Self Similar (Scale Free, Power Law) Networks (II)

E6083: lecture 5

Prof. Predrag R. Jelenković

Dept. of Electrical Engineering
Columbia University , NY 10027, USA
{predrag}@ee.columbia.edu

February 14, 2007

Outline

Description of the Preferential Attachment

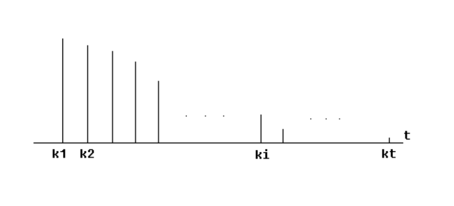
Basic model

- At time 0, one node is present
- At each discrete time step $t > 0$, a new vertex is added, with one undirected edge preferentially ($\propto k_i$ where k_i is the degree of the ith node) attached to one existing node (including itself).
- t_i represents the time when vertex i was added to the system.

More general model

Each new vertex has m edges linking to other vertexes, which allows multiple edges and also loops.

Informal derivation of preferential attachment



Then, at time t,

$$\frac{dk_i}{dt} = \frac{k_i}{2t} \Rightarrow k_i = \left(\frac{t}{t_i}\right)^{1/2}.$$

Thus, the degree D distribution is obtained by

$$\mathbb{P}[D > x] = \mathbb{P}[t_i < t/x^2] = \frac{t}{x^2(t+1)},$$

which implies $\mathbb{P}[D = x] \sim \frac{1}{x^3}$.

Rigorous treatment

Formal definitions

- $d_G(v)$ - total degree of vertex v in G.
- G_1^t - state of the graph at time $t \geq 1$ for $m = 1$
- G_1^1 - starting graph with one vertex and one loop (pointing to itself).
- Given G_1^{t-1}, form G_1^t by adding the vertex v_t together with a single edge directed from v_t to v_i, where i is chosen randomly with

$$\mathbb{P}[i = s \mid G_1^{t-1}] = \begin{cases} d_{G_1^{t-1}(v_s)}/(2t-1) & 1 \leq s \leq t-1 \\ 1/(2t-1) & s = t. \end{cases}$$

For $m > 1$, the graph is constructed by adding m edges one at a time. Equivalently, the process G_m^t can be obtained by coalescing every m vertexes of G_1^{mt} into one vertex.

Analysis due to Bollobás et al

Proving recipe

Used in many other papers.

Let the number of vertices of G_1^n with indegree equal to d be $X_n(d)$. The martingale

$$X_t = \mathbb{E}[X_n(d) \mid G_t]$$

satisfies that $\mid X_{t+1} - X_t \mid$ is bounded by two.

If $\{X_t\}_{n \geq t \geq 0}$ is a martingale with $\mid X_{t+1} - X_t \mid \leq c$ for $t = 0, 1, \cdots, n-1$, then,

$$\mathbb{P}[\mid X_n - X_0 \mid \geq x] \leq \exp\left(-\frac{x^2}{2c^2 n}\right).$$

Applying Azuma-Hoeffding inequality, we obtain that $X_n(d)$ is very concentrated around its mean, and thus only need to compute $\mathbb{E}[X_n(d)]$.

Compute $\mathbb{E}[X_n(d)]$

This needs some work, and it turns that for $m = 1$

$$\mathbb{E}[X_n(d)] \sim \frac{4n}{(d+1)(d+2)(d+3)},$$

and for $m > 1$,

$$\mathbb{E}[X_n^m(d)] \sim \frac{2m(m+1)n}{(d+m)(d+m+1)(d+m+2)}.$$

Is the topology unique? Is degree distribution enough? Intuition says no; however...(Kleinberg etc. 2005)

Erdos-Renyi Model: Isomorphism

Let $G_{rnd}(\infty, p)$ denote the probability distribution on graphs with vertex set $\mathbb{N}$, in which each edge (i, j) is included independently with probability $0 < p < 1$. There exists an infinite graph R, such that a random sample from $G_{rnd}(\infty, p)$ is isomorphic to R with probability 1.

