infty$.

Power law distributions of the form $p(k) \propto k^{\alpha}$

1) If $\alpha \geq 2$, then $<k>, <k>^2$ are indeed finite

2) If $1\alpha < 2$, then $<k>^2$ is not finite as we may get large C from the Configuration Model

2 Preferential Attachment Model

The Preferential Attachment Model is motivated by observation of the power law distributions in a variety of settings.

–In-links of web pages : fraction of web pages with k inlinks $\sim \frac{1}{k^2}$

–Fraction of phone numbers that receive k phone calls per unit time $\sim \frac{1}{k^3}$

–Fraction of papers that receive k citations $\sim \frac{1}{k^3}$

Definition 2 *The Preferential Attachment Model is used for generating a network that yields networks with degree distribution which satisfies power law. This is a discrete time model.*

·At time $t = 0$, there is a single isolated node in the graph. An input of an arbitrary graph can also be used in this model.

·At time t, a new node arrives and we place m (parameter of model) edges between new node and older nodes.

But how do we determine the other end points? The other end points of each new edge is chosen with probability proportional to its (older node) degree.

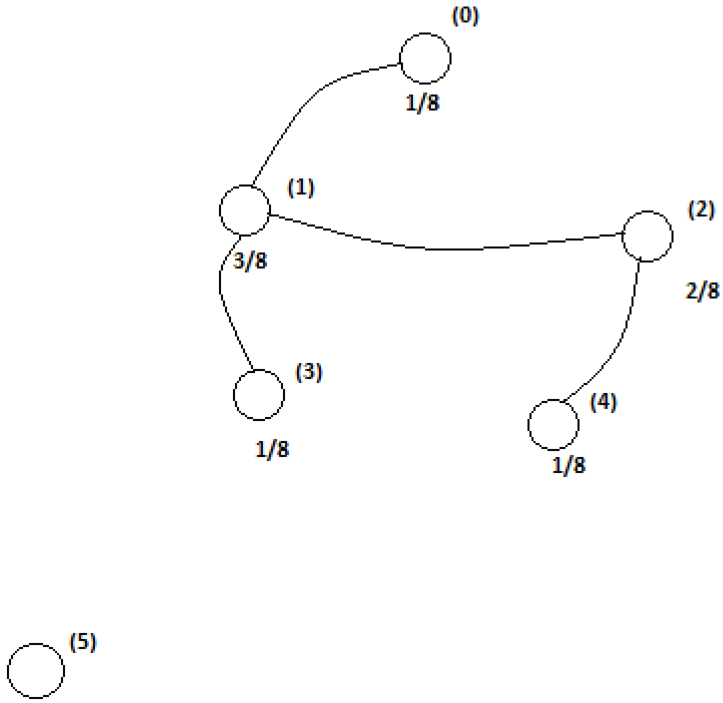

Figure 1: Probability node 5 connects to a node is the node's degree divided by the overall degree. For example, probability node 5 connects to node 2 is $\frac{2}{8}$

An experiment was conducted by Salgankik, Dodds, Watts on the unpredictability of "Rich get Richer" effects (based on if a node has high degree, the higher the probability of other nodes connecting to it). This experiment outlines how this notion of "Rich get Richer" is false. The experiment is outlined below.

Experiment:

–48 obscure songs with highly variable quality
–Number of downloads per song shown to users
–8 parallel copies of the website
–Users randomly sent to one of these 8 copies when they first visited
–Copies differ in aspects (Songs in random order, showing number of downloads etc.)

The copies of the website differed in ratings of the songs and the number of downloads. Not every website was a clone of itself. Even though all the copies were the same, the different aspects of the website caused the users to rate and download songs differently.

A lot of papers and experiments were conducted on degree distribution and power laws. A few older models are:

·Polya 1923 - Proposed Urn Models
·Yule 1925 - Proposed genetic diversity (Named Yule Process)

·Zipf 1949 Populations of cities
·Simon 1955 Wealth in economics
·Price 1965 “Networks of Scientific papers”
·Price 1976 “A general theory of bibliometric and other cumulative advantage processes”

References

[1] Mandelbrot. *Optimization Explanation*. 1950.

[2] Disouza et al. *Emergence of empared Preferential Attachment from optimization*. PNAS 2007.

[3] Barabasi, Albert. *The emergence of scaling in Random Networks*. Science 1999.