Adventures in random graphs: Models, structures and algorithms

Armand M. Makowski

ECE & ISR/HyNet
University of Maryland at College Park
armand@isr.umd.edu

LECTURE 4
Scale-free networks

Degrees revisited!

Consider a directed graph $G = (V, E)$ with $V = \{1, \ldots, n\}$ with link assignments $\{\chi_{ij},\ 1 \leq i, j \leq n\}$.

The **out-degree** of node i:

$$d_i^+ := \sum_{j=1,\ j \neq i}^{n} \chi_{ij}$$

The **in-degree** of node i:

$$d_i^- := \sum_{j=1,\ j \neq i}^{n} \chi_{ji}$$

Basic identity

$$\sum_{i=1}^{n} d_i^+ = \sum_{i=1}^{n} d_i^+ = |E|$$

For undirected graphs

For **undirected** graphs, the **degree** of node i is simply

$$d_i = d_i^+ = d_i^-, \quad i \in V$$

Basic identity

$$\sum_{i=1}^{n} d_i = |E|$$

or

$$\sum_{i=1}^{n} d_i = 2 \cdot \text{Total number of } \textbf{undirected} \text{ edges in } G$$

Empirical degree distribution in $G = (V, E)$

$$
\begin{aligned}
P(d) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left[d_i = d\right] \\
&= \frac{|\{i = 1, \ldots, n: \ d_i = d\}|}{n}, \quad d = 0, 1, \ldots, n-1
\end{aligned}
$$

so that

$$
d_{\text{Avg}} = \sum_{d=0}^{n-1} d \cdot P(d) = \frac{1}{n} \sum_{i=1}^{n} d_i
$$

No random graph model assumed – Data driven

The random node viewpoint

With ν denoting a rv **uniformly** distributed over the index set $\{1, \ldots, n\}$, define the **rv** given by

$$D = d_\nu$$

It is plain that

$$\begin{aligned}
\mathbb{P}\left[D = d\right] &= \sum_{i=1}^{n} \mathbb{P}\left[d_i = d,\ \nu = i\right] \\
&= \sum_{i=1}^{n} \mathbf{1}\left[d_i = d\right] \mathbb{P}\left[\nu = i\right] = P(d)
\end{aligned}$$

Random sampling interpretation of $P(\cdot)$!

The random edge viewpoint

Let ε be a rv **uniformly** distributed over the edge set E of V, and write $\varepsilon = (i, j)$ with $1 \leq i \neq j \leq n$. Next, **independently** of the outcome, select either node with probability 1/2, and denote the resulting node by μ. Set

$$D^\star = d_\mu$$

Can we compute the pmf

$$P^\star(d) = \mathbb{P}\left[D^\star = d\right], \quad d = 0, \ldots, n-1?$$

For each $d = 0, 1, \ldots, n-1$,

$$\begin{aligned}
\mathbb{P}[D^\star = d] &= \mathbb{P}[d_\mu = d] \\
&= \sum_{k=1}^{n} \mathbb{P}[\mu = k,\ d_\mu = d] \\
&= \sum_{k=1}^{n} \mathbb{P}[\mu = k,\ d_k = d] \\
&= \sum_{k=1}^{n} \mathbf{1}[d_k = d]\, \mathbb{P}[\mu = k]
\end{aligned}$$

By construction,

$$\mathbb{P}\left[\mu = k\right] = \frac{\frac{1}{2}d_+(k) + \frac{1}{2}d_-(k)}{|E|} = \frac{d(k)}{|E|}$$

so that

$$\begin{aligned}\mathbb{P}\left[D^\star = d\right] &= \sum_{k=1}^{n} \mathbf{1}\left[d_k = d\right] \frac{d(k)}{|E|} \\ &= \frac{d}{|E|} \cdot \sum_{k=1}^{n} \mathbf{1}\left[d_k = d\right] \\ &= \frac{d}{|E|} \cdot nP(d)\end{aligned}$$

$$P^\star(d) = \frac{n}{|E|} \cdot d\ P(d), \quad d = 0, 1, \ldots, n-1$$

Note that

$$\sum_{d=0}^{n-1} P^{\star}(d) = 1$$

since

$$\begin{aligned}
\sum_{d=0}^{n-1} d\ P(d) &= \sum_{d=0}^{n-1} d \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left[d_i = d\right] \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{d=0}^{n-1} d \cdot \mathbf{1}\left[d_i = d\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} d_i \\
&= \frac{|E|}{n}
\end{aligned}$$

Arbitrary degree distributions?

