Random Walks in Power Law Graphs

Jeffrey John Geevarghese
Department of Computer Science & Engineering
National Institute of Technology(NIT), Calicut,India.

February 17, 2011

1 Problem Setting

Most complex networks like citation networks, co-authorship networks, social and biological networks, etc have been found to follow a *power-law* distribution. Continuing with the earlier analysis on finding *short* paths in complex networks, we investigate whether this property holds true in power-law graphs as well. We describe here a simple *random walk* strategy that achieves this under suitable assumptions and scales sub-linearly with the size of the network.

Outline We describe the idea of generating functions briefly in Section 2 and in Section 3, we show how a simple random walk stategy can search the graph in sub linear time under suitable assumptions.

2 Generating Functions

The generating function for the distribution of vertex degrees k is denoted by $G_0(x)$ and is given by

$$G_0(x) = p_0 + p_1 x + p_2 x^2 + ... + p_k x^k + ...$$

or

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k,$$

where p_k is the probability that a randomly chosen vertex has degree k.

Now, given the generating function for a graph with arbitrary degree distribution, we can ask what is the average degree of a randomly chosen vertex? This is given by

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k.$$

It is easy to see that this quantity is the same as $G_0'(1)$.

Consider a power law graph. A power law graph is a graph such that the probability that a vertex has degree k varies as

$$P(k) \sim ck^{-\gamma},$$

or more precisely,

$$p_k = ck^{-\gamma}.$$

For such a graph, we can write the generating function as

$$G_0(x) = c\sum_{k=1}^{\infty} k^{-\gamma}x^k.$$

An empirical observation is that in these networks there is an *abrupt cut-off* in the degree after a while, and hence for the purpose of analysis, we denote by m the maximum degree of the graph. The above equation can be rewritten as

$$G_0(x) = c\sum_{k=1}^{m} k^{-\gamma}x^k.$$

Note the c constant in the above equation. It is a *normalization* constant that depends on m and γ such that

$$c\sum_{1}^{m} k^{-\gamma} = 1,$$

or $G_0(1) = 1$.

3 Random Walk Search Strategy

The search strategy is a simple random walk starting at a randomly chosen vertex of the graph. At each step, we proceed to any one of the neighbours of the current vertex chosen at *random*. We exclude the nodes already visited in the last step as well. Using the formalism of generating functions described above, we analyse this startegy as follows:

We first note that the average degree of a node chosen at random and the one arrived by following a random edge are different. A random edge arrives at a vertex with probability proportional to the degree of the vertex, i.e.

$$p'(k) \sim kp_k.$$

The correctly normalized distribution is given by

$$\frac{\sum_k kp_k x^k}{\sum_k kp_k} = x\frac{G_0'(x)}{G_0'(1)}.$$

If we want the number of outgoing edges from the vertex we arrived at, but not include the edge we just came on, we need to divide by one power of x. Hence the *number of new neighours* on each step of a random walk is given by the generating function

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)},$$

where $G_0'(1)$ is the average degree of a randomly chosen vertex as we showed above. Note here that $G_1(1) = 1$.

Now in power law graphs, because of the strong local clustering, it is reasonable to look at the *second neighbours* of a node. For example, in real social networks, one would have atleast some knowledge of one's friend's friends. Hence we now compute the number of second neighbours encountered on doing a random walk. But it is easy to see that the distribution function for the number of second neighbours encountered will be $G_1(G_1(x))$. Therefore, the average number of second neighbours would be given by

$$z_2 = [\frac{\partial}{\partial x} G_1(G_1(x))]_{x=1} = [G_1'(1)]^2.$$

We see that this quantity depends on $G_1'(1)$ which in turn depends on $G_0'(1)$, which we calculate for a fixed γ and m. Now we take the *first* assumption, that is, based on empirical evidence, we assume $2 < \gamma < 3$, and now the *second* assumption which is on the cutoff value m, that is $m \sim N^{\frac{1}{\gamma}}$.Under these assumptions and in the limit $\gamma \to 2$, we get

$$z_2 \sim \frac{N}{\ln^2 N}.$$

See Appendix A for a detailed derivation. Now, as the random walk proceeds node to node, each node reveals more of the graph. The search cost S is defined as the number of steps until approximately the whole graph is revealed, i.e.

$$S \sim \frac{N}{z_2}.$$

It is easy to see that the search cost $S \sim \ln^2 N$, meaning the search cost scales sublinearly with the size of the network. Keep in mind that these are in the limit of small exponents and under the cutoff assumption we mentioned earlier. There is also a variation of the random walk strategy in which the high degree nodes are chosen at each step instead of a random vertex, and this scales sublinearly as well.

References

[1] S.H. Strogatz M.E.J. Newman and D.J. Watts, *Random graphs with arbitrary degree distributions and their applications*, **Phys. Rev. E**, 64:026118, 2001.

[2] R.M. Lukose L.A. Adamic and A.R. Puniyani, *Search in power law networks*, **Phys. Rev. E**, 64:4613546143, 2001.

4 Appendix A

Claim $z_2 \sim \frac{N}{\ln^2 N}$ under the following assumptions :

- the cutoff m is such that $m \sim N^{\frac{1}{\gamma}}$.
- $2 < \gamma < 3$.

Proof: We have shown earlier that

$$z_2 = [\frac{\partial}{\partial x} G_1(G_1(x))]_{x=1} = [G_1'(1)]^2.$$

Now, since

$$G_0(x) = \sum_1^m c k^{-\gamma} x^k,$$

we have

$$G_0'(1) = \sum_1^m c k^{1-\gamma} \sim \int_1^m x^{\gamma-1} dx = \frac{1}{\gamma - 2}(1 - m^{2-\gamma}).$$

We had also shown that

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)},$$

so,

$$G_1'(1) = \frac{1}{G_0'(1)} \frac{\partial}{\partial x} \sum_1^m c k^{1-\gamma} x^{k-1},$$

which on simplification gives

$$G_1'(1) \sim \frac{1}{G_0'(1)} \frac{m^{3-\gamma}(\gamma-2) - 2^{2-\gamma}(\gamma-1) + m^{2-\gamma}(3-\gamma)}{(\gamma-2)(\gamma-3)}.$$

Now, because of our assumption that $2 < \gamma < 3$, the exponent $2 - \gamma$ is negative, hence we neglect those terms which gives us,

$$G_1'(1) = \frac{1}{G_0'(1)} \frac{m^{3-\gamma}}{3-\gamma}.$$

Substituting the earlier value of

$$G_0'(1) = \frac{1}{\gamma - 2}(1 - m^{2-\gamma}),$$

we obtain the final expression for z_2 as

$$z_2 = [\frac{(\gamma-2)}{(1-m^{2-\gamma})} \frac{(m^{3-\gamma})}{(3-\gamma)}]^2$$

Making using of the first assumption that $2 < \gamma < 3$, applying the limit $\gamma \to 2$ and $m \sim N^{\frac{1}{\gamma}}$, we get

$$z_2 \sim \frac{N}{\ln^2 N}.$$