M375T/M396C: Topics in Complex Networks **Spring 2013**

Lecture 12 — February 21

Lecturer: Ravi Srinivasan *Scribe: Chih-Hung Chen*

12.1 Preferential attachment models, Yule process

In this lecture, we would like to recap the Barabasi-Albert (BA) model mentioned in our previous section and move on to Yule process.

12.1.1 Barabasi-Albert (BA) model

Recap: The Barabasi-Albert model is an algorithm to generate random networks using a preferential attachment process. The network starts with an initial network of m nodes. One new node is added to the network at each time $t \in \mathbb{N}$. The preferential attachment process is stated as follows,

- With a probability $p \in [0,1]$, this new node connects to m existing nodes uniformly at random.
- With a probability $1-p$, this new node connects to m existing nodes with a probability proportional to the (in-)degree (degree for an undirected graph or in-degree for a directed graph) of node which it will be connected to.

In other words, it is more likely to see the new nodes to connect to the already highly linked nodes.

Remarks: The term of "connect to" has different meanings. Figure 12.1 (a) shows that edges have no orientation in an undirected graph. The degree of old node and new node are one in an undirected graph. On the other hand, edges have particular direction in an directed graph. The in-degree of a node is defined as the number of head endpoints it has. Figure 12.1 (b) shows that the old node has in-degree $d^{\text{in}} = 1$ and out-degree $d^{\text{out}} = 0$.

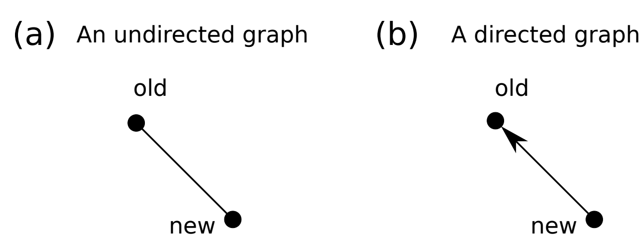

Figure 12.1. An undirected graph and a directed graph.

Properties: We will show that the degree distribution resulting from the BA model follows a power law of the form $d(k) \sim k^{-\beta}$, here k is the degree of nodes. The power law exponent $\beta = \frac{3-p}{1-p}$ in an undirected graph and $\beta = \frac{2-p}{1-p}$ in a directed graph.

We want to prove the properties above but first we think about the simplest case: when $p = 0$, i.e. $\beta = 3$ in an undirected graph. A heuristic argument is mentioned as follows:

12.1.2 An undirected graph, $\beta = 3$

Definition Let i be a node, $d_i(t)$ be the degree of node i at time $t \in \mathbb{N}$, and t_i be the time that node i appears.

Proof: We think of t as being continuous and take the derivative of $d_i(t)$ with respect to t to obtain,

$$\frac{d}{dt}\left(d_i(t)\right) = \frac{d_i(t)}{\sum\limits_j d_j(t)} = \frac{d_i(t)}{2t},$$

where $\sum\limits_j d_j(t)$ is the total degree of the entire network at time t and we know that the value is exactly $2t$ in an undirected graph. We integrate both sides with respect to t to obtain

$$\log\left(d_i(t)\right) - \log\left(d_i(t_i)\right) = \frac{1}{2}\log t - \frac{1}{2}\log t_i. \tag{12.1}$$

Note that $d_i(t_i) = 1$, then we have

$$\log\left(d_i(t_i)\right) = \log(1) = 0,$$

We substitute that into equation (12.1) to obtain

$$d_i(t) = \sqrt{\frac{t}{t_i}}.$$

If we assume that t_i is uniformly distributed between 0 and t, i.e. $t_i \sim \text{unif}(0, t)$, we see

$$\mathbb{P}\left(d_i(t) > k\right) = \mathbb{P}\left(\sqrt{\frac{t}{t_i}} > k\right) = \mathbb{P}\left(t_i < \frac{t}{k^2}\right) = \frac{1}{k^2}.$$

Therefore, we get a power-law degree distribution:

$$\mathbb{P}\left(d_i(t) = k\right) \approx -\frac{d}{dk}\mathbb{P}\left(d_i(t) > k\right) \sim \frac{1}{k^3}.$$

