Handout: Power Laws and Preferential Attachment

1 Preferential Attachment

Empirical studies of real world networks revealed that degree distribution often follows a heavy-tailed distribution, a power law. At that time, there were two kinds of network models: the Erdos-Renyi random graph $G_{n,p}$ and the Small World graphs of Watts and Strogatz. In both models the degrees were very close to the mean degree and there was little variation. Thus, there was the question of finding natural processes that could generate graphs with power law degree distributions. In 1999 Barabasi and Albert reinvented the model of Preferential Attachment (de Solla Price had introduced a similar model called Price's model in 1976, which in fact generalized the *Simon's model* introduced by Herbert Simon in 1955) that exhibited power-law degree distribution and renewed interest in the study of networks.

Definition 1 (Preferential Attachment (PA)) *Consider the sequence of directed graphs $\{G_t\}_{t \geq 0}$ where $G_t = (V_t, E_t)$, V_t is the vertex set and E_n the edge set. Given G_0, the graph G_{t+1} is constructed from G_t according to the following rule:*

1. *A new vertex v_{t+1} is introduced: $V_{t+1} = V_t \cup \{v_{t+1}\}$*

2. *We add a single directed edge from v_{t+1} to a vertex u in V_t, $E_{t+1} = E_t \cup \{(v_{t+1}, u)\}$, according to the following scheme.*

   - *(Uniform Attachment) With probability $p < 1$ we pick a vertex u uniformly at random.*
   - *(Preferential Attachment) With probability $q = 1 - p$ we pick a vertex $u \in V_t$ with probability proportional to its in-degree $q(u) \propto d_u$,*

The intuition behind the model stemmed from the general belief that power-laws correspond to underlying organization principles and feedback mechanisms. In particular, in the PA model the rich get richer effect is explicitly incorporated through the growth process. Since, then preferential attachment models and their variants have been extensively studied and their various descendants are employed to generate realistic network models.

The challenge with the PA model and similar growth models is to make precise quantitative predictions about them despite their intricate incremental construction. The basic tool that enables the analysis of such models is the Differential Equation Method.

1.1 Differential Equation Method

In order to analyze any complex process one needs to distinguish between the few variables that really matter and omit minor details or intricacies. In our case, incremental growth processes can be seen as mappings from one graph to a new graph. The hope is that each step will not change the properties of the graph too much and thus we hope to be able to make predictions. The ability

of analysis depends on whether we can track the evolution of the process using only a handful of manageable quantities. To this end, given the probabilistic nature of the PA model, the following concept is indispensable:

Definition 2 *A sequence of random variables (or vectors) $\{X(t)\}_{t\geq 0}$ is called Markov iff*

$$\mathbb{P}(X(t+1)|X(1),\ldots,X(t)) = P(X(t+1)|X(t)) \tag{1}$$

Intuitively, our aim is to find a (vector) X_t such that we can summarize the state of our process, in the sense that in order to track how X_{t+1} changes we only need the current values X_t and nothing else. This is essentially the usefulness of the Markov property. Next, we show how to utilize this concept to analyze the Preferential Attachment model through the *Differential Equation Method*. The basic steps of the method are:

1. **Markovian Dynamics:** Find a (vector) function $\mathbf{Z} = \mathbf{f}(G)$ of a graph G such that the sequence $\{\mathbf{Z}_t\}_{t\geq 0} = \{\mathbf{f}(G_t)\}_{t\geq 0}$ is (approximately) Markov.

2. **Conditional Change:** assuming that the state $\mathbf{Z}_t$ is known, compute the conditional expectation at time $t + 1$:
$$\mathbb{E}[\mathbf{Z}_{t+1} - \mathbf{Z}_t|\mathbf{Z}_t] = \mathbf{f}(\mathbf{Z}_t) \tag{2}$$

3. **Rate Equation:** If the function $\mathbf{f}$ is (approximately) linear $\mathbf{f}(\mathbf{Z}_t) = A_t\mathbf{Z}_t + \mathbf{b}_t$, set $\mathbf{z}(t) = \mathbb{E}[\mathbf{Z}_t]$ and take expectation with respect to $\mathbf{Z}_t$:

$$\mathbf{z}(t+1) - \mathbf{z}(t) = A_t\mathbf{z}(t) + \mathbf{b}_t \tag{3}$$

4. **Fluid limit:** consider the ordinary differential equation approximation for large $t \gg 1$:

$$\dot{\mathbf{z}}(t) = A_t\mathbf{z}(t) + \mathbf{b}(t) \tag{4}$$

5. **Solution of Ordinary Differential Equation (ODE):** use the boundary conditions and solve the ODE for $\mathbf{z}(t)$.

