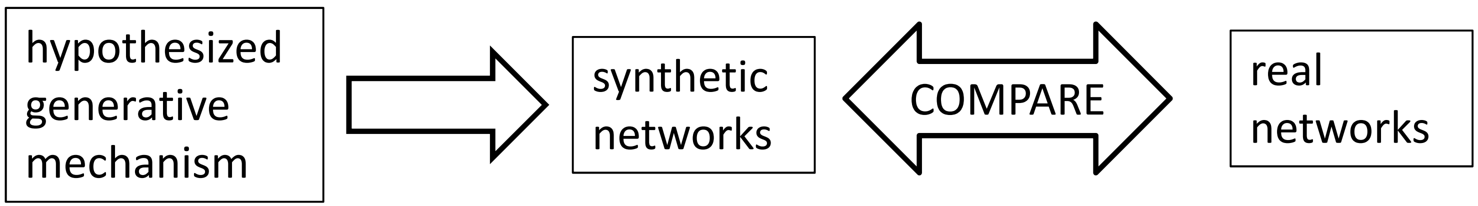
Complex Networks
models of network formation

2017.12.21(Thu)

contents of this chapter

- generative network models
  - models that offer explanations
  - why the network should have a power-law degree distribution?
- preferential attachment model
- vertex copying model

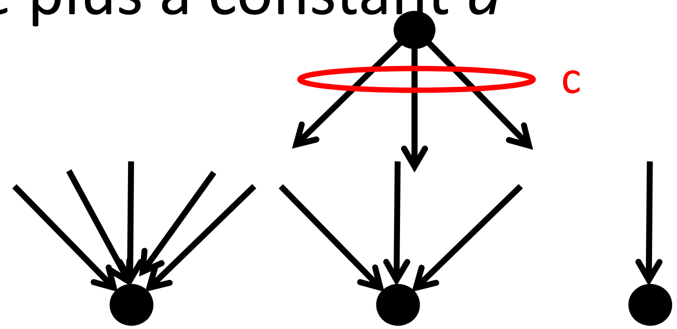


preferential attachment (1)

- many networks have degree distributions that approximately follow power laws

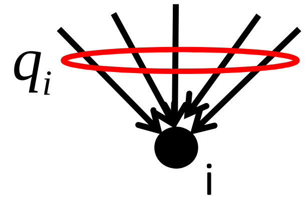
  Internet, WWW, citation networks, …

- Price's model of a citation network
  - a growing network of papers and their citations
  - vertices are continuously added but never destroyed
  - each paper cite on average c others
  - sited papers are chosen at random with probability proportional to their in-degree plus a constant *a*
- "rich-get-richer" effect



preferential attachment (2)

- it generate purely acyclic networks : every edge points backward in time



- q_i: in-degree of a vertex (instead of k_i^{in})
- $p_q(n)$: fraction of vertices that have in-degree q when the network contains n vertices
- citation to a particular vertex i is proportional to $q_i + a$

probability that a citation is to vertex i

$$\frac{q_i + a}{\sum_i (q_i + a)} = \frac{q_i + a}{n\langle q \rangle + na} = \frac{q_i + a}{n(c + a)}$$

- there are $np_q(n)$ vertices with degree q, and each new paper cites c others
  - expected number of new citations to all vertices with degree q is

$$np_q(n) \times c \times \frac{q + a}{n(c + a)} = \frac{c(q + a)}{c + a} p_q(n)$$

# of vertices with degree q

a new paper cite c papers

prob. that a citation is to vertices of degree q

preferential attachment (3)

- when we add a single new vertex to the network of n vertices
  - the number of vertices with in-degree q is

$$(n+1)p_q(n+1) = np_q(n) + \frac{c(q-1+a)}{c+a}p_{q-1}(n) - \frac{c(q+a)}{c+a}p_q(n)$$

the number of vertices previously of in-degree q

the number of vertices becoming in-degree q

the number of vertices becoming in-degree q+1

  - the appropriate equation for q=0 is

$$(n+1)p_0(n+1) = np_0(n) + 1 - \frac{ca}{c+a}p_0(n)$$

one vertex of in-degree zero

preferential attachment (4)

