Networks: Lectures 5 & 6
Toy Models of Network Formation

Heather A Harrington

Mathematical Institute
University of Oxford

HT 2017

What you're in for

- Week 1: Introduction and basic concepts
- Week 2: Small worlds
- **Week 3: Toy models of network formation**
- Week 4: Additional summary statistics and other concepts
- Week 5: Random graphs
- Week 6: Community structure and other mesocopic structures
- Week 7: Dynamical systems on networks
- Week 8: Other topics TBD

Contents

Motivation

- In addition to direct measurement of network properties (we've seen some of this and will see more), it is important to study networks using *generative models* (we've seen a bit of this in week 2 as well).
- We'll focus on *preferential attachment* (PA) models as an example.
- PA is a "rich-get-richer" scheme (e.g., highly cited papers are easier to find via Google, which in turn leads them to get even more citations).

Attachment kernel

- An *attachment kernel* gives a rule for how new objects in a network connect to existing objects.
- Suppose that we add one node at a time and that it forms an edge with a single existing node (there must also be a seed network in place or an initial condition).

Attachment kernel

New node connects to existing node i with probability

$$q_i = \frac{a_i}{\sum a_i};$$

a_i is the attachment kernel.

If q_i depends directly only on the properties of node i, we write: $a_i = a_i(\boldsymbol{\alpha_i}, \boldsymbol{\eta_i})$, where

- $\boldsymbol{\alpha_i}$ = vector of structural properties (e.g., degree, local clustering coefficient, etc).
- $\boldsymbol{\eta_i}$ = vector of other properties (e.g., a fitness value assigned to node i to represent different attributes).

Attachment kernel

More generally, could consider beyond node i (e.g., nearest neighbors) or exogenous to the network: $a_i = a_i(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{X})$ where

- $\boldsymbol{\alpha_{ij}}$ = matrix of structural properties being considered
- $\boldsymbol{\eta_{ij}}$ = matrix of other (e.g., "fitness") properties
- $\boldsymbol{X_{ij}}$ = matrix of exogenous properties

One could make even more complicated kernels, by including time-dependence. Often such generality is not explored (much simpler situations can exhibit rich behavior).

Note: There has been work on attachment mechanisms with > 1 edge added at a time, rewiring existing nodes, removing edges etc.

Seeds

To define a toy attachment model, we also need to specify a seed network or an initial condition.

Q: What happens to the "memory" of a seed as the size N of a network increases?

Linear preferential attachment

Consider an unweighted, undirected network (without self-edges or multi-edges).

Linear preferential attachment by degree has $a_i = k_i$.

$$\Rightarrow q_i = \frac{k_i}{\sum_i k_i} = \frac{k_i}{zN}$$

where $z = \langle k \rangle$ is mean degree.

Can write versions of this for directed networks by considering in- and out-degree separately.

de Solla Price's model

Concerned with citation network, so need to consider a directed network (new paper cites papers that already exist)

Network set-up.

- node: paper
- edge: a citation (directed edge from newer paper to older one).
- nodes cannot be removed, so this model gives DAGs.
- let $c :=$ mean number of papers cited by new paper (i.e., mean out degree).

de Solla Price's model

In Price's model, we add a node at every time-step. Each new node has on average c outgoing edges.

In this model,

$$q_i = \frac{k_i + a}{\sum_i (k_i + a)},$$

where this q_i is an affine preferential attachment, and the new bonus parameter $a > 0$ allows papers without citations to eventually be cited.

Note: a paper can cite another paper multiple times in this model.

de Solla Price's model

Let's write down equations that govern the in-degree (i.e., number of citations) of nodes in terms of c and a. Let $k_i = k_i^{in}$ denote the in-degree of node i.

Let $p_k(N) :=$ fraction of nodes with in-deg k when network has N nodes.