Isomorphism of infinite limits for PA scale-free graphs: $m = 1, 2$

For $d = 1, 2$, there is a graph R such that a random sample from $G_{rnd}(\infty, p)$ is isomorphic to R with probability 1.

Infinite limits of PA scale-free graphs: $m \geq 3$

For each out-degree $m \geq 3$, it is not the case that two independent random samples from $G_{PA}(\infty, p)$ are isomorphic with probability 1.

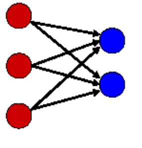
Preferential Attachment: Continued...

What if attaching proportional to k_i^α?

- Only linear preferential attachment yields power-law graphs.
- If $\alpha > 1$, eventually one person gets all the links.
  There is a finite time after which no one else gets anything!
- If $\alpha < 1$, the degree distribution follows a stretched exponential.

Limitation of preferential attachment

- Outdegree $= m$ for directed graph.
- Global information.
- Number of nodes increases linearly.
- In a large scale experimental study by Kumar et al, they observed that the Web contains a large number of small bipartite cliques (cores)



a $K_{3,2}$ clique

One generalization

Both in-degree and out-degree are power law(Bollobás et al)

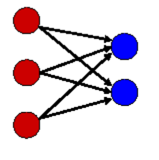
A directed graph which grows by adding single edge at discrete time steps. At each such step a vertex may or may not also be added. Let $\alpha, \beta, \gamma, \delta_{in}$ and δ_{out} be non-negative real numbers, with $\alpha + \beta + \gamma = 1$.

1. With probability α, add a new vertex v together with an edge from v to an existing vertex w, where w is chosen proportionally to $d_{in} + \delta_{in}$.
2. With probability β, add an edge from an existing vertex v to an existing vertex w, where v and w are chosen independently, v according to $d_{out} + \delta_{out}$, and w according to $d_{in} + \delta_{in}$.
3. With probability γ, add a new vertex w and an edge from an existing vertex v to w, where v is chosen according to $d_{out} + \delta_{out}$.

Copying Model

- Parameters:
  1. The out-degree d (constant) of each node
  2. Probability p.
- The process:
  1. Nodes arrive one at the time
  2. A new node selects uniformly one of the existing nodes as a prototype
  3. The new node creates d outgoing links. For the *ith* neighbor of the prototype node
     - with probability p it connects to the *ith* neighbor of the prototype node
     - with probability $1 - p$ it selects the target connection uniformly at random among all the existing nodes

- Power law degree distribution with exponent $\beta = (2 - p)/(1 - p)$
- Number of bipartite cliques of size K_{id} is ne^{-i}



The model has also found applications in biological networks: copying mechanism in genes

1-Step Random Walk with Self-Loop (A. Blum, et al 2006)

Given k and p, at time t, vertex v makes k connections to the existing graph by repeating the following process k times:

1. Pick an existing node v uniformly at random from $\{v_0, \cdots, v_{t-1}\}$.
2. With probability p stay at v; with probability $1-p$ take a $1-$step walk to a random neighbor of v.
3. Add an edge from v to the node where the random walk stops.

In the directed version, the edges added are directed from v_t into the existing graph. In the undirected version, edges are undirected.

analyzing 1-step walk

$D_i(t)$ is the expected number of nodes with in-degree i at step t, then,

$$\begin{aligned} D_i(t+1) = D_i(t) &+ \\ &\frac{pk}{t}(D_{i-1}(t) - D_i(t)) + \\ &\frac{(1-p)k}{t}((i-1)D_{i-1}(t) - iD_i(t))\frac{1}{k} \end{aligned}$$

Substitute $D_i(t) = c_i t$ in the above equation,

$$c_i = pk(c_{i-1} - c_i) + (1-p)((i-1)c_{i-1} - ic_i),$$

to get $c_i \sim Ci^{-\frac{2-p}{1-p}}$. (same as the copying model)

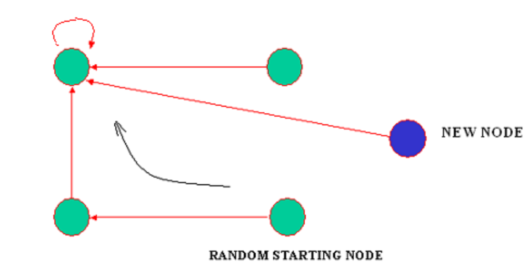
Geometric random walk (A. Blum et al 2006)