- taking the limit $n \to \infty$ and using $p_q = p_q(\infty)$

$$p_q = \frac{c}{c+a}\left[(q-1+a)\,p_{q-1} - (q+a)\,p_q\right] \text{ for q>=1}$$

$$p_0 = 1 - \frac{ca}{c+a}\,p_0 \qquad \text{for q=0}$$

$$\Rightarrow \quad p_q = \frac{q+a-1}{q+a+1+a/c}\,p_{q-1}$$

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

$$p_1 = \frac{a}{a+2+a/c}\,p_0 = \frac{a}{(a+2+a/c)}\frac{(1+a/c)}{(a+1+a/c)}$$

$$p_2 = \frac{a+1}{a+3+a/c}\,p_1 = \frac{(a+1)a}{(a+3+a/c)(a+2+a/c)}\frac{(1+a/c)}{(a+1+a/c)}$$

$$p_3 = \frac{a+2}{a+4+a/c}\,p_2 = \frac{(a+2)(a+1)a}{(a+4+a/c)(a+3+a/c)(a+2+a/c)}\frac{(1+a/c)}{(a+1+a/c)}$$

$$p_q = \frac{(q+a-1)(q+a-2)\ldots a}{(q+a+1+a/c)\ldots(a+2+a/c)}\frac{(1+a/c)}{(a+1+a/c)}$$

preferential attachment (5)

- gamma function $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$

$$\Gamma(x+1)=\int_0^\infty t^x e^{-t}dt=-\left[t^x e^{-t}\right]_0^\infty + x\int_0^\infty t^{x-1}e^{-t}dt = x\Gamma(x)$$

$$\frac{\Gamma(x+n)}{\Gamma(x)}=(x+n-1)(x+n-2)\ldots x$$

$$p_q=(1+a/c)\frac{\Gamma(q+a)\Gamma(a+1+a/c)}{\Gamma(a)\Gamma(q+a+2+a/c)}$$

- multiply both the numerator and the denominator by $\Gamma(2+a/c)=(1+a/c)\Gamma(1+a/c)$

degree distribution is determined by the behavior of this function

$$p_q=\frac{\Gamma(q+a)\Gamma(2+a/c)}{\Gamma(q+a+2+a/c)}\times\frac{\Gamma(a+1+a/c)}{\Gamma(a)\Gamma(1+a/c)}=\frac{B(q+a,2+a/c)}{B(a,1+a/c)}$$

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

Euler's beta function

preferential attachment (6)

- for large q and fixed a and c

$$\Gamma(x) \cong \sqrt{2\pi} e^{-x} x^{x-\frac{1}{2}}$$

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \cong \frac{e^{-x} x^{x-\frac{1}{2}}}{e^{-(x+y)} (x+y)^{x+y-\frac{1}{2}}} \Gamma(y)$$

$$\because e^{x} = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^{n}$$

$$(x+y)^{x+y-\frac{1}{2}} = x^{x+y-\frac{1}{2}}\left[1+\frac{y}{x}\right]^{x+y-\frac{1}{2}} \cong x^{x+y-\frac{1}{2}} e^{y}$$

$$B(x,y) \cong \frac{e^{-x} x^{x-\frac{1}{2}}}{e^{-(x+y)} x^{x+y-\frac{1}{2}} e^{y}} \Gamma(y) = x^{-y}\Gamma(y)$$

beta function B(x,y) as a power law for large value of x with exponent y

$$p_q = \frac{B(q+a, 2+a/c)}{B(a, 1+a/c)} \approx (q+a)^{-\alpha} \approx q^{-\alpha} \qquad \alpha = 2 + \frac{a}{c}$$

computer simulation of Price's model

- by performing computer simulations, we can calculate
  - network quantities, such as path lengths, correlations, clustering coefficients
  - solution of dynamical models, percolation processes, opinion formation models, and others

Method proposed by Krapivsky and Redner (1)