Each existing node i attracts an incoming node with probability proportional to its in-degree k_i plus a, an initial fitness:

$$q_i = \frac{k_i + a}{\sum_i (k_i + a)} = \frac{k_i + a}{Na + N\langle k \rangle} = \frac{k_i + a}{N(a + c)},$$

where $z = \frac{1}{N} \sum_i k_i = \langle k \rangle$. Since each new paper cites c papers on average, the expected number of new citations to node i when a new node is added is $c \left(\frac{a+k_i}{N(a+c)} \right)$.

de Solla Price's model

$Np_k(N)$ nodes have in-degree k, so the expected number new citations (i.e., incoming edges) to all nodes with in-degree k is

$$Np_k(N) \times c \times \frac{k+a}{N(c+a)} = p_k(N)\frac{c(k+a)}{c+a}. \tag{1}$$

To study dynamics of Eq. 1, we write a *master equation* for the evolution of the in-degree.

When adding exactly one node to the $N-$node network, the number of nodes with in-degree k increases by 1 for every node that previously had degree $k-1$ and receives a new citation. The expected number of such nodes is

$$\frac{c(k-1+a)}{c+a}p_{k-1}(N)$$

Similar terms for nodes that go from in-degree k to in-degree $k+1$ (i.e. Eq. 1).

Preferential attachment

The evolution of the number of nodes with degree $k > 0$ is given by the discrete-time dynamical process:

$$(N+1)p_k(N+1) = \underbrace{Np_k(N)}_{[prev.\ \#k]} + \underbrace{p_{k-1}(N)\frac{c(k-1+a)}{c+a}}_{\#[(k-1)\to k]} - \underbrace{p_k(N)\frac{c(k+a)}{c+a}}_{\#[(k)\to k+1]}.$$

This is the expected number of nodes with in-degree k when the network has $N+1$ total nodes.

Note: the maximum degree obtainable at time N is $c(N-1)$.

When $k = 0$ we have:

$$(N+1)p_0(N+1) = Np_0(N) + 1 - p_0(N)\frac{ca}{c+a}.$$

Note: can't go from $-1 \to 0$ citations, so the corresponding term isn't present for $k = 0$ case. The new node automatically has in-degree 0.

Preferential attachment

A master equation is an equation of evolution of the following form:

new # things = old # things + ways to add more things - ways to remove things,

for each state, e.g., a discrete set as above.

Consider $N \to \infty$ to calculate an asymptotic form for the degree distribution (abuse of notation $p_k = p_k(\infty)$).

For every N we can obtain $p_k(N)$, given $p_k(0)$ by iteratively solving N equations.

There are also continuous-time versions of these equations, which are often convenient to use.

In the large-graph case (i.e., as $N \to \infty$) we assume that $p_k(N) \to p_k$ which gives

$$p_0 = 1 - p_0 \frac{ca}{c+a}, \quad k = 0$$

$$p_k = \frac{c}{c+a}\left[(k-1+a)p_{k-1} - (k+a)p_k\right], \quad k \geq 1.$$

In general, we can't necessarily solve such an equation (might need to do numerical simulations of Eq. 1. Here, however we can rearrange and calculate the solutions:

$$p_0 = \frac{1 + \frac{a}{c}}{a + 1 + \frac{a}{c}},$$

$$p_k = \frac{k + a + 1}{k + a + 1 + \frac{a}{c}} p_{k-1}.$$

Solving iteratively for p_k we get:

$$p_k = \frac{\prod_{i=1}^{k}(k+a-i)}{\prod_{i=1}^{k}(i+1+a+\frac{a}{c})}\frac{1+\frac{a}{c}}{a+1+\frac{a}{c}} \quad = \left(1+\frac{a}{c}\right)\frac{\prod_{i=1}^{k}(k+a-i)}{\prod_{i=0}^{k}(i+1+a+\frac{a}{c})}.$$

Recall the *gamma function* [1]: $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$ with properties:

$$\Gamma(x+1) = x\Gamma(x) \quad \forall x > 0 \quad (\Rightarrow \Gamma(x) = (x-1)! \text{ for } x \in \mathbb{N}.)$$

$\frac{\Gamma(x+n)}{\Gamma(x)} = (x+n-1)(x+n-2)\ldots x,$ (Pochhammer's function).
Using Pochhammer's function, we can re-write terms of p_k:

$$\prod_{i=1}^{k}(k+a-i) = \frac{\Gamma(k+a)}{\Gamma(a)}, \quad \prod_{i=0}^{k}(i+1+a+\frac{a}{c}) = \frac{\Gamma(1+\frac{a}{c}+a+(k+1))}{\Gamma(1+\frac{a}{c}+a)},$$

$$\Rightarrow p_k = \left(1+\frac{a}{c}\right)\frac{\Gamma(k+a)\Gamma(a+1+a/c)}{\Gamma(a)\Gamma(k+a+2+a/c)}$$