6. **Concentration:** argue that the probability that the random vector $\mathbf{Z}_t$ deviates significantly from its expectation $\mathbf{z}(t)$ is very small.

The last step is beyond the scope of the class at is very technically involved. What we gain by using the differential equation method is that we are able to work effectively with *expected deterministic quantities* and track their changes instead of working with random variables. That is the power of the method. What allows us to carry out Step 2 is the Markov property and Step 3 is possible because of linearity of expectation and the following property.

Lemma 1 (Tower Property) *Given random variables X, Y, it holds that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$*

Proof: Here, we prove the lemma only for discrete random variables but the general case follows easily.

$$
\begin{align}
\mathbb{E}[X] &= \sum_x x P(X = x) \tag{5}\\
&= \sum_x x \sum_y P(X = x, Y = y) \tag{6}\\
&= \sum_x x \sum_y P(X = x | Y = y) P(Y = y) \tag{7}\\
&= \sum_y P(Y = y) \sum_x x P(X = x | y = y) \tag{8}\\
&= \sum_y P(Y = y) \mathbb{E}[X | Y = y] \tag{9}\\
&= \mathbb{E}[\mathbb{E}[X | Y]] \tag{10}
\end{align}
$$

where in Equation (6) we used the law of total probability and Bayes rule in Eq. (7). ∎

Next, we will concretely instantiate the above framework in our analysis of Preferential Attachment.

1.2   Power Law degree distribution of PA graphs

Before starting the analysis we first compute the normalizing constant involved in the Preferential Attachment step of the growth process. In our model, at time $t \geq 1$ there are exactly t vertices and $t - 1$ directed edges. The exact probability of connecting to a node of degree k is:

$$
\frac{1}{Z_t} \sum_u \mathrm{d}_u = \frac{1}{Z_t} |E_t| = 1 \Rightarrow Z_t = |E_t| = (t - 1) \approx t
$$

We start now with the first step of the method:

Markovian Dynamics: Let $D_k(t)$ be the number of vertices of in-degree k of the graph G_t and consider the random vector $\mathbf{D}_t = (D_0(t), \ldots, D_{t-1}(t))$. A moment's thought reveals that: (i) $\frac{\mathbf{D}_t}{t}$ is the in-degree distribution of the graph G_t, (ii) the sequence $\{\mathbf{D}_t\}_{t \geq 0}$ is Markov with respect to the preferential attachment process.

Conditional Change: To calculate the expected change of $\mathbf{D}_{t+1}$ given $\mathbf{D}_t$ we focus on $D_k(t+1)$ the number of nodes with a particular degree k:

- **Increase:** $D_k(t + 1)$ can only increase by one if the new vertex inserted at time t connects to one vertex with degree $k - 1$ at time t. This happens with probability:

$$
p \frac{D_{k-1}(t)}{t} + q(k - 1) \frac{D_{k-1}(t)}{t} \tag{11}
$$

  as with probability p we would have to select one of $D_{k-1}(t)$ out of t vertices in the graph, and with probability q the total "weight" of the vertices in the preferential attachment scheme is given by $(k - 1)D_{k-1}(t)$.

- **Decrease:** respectively $D_k(t+1)$ can only decrease by one if the new edge inserted lands on one of the vertices of degree k. This happens with probability:

$$p\frac{D_k(t)}{t} + qk\frac{D_k(t)}{t} \tag{12}$$

Rate Equation: Define $d_k(t) = \mathbb{E}[D_k(t)]$ to be the expected number of vertices having in-degree k at time t. Adding Eq. (11), subtracting Eq. (12) and taking expectations, we obtain:

$$d_k(t+1) - d_k(t) = p\left(\frac{d_{k-1}(t) - d_k(t)}{t}\right) + q\left(\frac{(k-1)d_{k-1}(t) - kd_k(t)}{t}\right) \tag{13}$$

Fluid Limit: For $t \gg 1$, the left hand side approximates the derivative:

$$\frac{d}{dt}d_k(t) = p\left(\frac{d_{k-1}(t) - d_k(t)}{t}\right) + q\left(\frac{(k-1)d_{k-1}(t) - kd_k(t)}{t}\right) \tag{14}$$

The previous equation holds for all $k \geq 1$, for $k = 0$ we have:

$$\frac{d}{dt}d_0(t) = p\left(1 - \frac{d_0(t)}{t}\right) + q\left(1 - \frac{d_0(t)}{t}\right) \tag{15}$$

as the only way to increase the number of vertices of degree 1 is for the new vertex to connect to a vertex with degree greater than one.