Degrees in undirected graphs

Is it always the case that non-negative integers $(d_i,\ i \in V)$ are the degree sequence of a graph $G = (V, E)$?

Erdős and Gallai (1960): The positive integers $d_1, \ldots, d_n$ are the degrees of a simple graph with n vertices if and only if

$$d_1 + \ldots + d_n \quad \text{even}$$

and

$$\sum_{j=1}^{k} d_{(j)} \leq k(k-1) + \sum_{\ell=k+1}^{n} \min\left(k, d_{(\ell)}\right), \quad k = 1, \ldots, n$$

with $d_{(1)}, \ldots, d_{(n)}$ denoting $d_1, \ldots, d_n$ in decreasing order.

A simple fact

Theorem 1 *With i.i.d.. $\mathbb{N}_0$-valued rvs $\{D, D_n,\ n = 1, 2, \ldots\}$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left[D_1 + \ldots + D_n \text{ even }\right] = \frac{1}{2}$$

provided

$$0 < \mathbb{P}\left[D \text{ even }\right] < 1$$

Proof: For each $n = 1, 2, \ldots$, note that

$$\mathbb{E}\left[(-1)^{D_1+\ldots+D_n}\right] = \mathbb{E}\left[(-1)^{D}\right]^n = (2\mathbb{P}\left[\text{D even }\right] - 1)^n$$

while

$$\mathbb{E}\left[(-1)^{D_1+\ldots+D_n}\right] = 2\mathbb{P}\left[D_1 + \ldots + D_n \text{ even }\right] - 1$$

In many deployed networks, data analysis shows a Pareto-like behavior, e.g.,

$$P(d) \sim d^{-\gamma}, \quad d = 1, \ldots, n-1$$

for some $\gamma > 2$.

Why generating models that capture this property?

- From local rules to global properties!
- Impact on the behavior of dynamic processes on such networks

---

- Growth models
- Optimization-based models
- Threshold random graphs
- !!!

Growth models

What are growth models?

Sequence of **random** graph models

$$\{\mathbb{G}_t,\ t = 0, 1, \ldots\}$$

with rules/mechanisms

$$V_{t+1} \leftarrow V_t$$

and

$$\mathbb{G}_{t+1} \leftarrow (\mathbb{G}_t, V_{t+1})$$

Usually **Markov**-like random mechanisms with the requirements that

$$V_t \subseteq V_{t+1}, \quad t = 0, 1, \ldots$$

and

$$\mathbb{G}_t = \text{Subgraph of } \mathbb{G}_{t+1}, \quad t = 0, 1, \ldots$$

Two basic rules

- Preferential attachment – "The rich get richer"
- Copying

Very long history with many more models recently proposed:

- Yule (1925)
- Barabási and Albert (1999)
- Kleinberg et al.
-

Key features

- Nodes are non-homogeneous – Infinite node population
- Yet not a dynamical network model
- No node deletion and no edge deletion!
- Interest in asymptotic behavior of the sample degree distributions

$$\frac{1}{|V_t|}\sum_{i\in V_t}\mathbf{1}\left[D_t(i)=d\right], \qquad \begin{array}{l} d=0,1,\ldots \\ t=0,1,\ldots \end{array}$$

where

$$D_t(i) = \text{Degree of node } i \text{ in } \mathbb{G}_t$$

The BA model (Simplified version)

We are given

- A connected graph $G = (V, E)$ with
$$V = \{-(K-1), \ldots, -1, 0\}$$
for some integer K in $\mathbb{N}$
- An **attachment** function, i.e., a monotone increasing mapping $\varphi : \mathbb{N}_0 \to \mathbb{R}_+$ with $\varphi(u) = 0$ only if $u = 0$, i.e.,
$$\varphi(u) = u^\nu \quad (\nu > 0) \quad \text{or} \quad \varphi(u) = e^{\lambda u} \quad (\lambda > 0)$$