[1] To see more properties of the gamma function see `http://dlmf.nist.gov/`

Preferential attachment

Another special function is the *beta function*

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Using $\Gamma(x+1) = x\Gamma(x)$ property, we let $x = 1 + \frac{a}{c}$, which gives

$$\Rightarrow \Gamma\left(2 + \frac{a}{c}\right) = \left(1 + \frac{a}{c}\right)\Gamma\left(1 + \frac{a}{c}\right).$$

Multiply p_k by this expression on the numerator and denominator:

$$\begin{aligned} p_k &= \cancel{\left(1 + \frac{a}{c}\right)} \frac{\Gamma(k+a)\Gamma(a+1+a/c)}{\Gamma(a)\Gamma(k+a+2+a/c)} \cdot \frac{\Gamma\left(2 + \frac{a}{c}\right)}{\cancel{\left(1+\frac{a}{c}\right)}\Gamma\left(1 + \frac{a}{c}\right)}, \\ &= \frac{\Gamma(k+a)\Gamma\left(2 + \frac{a}{c}\right)}{\Gamma(k+a+2+a/c)} \cdot \frac{\Gamma(a+1+a/c)}{\Gamma(a)\Gamma\left(1 + \frac{a}{c}\right)}, \\ &= \frac{B(k+a, 2+a/c)}{B(a, 1+a/c)}. \end{aligned}$$

Fix a and c, what happens as k gets large? We focus on numerator of p_k.

To better understand what $B(k+a, 2+a/c)$ looks like, we can use *Stirling's formula* $\Gamma(x) \sim \sqrt{2\pi} e^{-x} x^{x-\frac{1}{2}}$, $x \to \infty$ which applied to the expression above gives

$$B(x,y) \sim \frac{e^{-x} x^{x-(1/2)}}{e^{-(x+y)}(x+y)^{x+y-(1/2)}} \Gamma(y), x \to \infty.$$

Now we use

$$\begin{aligned}(x+y)^{x+y-(1/2)} &= x^{x+y-(1/2)} \left[1+\frac{y}{x}\right]^{x+y-(1/2)} \\ &\sim x^{x+y-(1/2)} e^{y}, \quad \text{as } x \to \infty\end{aligned}$$

$$B(x,y) \sim \frac{e^{-x} x^{x-(1/2)}}{e^{-(x+y)}(x+y)^{x+y-(1/2)}} \Gamma(y) = \Gamma(y) x^{-y} \quad \text{as } x \to \infty.$$

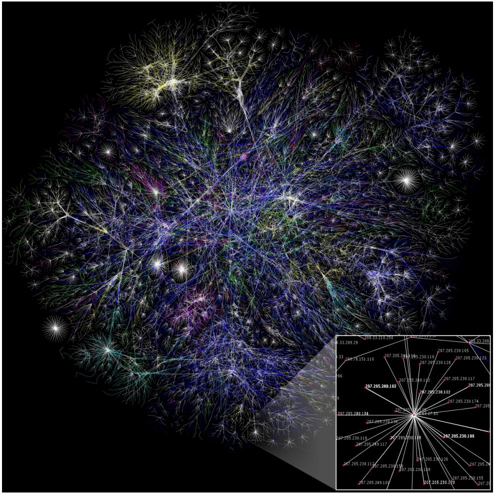
$\therefore$ beta function has a power-law decay for large x.
$\therefore$ in-degree distribution has power-law tail

$$p_k \sim (k+a)^{-\beta} \Rightarrow p_k \sim k^{-\beta} \text{ for } k \gg a.$$

The power is $\beta = 2 + \frac{a}{c} > 2$.

Example

The Internet is thought to grow following preferential attachment:



Barabási-Albert model

A special case of the Price model is known as the Barabási-Albert (BA) model.

$$p_k \sim k^{-(2+\frac{a}{c})}$$

In this model, $a = c \Rightarrow p_k \sim k^{-3}$

It is also has undirected edges rather than directed edges.

Generalizations of PA

We can also study the in-degree distribution as a function of time (not just the infinite-time limit).