Given k and p, at time t, vertex v makes k connections to the existing graph by repeating the following process k times:

1. Pick an existing node v uniformly at random from $\{v_0, \cdots, v_{t-1}\}$.
2. Flip a coin of bias p.
3. If the coin comes up heads add an edge from v_t to the current node and stop.
4. If the coin comes up tails, move to a random neighbor of the current node and go back to (2).

In the directed version, the edges added are directed from v_t into the existing graph. In the undirected version, edges are undirected.

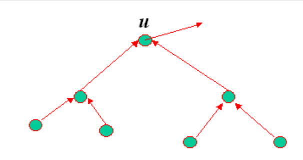
First and second toss = tail, second toss = head.



Existing results (only for directed graphs)

Let $l_i(u)$ be the number of level i descendents of node u. For example, $l_1(u) = \#$ of children, $l_2(u) = \#$ of grandchildren, etc.

Let $\beta = (\beta_1, \beta_2, \cdots)$ be a sequence of real numbers with $\beta_1 = 1$.
Define virtual degree of u with respect to β to be
$v_\beta(u) = 1 + \beta_1 l_1(u) + \beta_2 l_2(u) + \beta_3 l_3(u) + \cdots$.



$v(u) = 1 + 2\beta l + 4\beta_2 + 0\beta_3 + 0\beta_4 + \cdots$.

Existing results

Theorem: There always exist β_i, dependent on p with $\beta_0 = 1$, $\mid \beta_i \mid \leq 1$ such that the expected increase in $v(u)$ from step t to $t+1$ is $v(u)p/t$. Furthermore, $\beta_i == O(\rho^i), 0 < \rho < 1$.

Let $v_t(u)$ be the virtual degree of node u at time t and t_u be the time when node u first appears. Then, for any node u and time $t \geq t_u$,

$$\mathbb{E}[v_t(u)] = \Theta((t/t_u)^p).$$

Proof: $\mathbb{E}[v_t(u)] = (1 + \frac{p}{t-1})\mathbb{E}[v_{t-1}(u)]$, hence,
$\mathbb{E}[v_t(u)] = \prod_{i=t_u}^{t}(1 + p/i) = \Theta\left(\left(\frac{t}{t_u}\right)^p\right)$

Existing results

For real degree $d(u)$

Let $d_t(u)$ be the virtual degree of node u at time t and t_u be the time when node u first appears. For any node u and time $t \geq t_u$,

$$\mathbb{E}[d_t(u)] \geq \Omega((t/t_u)^{p(1-p)}).$$

Proof

Observe that

$$\mathbb{E}[d_{t+1}(u)] \geq \mathbb{E}[d_t(u)] + \frac{p(1-p)}{t}\mathbb{E}[d_t(u)],$$

then...

Only partial results are proved for the random walk model.

Scale free network caused by random walk

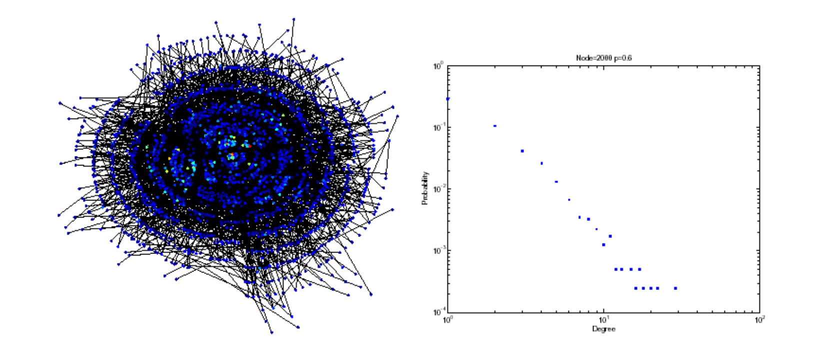


Figure: Node=2000, Random Walk $p = 0.6$.