- probability that an edge attaches to vertex i

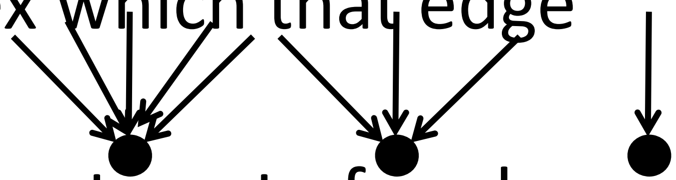
$$\theta_i = \frac{q_i + a}{n(c+a)}$$

- the edge is attached to a vertex
  - with probability ϕ: the vertex is chosen in proportion to its in-degree $\frac{q_i}{\sum_j q_j} = \frac{q_i}{nc}$
  - with probability 1-ϕ: the vertex is chosen uniformly at random $\frac{1}{n}$
- total probability of attaching to vertex i

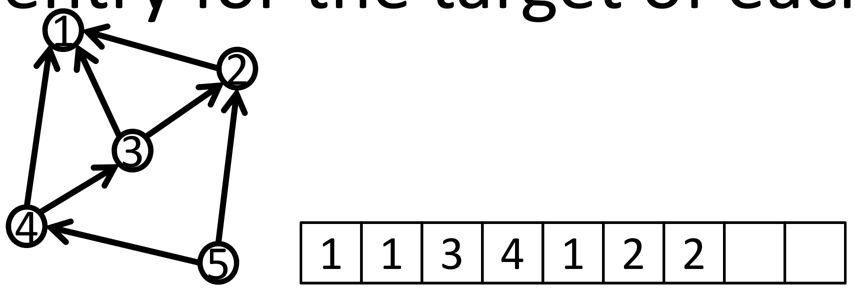
$$\theta_i' = \phi \frac{q_i}{nc} + (1-\phi)\frac{1}{n} \qquad \phi = \frac{c}{c+a} \Rightarrow \theta_i' = \frac{c}{c+a}\frac{q_i}{nc} + \left(1 - \frac{c}{c+a}\right)\frac{1}{n} = \frac{q_i + a}{n(c+a)}$$

Vertex label list used in the simulation of Price's model

- With probability c/(c+a) choose a vertex in strict proportion to in-degree. Otherwise, choose a vertex uniformly at random from the set of all vertices.
- Choosing in proportion to in-degree is equivalent to picking an edge in the network uniformly at random and choosing the vertex which that edge points to.



- the list contains one entry for the target of each edge in the network



Algorithm for creating a new edge

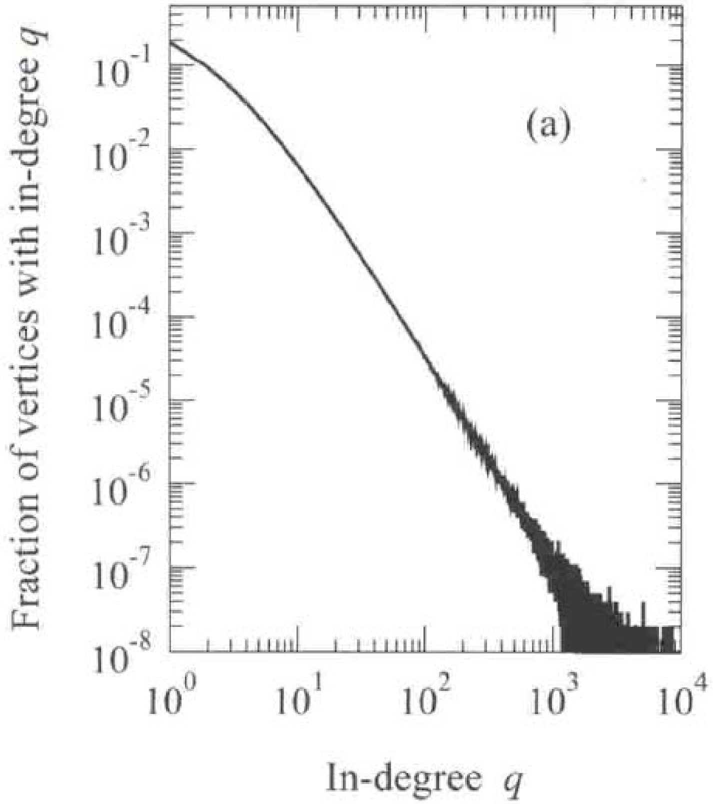
1. Generate a random number r in the range $0 \le r < 1$
2. If $r < c/(c+a)$, choose an element uniformly at random from the list of targets
3. Otherwise, choose a vertex uniformly at random from the set of all vertices
4. Create an edge linking to the vertex thus selected, and add that vertex to the end of the list of targets