Let $p_k(t, N)$ = mean fraction of nodes created at time t and which have in-degree k when the total number of nodes is N.

We are in event time, so first node is created at $t = 1$ and most recent nodes are created at $t = N$.

We'll use a master equation to study the dynamics of $p_k(t, N)$.

Generalizations of PA

When we add a new node, the expected number of new edges acquired by existing nodes with in-degree k is:

$$N p_k(t,N) \times c \times \frac{a+k}{N(a+c)} = \frac{c(a+k)}{a+c} p_k(t,N),$$

which is independent of the creation time of those nodes. We obtain the following master equation (for $k \geq 1$ and $k = 0$, respectively):

$$(N+1)p_k(t,N+1) = Np_k(t,N) + p_{k-1}(t,N)\frac{c(k-1+a)}{c+a} - p_k(t,N)\frac{c(k+a)}{c+a},$$

$$(N+1)p_0(t,N+1) = Np_k(t,N) + \delta_{t,N+1} - \frac{ca}{c+a}p_0(t,N),$$

where $\delta_{t,N+1} = \begin{cases} 1, t = N+1 \text{ i.e., a new in-degree 0 node is created,} \\ 0, \text{else.} \end{cases}$

There is a problem as $N \to \infty$ because fraction of nodes created at time t becomes 0 (only one node is created at a given time).

Generalizations of PA

We rescale time using $\tau := \frac{t}{N} \in [0, 1]$.

$\therefore$ oldest nodes created at $\tau = 0$ and newest nodes created at $\tau = 1$

Since we've rescaled time, we also need to change from $p_k(t, N)$ to a function $\pi_k(\tau, N)$ such that $\pi_k(\tau, N)d\tau$ is the fraction of nodes that have in-degree k that fall in the interval $[\tau, \tau + d\tau]$.

The number of nodes in that interval $d\tau$ is $Nd\tau \implies \pi_k d\tau = p_k \times Nd\tau$, giving

$$\pi_k(\tau, N) = Np_k(t, N)$$

Note: π_k does not vanish as $N \to \infty$. Also τ is not longer constant for a given node.

Generalizations of PA

The master equation becomes:

$$\pi_k\left(\frac{N}{N+1}\tau, N+1\right) = \pi_k(\tau, N) + \frac{\pi_{k-1}(\tau, N)}{N}\frac{c(k-1+a)}{c+a} - \frac{\pi_k(\tau, N)}{N}\frac{c(k+a)}{c+a}.$$

Taking $N \to \infty$ and abusing notation ($\pi_k(\tau) = \tau_k(\tau, \infty)$) and defining $\epsilon := \frac{1}{N}$ yields:

$$\frac{\pi_k(\tau) - \pi_k(\tau - \epsilon\tau)}{\epsilon} + \frac{c}{a+c}\left[(a+k-1)\pi_{k-1}(\tau) - (a+k)\pi_k(\tau)\right] + \mathcal{O}(\epsilon^2) = 0.$$

As $N \to \infty$, so LHS $\to \tau\frac{d\pi_k}{d\tau}$

$$\tau\frac{d\pi_k}{d\tau} + \frac{c}{a+c}\left[(a+k-1)\pi_{k-1}(\tau) - (a+k)\pi_k(\tau)\right] + \mathcal{O}(\epsilon^2) = 0, \;\; k \geq 1, \quad (2)$$

$$\tau\frac{d\pi_0}{d\tau} - \frac{ca}{a+c}\pi_0(\tau) = 0, \;\; k = 0.$$

Notice Eqs. 2 is valid for $\tau = [0, 1)$. The boundary conditions are the final conditions at $\tau = 1$. These are $\pi_0(1) = 1$ and $\pi_k(1) = 0$ for $k \geq 1$.