□

12.1.3 A directed graph, $\beta = 2$

We have seen that the BA preferential attachment process generates an undirected graph with a degree distribution $\sim k^{-3}$. Now, we want to show the power law exponent $\beta = 2$ in a directed graph. Again, we start with the simplest case: when $p = 0$. Let $d_i(t)$ be the in-degree of node i at time t. Notice that the total in-degree of the entire network $\sum_j d_i^{\text{in}}(t) = t$ in this case. A similar argument is used and notice that all one half factors are gone this time. Consequently, we obtain

$$d_i^{\text{in}}(t) = \frac{t}{t_i}.$$

This leads to a power-law degree distribution with the exponent $\beta = 2$, i.e.

$$\mathbb{P}\left(d_i(t) = k\right) \sim \frac{1}{k^2}.$$

12.1.4 Connections to Yule process

The BA model is actually a special case of a general model, Yule process. The dynamics of a simple balls and bins reinforcement process is also related to the Plya's urn model, which is the opposite to sampling without replacement.

Here, we think of a species as a ball contained in a bin which refers to its associated genus. We assume that a mutation happens in one species which is picked uniformly at random at each time. This species gives a mutation

- with a probability p, and this mutation creates a new genus. In other words, a new bin is created and containing one ball.
- with a probability $1 - p$, and this mutation creates a new species in the same genus. In other words, one new ball created in the same bin.

Figure 12.2 shows an example of Yule process.

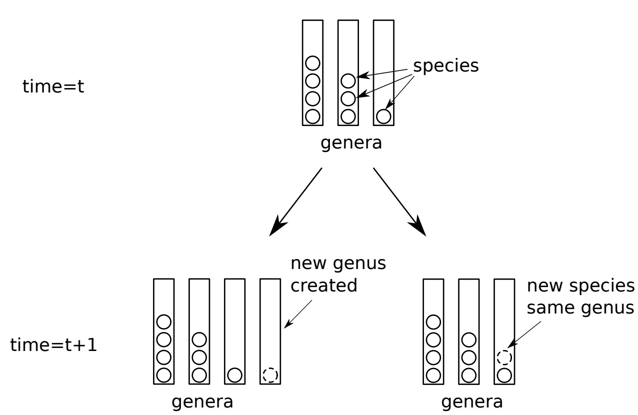

Figure 12.2. Sketch: an example of Yule process.

Theorem 12.1. *Let $X_i(t)$ be the number of genera containing exactly i species, here $i = \{1, 2, 3, ...\}$. The following three statements hold:*

(i) For each $i \geq 1$, $\exists\, C_i$ s.t. $\frac{X_i(t)}{t} \to C_i$ as $t \to \infty$.

(ii) $C_1 = \frac{p}{2-p}$, and $C_i = C_{i-1}\left(1 - \beta/i + O(i^{-2})\right)$ with $\beta = \frac{2-p}{1-p}$.

(iii) This implies that $\log(\frac{C_i}{C_1}) \sim -\beta \log(i) \Rightarrow C_i \sim i^{-\beta}$.

Proof of *(i)*: We think about this by the following two steps:

(1) Show that $\mathbb{E}[X_i(t)]/t \to C_i$ and this implies that $X_i(t)/t \to C_i$ as $t \to \infty$.

(2) A concentration result $X_i(t) \approx \mathbb{E}[X_i(t)]$ for large t.

Let $N(t)$ be the number of species at time t and we know $N(t) = N(0) + t$, here $N(0) = 1$ if the system starts with one species. The evolution equation for $X_1(t)$ is

$$X_i(t+1) - X_i(t) = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } (1-p)\frac{iX_i(t)}{N(t)} \\ 0 & \text{with probability } 1 - \left(p + (1-p)\frac{iX_i(t)}{N(t)}\right) \end{cases} \tag{12.2}$$

Let's look at $i = 1$ first. X_1 is the number of genus with exactly 1 species. Based on equation (12.2), we have a recurrence relation for the expected value of X_1, which is:

$$\mathbb{E}[X_1(t+1)] - \mathbb{E}[X_1(t)] = p - (1-p)\frac{\mathbb{E}[X_1(t)]}{N(t)} \tag{12.3}$$

$$\tag{12.4}$$

Let $\Delta_1(t) = \mathbb{E}[X_1(t)] - C_1 t$. Plugging it into equation (12.3), we see

$$\begin{aligned} \Delta_1(t+1) - \Delta_1(t) &= \mathbb{E}[X_1(t+1)] - \mathbb{E}[X_1(t)] - C_1 && (12.5) \\ &= p - (1-p)\frac{\mathbb{E}[X_1(t)]}{N(t)} - C_1 && (12.6) \\ &= p - (1-p)\frac{\Delta_1(t) + C_1 t}{N(t)} - C_1 && (12.7) \end{aligned}$$