Recursively

For $t = 0$:

$$\mathbb{G}_0 = G$$

For $t = 0, 1, \ldots$: With $\mathbb{G}_t = (V_t, \mathbb{E}_t)$ available, we construct $\mathbb{G}_{t+1} = (V_{t+1}, \mathbb{E}_{t+1})$ as follows:

- A new **node**, labelled $t + 1$, arrives so that

$$V_{t+1} = V_t \cup \{t + 1\}$$

- A new **edge** is created, linking node $t + 1$ to one of the existing nodes, say X_{t+1}, in V_t so that

$$\mathbb{E}_{t+1} = \mathbb{E}_t \cup \{(X_{t+1}, t + 1)\}$$

The preferential attachment rule

The nodes $\{X_{t+1},\ t = 0, 1, \ldots\}$ are randomly generated as follows:
At time $t + 1$,

$$\mathbb{P}\left[X_{t+1} = x | X_1, \ldots, X_t\right] = \frac{\varphi(D_t(x))}{\sum_{y \in V_t} \varphi(D_t(y))}, \quad x \in V_t$$

with

$$D_t(i) = \text{Degree of node } i \text{ in } \mathbb{G}_t, \quad i \in V_t$$

Note that

$$D_t(i) > 0, \qquad \begin{array}{c} i = -(K-1), \ldots, t \\ t = 0, 1, \ldots \end{array}$$

A basic recursion

For each $t = 0, 1, \ldots,$

$$D_{t+1}(i) = \begin{cases} D_t(i) + \mathbf{1}\left[X_{t+1} = i\right] & \text{if } i \in V_t \\ 1 & \text{if } i = t+1 \end{cases}$$

with

$$D_0(i) = \text{Degree of node } i \text{ in } G, \quad i \in V_0$$

Obviously, by construction, we have

$$|\mathbb{E}_{t+1}| = |\mathbb{E}_t| + 2, \quad t = 0, 1, \ldots$$

so that

$$|\mathbb{E}_t| = |E_0| + 2t, \quad t = 0, 1, \ldots$$

Hence

$$\sum_{i \in V_t} D_t(i) = |E_0| + 2t, \quad t = 0, 1, \ldots$$

The linear case ($\varphi(u) = u,\ u \geq 0$)

The nodes $\{X_{t+1},\ t = 0, 1, \ldots\}$ are now randomly generated as follows: At time $t + 1$,

$$\begin{aligned}\mathbb{P}\left[X_{t+1} = x | X_1, \ldots, X_t\right] &= \frac{D_t(x)}{\sum_{y \in V_t} D_t(y)} \\ &= \frac{D_t(x)}{|E_0| + 2t}, \quad x \in V_t\end{aligned}$$

For $t = 0, 1, \ldots$, define the counting rvs

$$N_t(d) = \sum_{x \in V_{t+1}} \mathbf{1}\left[D_t(x) = d\right], \quad d = 1, 2, \ldots$$

Note that $N_t(0) = 0$.

For $d = 1$:

$$N_{t+1}(1) = \begin{cases} N_t(1) & \text{if } D_t(X_{t+1}) = 1 \\ N_t(1) + 1 & \text{if } D_t(X_{t+1}) > 1 \end{cases}$$

so that

$$N_{t+1}(1) = N_t(1) + \mathbf{1}\left[D_t(X_{t+1}) \neq 1\right]$$

$$\begin{aligned}
\mathbb{E}\left[N_{t+1}(1)|\mathcal{F}_t\right] &= N_t(1) + \mathbb{P}\left[D_t(X_{t+1}) \neq 1|\mathcal{F}_t\right] \\
&= N_t(1) + 1 - \mathbb{P}\left[D_t(X_{t+1}) = 1|\mathcal{F}_t\right] \\
&= N_t(1) + 1 - \sum_{x \in V_t} \mathbf{1}\left[D_t(x) = 1\right] \frac{D_t(x)}{|E_0| + 2t} \\
&= N_t(1) + 1 - \sum_{x \in V_t} \mathbf{1}\left[D_t(x) = 1\right] \frac{1}{|E_0| + 2t} \\
&= N_t(1) + 1 - \frac{N_t(1)}{|E_0| + 2t} \\
&= N_t(1) \cdot \left(1 - \frac{1}{|E_0| + 2t}\right) + 1
\end{aligned}$$