Solution of ODE: To solve the differential equation we assume the following form for the solution $d_k(t) = p_k \cdot t$. Substituting in equations Eq. (14) and Eq. (15) we get:

$$(1 + p + kq)\, p_k = (p + (k-1)q)\, p_{k-1} \tag{16}$$

$$(1 + p + q)\, p_0 = 1 \tag{17}$$

$$(1 + p + 1 - p)\, p_0 = 1 \tag{18}$$

The above equations allow us to form a recurrence for p_k:

$$p_k = \left(\frac{p + (k-1)q}{1 + p + kq}\right)p_{k-1} \tag{19}$$

$$= \left(1 - \frac{1+q}{1+p+kq}\right)p_{k-1} \tag{20}$$

$$\approx \left(1 - \frac{(1+q)/q}{k}\right)p_{k-1} \tag{21}$$

$$\approx \left(\frac{k-1}{k}\right)^{\frac{1+q}{q}}p_{k-1} \tag{22}$$

From (18) we see that $p_0 = \frac{1}{2}$, so iterating (22) we get that:

$$p_k \approx k^{-\frac{1+q}{q}} = k^{-\frac{2-p}{1-p}} = k^{-1-\frac{1}{1-p}} \tag{23}$$

That is we have shown heuristically that the degree distribution of the preferential attachment model follows a power law with exponent $\alpha = \frac{2-p}{1-p} = 1 + \frac{1}{1-p}$.

2 Power Laws

Statistical modeling of data involves gathering enough data such that reasonable models can be built and from there informed inferences can be carried out. One can make a very broad distinction between different random variables as to the extent that they are close their expected behavior (value). People speak broadly of heavy-tailed phenomena where extremes are possible (even frequent) and light tailed phenomena where things are well behaved within limits. The prototypical example of heavy-tailed behavior are power-law distributed random variables.

Definition 3 *A random variable X is said to follow a power law with* exponent α *if for large* $x \gg 1$, $\mathbb{P}(X = x) \propto x^{-\alpha}$.

The past decades numerous empirical studies have identified power laws in almost every imaginable aspect of the world. In our course, power-laws will show up when looking at various statistics of networks, most predominantly the degree-distribution.

There are two flavors of power-laws one for continuous and one for discrete data. For the purposes of this note, we will focus almost exclusively on continuous random variables and simply allude to the fact that discrete data can be very accurately simulated by appropriately binning (rounding to the closer integer) the continuous data. We start with some calculations.

2.1 Normalizing Constant

A continuous random variable would have a density $p(x)dx = P(x \leq X \leq x + dx) = Cx^{-\alpha}$. Since the density diverges for $x \to 0$ we need to impose a lower bound x_{min} on the values of X. For all $\alpha > 1$ and given $x_{min} > 0$ the normalization constant $C(\alpha, x_{min})$ is given by:

$$\int_{x_{min}}^{\infty} p(x)dx = C \int_{x_{min}}^{\infty} x^{-\alpha}dx = C\frac{1}{\alpha - 1}\left[x^{-a+1}\right]_{\infty}^{x_{min}} = 1 \Rightarrow C(\alpha, x_{min}) = \frac{\alpha - 1}{(x_{min})^{\alpha-1}}$$

The final form of a continuous (pure) power law distribution is:

$$p(x) = \frac{\alpha - 1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\alpha} \tag{24}$$

For a discrete random variable following a power law $p(k) = Dk^{-\alpha}$ we get the following expression for the normalizing constant $D(\alpha, x_{min})$:

$$\sum_{x_{min}}^{\infty} p(k) = D \sum_{n=0}^{\infty} (x_{min} + n)^{-\alpha} = 1 \Rightarrow D(\alpha, x_{min}) = \frac{1}{\zeta(\alpha, x_{min})}$$

The function $\zeta(\alpha, x_{min}) = \sum_{n=0}^{\infty}(n + x_{min})^{-\alpha}$ is called the Hurwitz zeta function. The final form for a discrete (pure) power law distribution is:

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})} \tag{25}$$

2.2   Tail Distribution

Power laws and heavy-tailed distributions in general are typically contrasted to the Normal distribution. The pronounced difference is in their tail distribution.