After some algebra, we have a recurrence relation for Δ_1 as follows,

$$\begin{aligned} \Delta_1(t+1) &= \Delta_1(t)\left(1 - \frac{1-p}{N(t)}\right) + (p - C_1) - \frac{(1-p)C_1 t}{N(t)} \\ &= \Delta_1(t)\left(1 - \frac{1-p}{N(t)}\right) + \frac{(p - C_1)N(0)}{N(t)} + \frac{[(p - C_1) - (1-p)C_1]t}{N(t)} && (12.8) \end{aligned}$$

Let $\gamma_t = 1 - \frac{1-p}{N(t)}$ and $s_t = \frac{(p-C_1)N(0)}{N(t)}$, and choose $C_1 = \frac{p}{2-p}$ to make the last term vanish. Equation (12.8) becomes

$$\Delta_1(t+1) = \gamma_t \Delta_1(t) + s_t.$$

Next, we will say that $\Delta_1(t+1)$ is bounded above by $\log(t)$ (up to some constant factor). Given that $|\gamma_t| \leq 1$, it is easily shown that

$$\begin{aligned}
|\Delta_1(t+1)| &\leq |\gamma_t||\Delta_1(t)| + |s_t| \\
&\leq |\Delta_1(t)| + |s_t| \\
&\leq |\Delta_1(t-1)| + |s_{t-1}| + |s_t| \\
&\leq |\Delta_1(t-2)| + |s_{t-2}| + |s_{t-1}| + |s_t| \\
&\leq \ldots \\
&\leq |\Delta_1(0)| + \sum_{v=1}^{t} |s_v|.
\end{aligned}$$

Note that the first term $|\Delta_1(0)|$ is a constant and $s_t = \frac{(p-C_1)N(0)}{N(t)} \sim \frac{A_0}{t}$, here $A_0 = (p - C_1)N(0)$, which is a constant. In addition, we can show that

$$\sum_{v=1}^{t} |s_v| \sim \sum_{v=1}^{t} \frac{A_0}{v} \approx A_0 \log(t) \quad \text{if } t \text{ is large.}$$

This implies that $\Delta_1(t)$ grows no faster than $\log(t)$ as $t \to \infty$, then we have

$$|\Delta_1(t)| = O(\log(t)) \Rightarrow \frac{|\Delta_1(t)|}{t} \to 0 \quad \text{as } t \to 0.$$

That is

$$\frac{\mathbb{E}[X_1(t)]}{t} \to C_1 \quad \text{as } t \to 0.$$

Similarly, we can use this argument to show that this holds for all i. We have already finished the proof for step (1) but we will not go through the proof for step (2) in this lecture. Please see the attached notes for details.

Proof of (*iii*): From (*ii*), we have

$$\begin{aligned}
C_i &= C_{i-1} \left(1 - \beta/i + O(i^{-2})\right) \\
&= C_1 \left[\prod_{j=1}^{i} \left(1 - \beta/j + O(j^{-2})\right)\right].
\end{aligned}$$

Take the log of both sides to obtain,

$$\log C_i = \log C_1 + \sum_{j=1}^{i} \log\left(1 - \beta/j + O(j^{-2})\right)$$

Note that $\log(1+x) \approx x$ for small x, so we have $\log\left(1-\beta/j+O(j^{-2})\right) \approx -\beta/j+O(j^{-2})$ for large j. Now we can say

$$\begin{aligned}\log C_i &\approx \log C_1 + \sum_{j=1}^{i}\left(-\beta/j+O(j^{-2})\right)\\ &\approx \log C_1 - \beta \log i\\ &= \log C_1\left(i^{-\beta}\right).\end{aligned}$$

It is found that $\frac{C_i}{C_1} \sim i^{-\beta}$. So we have finished the proof for (iii).