$$\mathbb{E}\left[N_{t+1}(1)\right] = \mathbb{E}\left[N_t(1)\right] \cdot \left(1 - \frac{1}{|E_0| + 2t}\right) + 1$$

For $d > 1$:

$$N_{t+1}(d) = N_t(d) + \mathbf{1}\left[D_t(X_{t+1}) = d-1\right] - \mathbf{1}\left[D_t(X_{t+1}) = d\right]$$

$$\begin{aligned}
&\mathbb{E}\left[N_{t+1}(d)|\mathcal{F}_t\right] \\
&\quad= N_t(d) + \mathbb{P}\left[D_t(X_{t+1}) = d-1|\mathcal{F}_t\right] - \mathbb{P}\left[D_t(X_{t+1}) = d|\mathcal{F}_t\right] \\
&\quad= N_t(d) + \sum_{x \in V_t} \mathbf{1}\left[D_t(x) = d-1\right] \mathbb{P}\left[X_{t+1} = x|\mathcal{F}_t\right] \\
&\qquad\quad - \sum_{x \in V_t} \mathbf{1}\left[D_t(x) = d\right] \mathbb{P}\left[X_{t+1} = x|\mathcal{F}_t\right] \\
&\quad= N_t(d) + \sum_{x \in V_t} \mathbf{1}\left[D_t(x) = d-1\right] \frac{d-1}{|E_0| + 2t}
\end{aligned}$$

$$
\begin{aligned}
&\quad - \sum_{x \in V_t} \mathbf{1}\left[D_t(x) = d\right] \frac{d}{|E_0| + 2t} \\
&= N_t(d) + \frac{d-1}{|E_0|+2t} \cdot N_t(d-1) - \frac{d}{|E_0|+2t} \cdot N_t(d) \\
&= N_t(d)\left(1 - \frac{d}{|E_0|+2t}\right) + \frac{d-1}{|E_0|+2t} \cdot N_t(d-1)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left[N_{t+1}(d)\right] \\
&\quad = \mathbb{E}\left[N_t(d)\right]\left(1 - \frac{d}{|E_0|+2t}\right) + \frac{d-1}{|E_0|+2t} \cdot \mathbb{E}\left[N_t(d-1)\right]
\end{aligned}
$$

Theorem 2 *The limits*

$$\lim_{t\to\infty} \frac{\mathbb{E}\,[N_t(d)]}{|V_t|} = \pi(d), \quad d = 1, 2, \ldots$$

all exist and satisfy

$$\pi(d) = \frac{d-1}{d+2}\pi(d-1), \quad d = 2, 3, \ldots$$

with $\pi(1) = \frac{2}{3}$.

Easy to see that

$$\pi(d) \sim \frac{4}{d^3}$$

A stronger results is available

Theorem 3 *For each $d = 1, 2, \ldots$, we have*

$$\lim_{t\to\infty} \frac{N_t(d)}{|V_t|} = \pi(d) \quad a.s.$$

with $\pi(d)$ as given earlier.

More general models (I)

With α in $[0, 1)$, the **new** node labeled $t+1$ is attached to one of the existing nodes, say X_{t+1} in V_t,

- With probability α, **uniformly** amongst the nodes in V_t
- With probability $1-\alpha$, according to a **preferential** attachment mechanism

---

With a linear attachment function, Theorems 2 and 3 still hold with

$$\pi(d) \sim d^{-\gamma} \quad \text{with} \quad \gamma = \frac{3-\alpha}{1-\alpha}$$

More general models (II)

Add m new node at each time $t = 0, 1, \ldots$ with $m \geq 1$