Definition 4 *The tail distribution $T(x)$ of a random variable x is defined as $T(x) = \mathbb{P}(X \geq x)$.*

Definition 5 *A random variable is heavy-tailed if $\forall \lambda > 0, \lim_{x \to \infty} \left[\frac{T(x)}{e^{-\lambda x}} \right] = \infty$.*

From the definition it is easy to see that Normal distribution does not have heavy-tails whereas power laws do. The tail distribution for continuous and discrete power-law distributions are:

$$T_c(x) = \mathbb{P}(X \geq x) = \int_x^\infty \frac{\alpha - 1}{x_{min}} \left(\frac{z}{x_{min}} \right)^{-\alpha} dz = \left(\frac{x}{x_{min}} \right)^{-\alpha + 1}$$

$$T_d(x) = \mathbb{P}(X \geq x) = \sum_{z=x}^\infty \frac{z^{-\alpha}}{\zeta(\alpha, x_{min})} = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})}$$

2.3   Moments

One of the consequences of power-law behaviour is that depending on the exponent α certain moments diverge.

$$
\begin{align}
\mathbb{E}[X^k] &= \int_{x_{min}}^\infty p(x) x^k dx \tag{26} \\
&= \int_{x_{min}}^\infty \frac{\alpha - 1}{x_{min}^{1-\alpha}} x^{k-\alpha} dx \tag{27} \\
&= \begin{cases} \frac{\alpha - 1}{\alpha - k - 1} (x_{min})^k & k < \alpha - 1 \\ \infty & k \geq \alpha - 1 \end{cases} \tag{28}
\end{align}
$$

For instance if $\alpha \leq 3$ the variance diverges, whereas for $\alpha \leq 2$ the mean diverges.

2.4   Generation

For a continuous CDF, the function $F(x)$ is one-to-one. Using the transformation method we can generate power-law distributed data using only a uniform random variable:

$$
\begin{align}
U = 1 - \left(\frac{X}{x_{min}} \right)^{-\alpha + 1} \quad &\Rightarrow \quad \ln(1 - U) = (1 - \alpha)[\ln(X) - \ln(x_{min})] \tag{29} \\
&\Rightarrow \quad -\frac{1}{\alpha - 1} \ln(1 - U) + \ln(x_{min}) = \ln(X) \tag{30} \\
&\Rightarrow \quad X = x_{min} (1 - U)^{-\frac{1}{\alpha - 1}} \tag{31}
\end{align}
$$

2.5 MLE for the Scaling Parameter

Due to the peculiar statistical properties of power-law random variables, caution must be practiced when estimating the parameters. The best way to estimate the scaling exponent is to use Maximum Likelihood Estimation (MLE). We first derive an expression for the log-likelihood for n independent samples:

$$
\begin{aligned}
L(\alpha) &= \ln p(x_1, \ldots, x_n) & (32)\\
&= \ln \left[\prod_{i=1}^{n} p(x_i) \right] & (33)\\
&= \sum_{i=1}^{n} \ln[p(x_i)] & (34)\\
&= \sum_{i=1}^{n} \left[\ln(\alpha - 1) - \ln(x_{min}) - \alpha \ln \left(\frac{x_i}{x_{min}} \right) \right] & (35)\\
&= n[\ln(\alpha - 1) - \ln(x_{min})] - \alpha \sum_{i=1}^{n} \ln \left(\frac{x_i}{x_{min}} \right) & (36)
\end{aligned}
$$

Next we are going to find the value $\hat{a}$ that maximizes the likelihood:

$$
\left. \frac{\partial L(\alpha)}{\partial \alpha} \right|_{\alpha = \hat{\alpha}} = 0 \quad \Leftrightarrow \quad \frac{n}{\hat{\alpha} - 1} - \sum_{i=1}^{n} \ln \left(\frac{x_i}{x_{min}} \right) = 0 \tag{37}
$$

$$
\Leftrightarrow \quad \hat{\alpha} = 1 + n \left[\sum_{i=1}^{n} \ln \left(\frac{x_i}{x_{min}} \right) \right]^{-1} \tag{38}
$$