Generalizations of PA

Want to solve the $k = 0$ equation

$$\tau \frac{d\pi_0}{d\tau} - \frac{ca}{a+c}\pi_0(\tau) = 0,\ \ k = 0,$$

in closed form and ideally we use that to get solutions for $k \geq 1$ iteratively:

$$\pi_0(\tau) = \tau^{ca/(a+c)}$$

$\Rightarrow \pi_k(\tau) = \frac{1}{k!}\left[a(a+1)\cdots(a+k-1)\right]\tau^{ca/(a+c)}\left(1-\tau^{c/(a+c)}\right)^k$

Again using $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ and $B(x, y) \sim x^{-y}$ as $x \to \infty$ gives:

$$\pi_k(\tau) \sim k^{a-1}(1-\tau^{c/(a+c)})^k \text{ as } k \to \infty.$$

Generalizations of PA

$$\pi_k(\tau) \sim k^{a-1}(1-\tau^{c/(a+c)})^k \text{ as } k \to \infty.$$

We see there is exponential decay (for $|\tau| < 1$ implies the second term < 1) after some growth from the term k^{a-1}. Over all τ we see power-law behavior as before.

Exponential decay is slower for small τ, which means older nodes are more likely to have high in-degree than younger nodes.

Q: Is this true in general (e.g. Yahoo is older than Google in WWW)?

$\therefore$ if we wanted to use this as a model of structures of the WWW, we're clearly missing ingredients (an important one is node fitness)!

Edge removal

Consider an example in which undirected edges are added preferentially as in the BA model (number of new connections for each new node is exactly $c = a \in \mathbb{Z}_+$).

Then the degree of each node in BA mechanism is exactly c (not the average), so degree of each node is $k_i = k_i + c$ where $k_i =$ in-degree from the de Solla Price model (note change of notation).

Also, edges are *removed* uniformly at random during update steps. Probability that one of the removed edge's stubs is attached to node i is then proportional to k_i.

$$\text{prob(node } i \text{ looses an edge)} = \frac{2k_i}{\sum_i k_i}$$

(factor of 2 because each edge has 2 stubs)

Edge removal

Suppose at each step, degree c nodes are added to the network and a mean of v edges are removed for each node that is added (assume $v < c$).

On average, the network has $N(c - v)$ edges (for N nodes).

One can derive the master equation (for $k \geq 1$):

$$(N+1)p_k(N+1) = Np_k(N) + \delta_{kc} + p_{k-1}(N)\frac{c(k-1)}{2(c-v)} + p_{k+1}(N)\frac{2v(k+1)}{2(c-v)} - p_k(N)\frac{(c+2v)k}{2(c-v)},$$

For $k = 0$, the term proportional to $(k-1)$ isn't there (because no nodes with degree -1), all other are present.

Take $N \to \infty$ to study asymptotic properties of degree distribution as usual.

Nodes with varying quality

It is desirable to develop models in which different nodes have inherently different fitnesses (more than structural features).

e.g., *Bianconi-Barabási (BB) model*

As in BA model, add nodes one at a time via undirected edges to c existing nodes.

Node i has scalar fitness $\eta_i \in \mathbb{R}$ assigned when node is created (drawn from a continuous probability distribution $p(\eta)$.

The c new stubs each attach to an existing node with probability $\propto a(k, \eta)$ (e.g., $a(k, \eta) = \eta k$).

Network rewiring in bipartite network

So far, we've considered network growth (structural and fitness properties), now rewiring.

e.g., *Evans-Plato model*

This is a bipartite network with U people and M artifacts.

At time t, each of U people is connected via an undirected, unweighted edge to one of M artifacts (e.g., renting movies).

One can study this kind of setup with a rewiring rule for changing edges.

Network optimization

Consider network structures that arise through some optimization process. One constructs some sort of *cost function* and see what network structures minimize it.

e.g., Gastner-Newman model for transportation networks: Let t_{ij} = time to get from city i to city j.

$$= b_1 + b_2 r_{ij}$$

where b_1 is total time spent in airports, b_2 is total time traveled (proportional to distance), and r_{ij} is geographic distance between i and j.

Location of nodes can come from probability distribution (e.g., placed via some random process in square $[0,1] \times [0,1]$) or from empirical data (e.g., directly from a map).

Network optimization

Cost function is $E(m, L) = \lambda m + (1 - \lambda)L, \ \lambda \in [0, 1]$, where L is the mean geodesic distance between node pairs given by time travel t_{ij} and m is the number of edges.

Results (briefly):

- small b_1, large b_2: networks "close" to planar (like a road network),
- large b_1, small b_2: hub-and-spoke networks (like airline network) with many low-degree nodes and few high-degree nodes.

Q: How does the Gastner-Newman model behave for different values of λ?