Stat 206 (Spring 2003): Random Graphs and Complex Networks **01/28/03**

Lecture 4: Proportional Attachement models and the Yule process

Lecturer: David Aldous *Scribe: Rahul Jain*

In this lecture, we see two preferential attachment models of random graphs treated heuristically via the Yule branching process.

4.1 Yule process - a branching process in continuous time

In the Yule process, individuals live forever. And for each individual living at time t during $[t, t + dt]$ there is a chance $1dt$ to have a child. So, each individual gives birth at rate 1. Define

$$Z_t = \#\text{individuals at time } t, Z_0 = 1.$$

Let W_1 be the waiting time for first birth, which is exponentially distributed with rate 1. Let W_k be the interarrival time between the birth of the $(k-1)$th and the kth individual. Then, since there are $k - 1$ individuals and each gives birth with rate 1, and exponential distribution has memoryless property, $W_k \sim \text{EXP}(k)$. Also note that W_k is independent of W_{k+1}.

Then, we can define $Z_t = \min\{k : \sum_1^k W_i > t\}$. Yule (1924) showed that
(i) $EZ_t = e^t$.
(ii) $Z_t \sim \text{GEOM}(e^{-t})$.
(iii) $Z_t/e^t \to W$ a.s., with $W \sim \text{EXP}(1)$.
Yule set up and solved the following differential equation:

$$\frac{dP(Z_t = i)}{dt} = -iP(Z_t = i) + (i-1)P(Z_t = i-1)$$

Here, the first term on RHS is because there may be i individuals at time $t - h$ and there is no birth in $[t - h, t]$, and the second term is if there are $i - 1$ individuals at time $t - h$ and there is one birth in $[t - h, t]$ with rate $i - 1$.

Kendall gave a clever bijection between the Yule process and the Poisson point process. A Poisson point process is a spatial random process of points (say in R^2).

We can construct the bijection in the following way: Let the first point A be at (X, Y). Then, X has distribution EXP(1). The Y-axis represents the e^t-axis in the Yule process. Consider only points right of point A. B considered child of A.... Let $N(y)$ be the number of points above level y. Then,...

Proposition 4.1 *(1) The Process* $(N(e^t), t \geq 0)$ *is exactly same as* $(Z_t, t \geq (0)$.
(2) For large y, $N(y)$ *= number of points in the rectangle*$(W, y) \simeq Wy$.

The tree structure here is what is mimicing the Yule process.

4.2 Proportional attachment model I

We create a digraph by adding new vertices chosen randomly but with probabilities proportional to $1 + in - degree$. Fix $m \geq 0$. Start with an arbitrary graph G_m on m vertices. To construct the graph, consider a deck of m cards. For each vertex V, deal cards until get m distinct labels $V_1, ..., V_m$. For vertex V, create m edges $V \to V_i$. Create $(m+1)$ new cards: $V_1, ..., V_m$.

We get a random graph G_n on n vertices. Study $D_n = 1 + in - degree$ of uniform random vertex of G_n.

Claim 4.2 *As $n \to \infty, P(D_n = i) \to P(D = i)$, and distribution of D $GEOM(e^{-\frac{mT}{m+1}}), T \sim EXP(1)$. Then, as $i \to \infty, P(D = i) \sim c_m i^{-(2+\frac{1}{m})}$.*

To see the last part, note that $Y \sim$ GEOM(p), so $P(D > i|T) = (1 - e^{-\frac{mT}{m+1}})^i$ which gives

$$\begin{aligned} P(D > i) &= \int_0^\infty (1 - e^{-\frac{mt}{m+1}})^i e^{-t} dt \\ &= \frac{m+1}{m} \frac{\Gamma(i+1)\Gamma(1+1/m)}{\Gamma(i+2+1/m)} \end{aligned}$$

For m=1, $P(D = i) = \frac{4}{i(i+1)(i+2)} \sim i^{-3}$.

Heuristics: Make e^t vertices arrive by time t. Consider vertex V arriving at time t. $D(t) = 1 + in - degree =$ #cards V in the deck.

To calculate probability that D(t) increases by 1 (new in-edges) in $[t, t+\delta]$, note that there are $(m+1)e^t$ cards in deck by time t, and δe^t new vertices appear in the time interval. Each new vertex has m picks, and $D(t)$ cards have label V, Thus, the probability is given by

$$\frac{m\delta e^t D(t)}{(m+1)e^t} = \frac{m}{m+1} D(t)\delta.$$

$D(t)$ is Yule process, rate $\frac{m}{m+1}$, i.e. $D(t)$ has GEOM($e^{-\frac{mt}{m+1}}$) distribution.

Also note that "age" of any arrival (i.e., how long ago a given point was born) is $EXP(1)$.

4.3 Proportional attachment model II

Now, we create a digraph by adding new vertices chosen randomly but with probabilities proportional to total degree. Fix out-degree $m \geq 0$. For each new vertex, choose m end-vertices with probabilities proportional to total degree. For $m = 1$, it is same as previous model. Study D_n, the total degree of a random vertex in G_n.

Claim 4.3 *As $n \to \infty$, $P(D_n = i) \to P(D = i)$, and distribution of D $\sum_{i=1}^m GEOM(e^{-\frac{1}{2}T})$, $T \sim EXP(1)$, and the geometries are independent.*

Fix m: start with arbitrary digraph G_m on m vertices. For each edge $V \to W$, deal cards until get m distinct labels: $V_1, ..., V_m$. Create new vertex V and m edges $V \to V_i$. Create $2m$ new cards.

Heuristics: Make e^t vertices arrive by time t. Consider typical vertex V. Create m cards V, assign m different colors. Later, when an edge $W \to V$ is created, make the new V and copy the color of chosen V-card. $D_Y(t) =$ #Yellow V cards at time t.

Claim 4.4 *For a random vertex, $D_Y is GEOM(e^{-T/2}), T \sim EXP(1)$. The process is same for arbitrary m, as for $m = 1$.*

Example: Calculate $P(D = i) = \frac{2m(m+1)}{(i+m)(i+m+1)(i+m+2)}$, $i = m, m+1$. This gives

$$P(D = i) \sim c_m i^{-3}.$$

For rigorous proof, see paper by Bollobas, et al [BRST02]

References

[BRST02] B.Bollobas, O.Riordan, J.Spencer, G.Tusnady), "The degree sequence of a scale-free random graph process", *Random structures and algorithms*, 18:3, 2001, pp. 279-290.