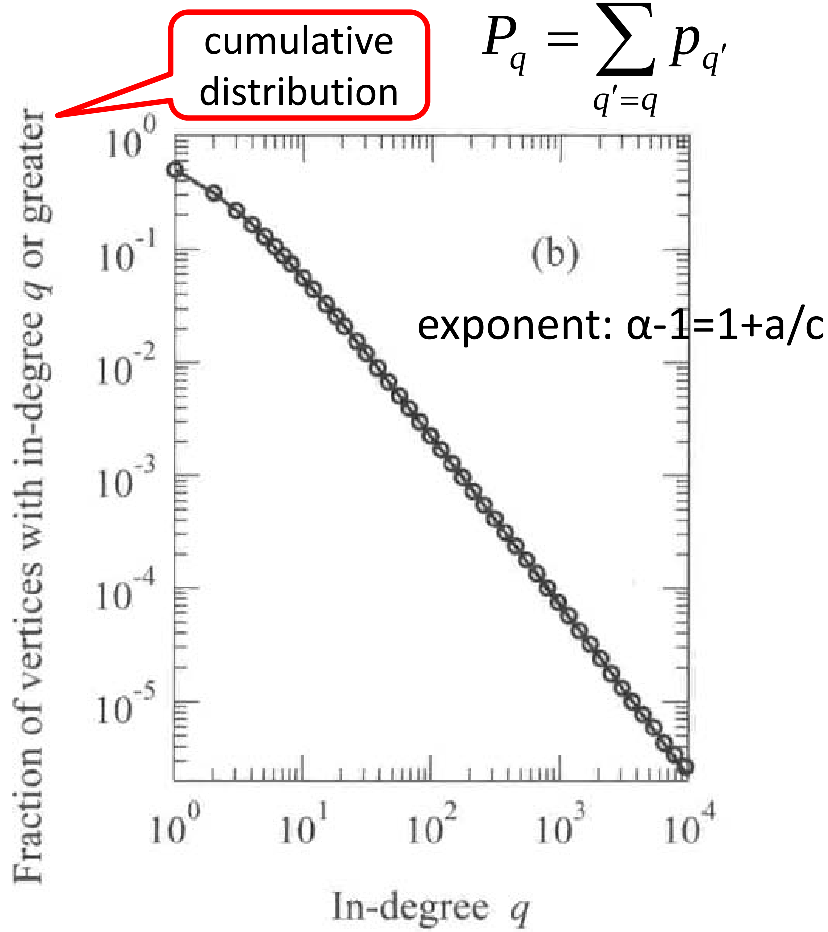
- Each step can be accomplished in constant time, so the growth of a network of n vertices can be accomplished in time $O(n)$

Degree distribution of a 100-million-node network

- power-law in the tail of the distribution

$$P_q = \sum_{q'=q}^{\infty} p_{q'}$$





analysis of cumulative distribution function

$$\alpha - 1 = 1 + \frac{a}{c}$$

$$\alpha = 2 + \frac{a}{c}$$

- exponent of P_q is less than that of the degree distribution

$$B(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1} du$$

$$P_q = \sum_{q'=q}^{\infty} p_{q'} = \sum_{q'=q}^{\infty} \frac{B(q'+a, 2+a/c)}{B(a, 1+a/c)} = \frac{1}{B(a,1+a/c)} \sum_{q'=q}^{\infty} \int_0^1 u^{q'+a-1}(1-u)^{1+a/c} du$$

$$= \frac{1}{B(a,1+a/c)} \int_0^1 u^{a-1} \sum_{q'=q}^{\infty} u^{q'} (1-u)^{1+a/c} du$$

$$\sum_{q'=q}^{\infty} u^{q'} = u^q + u^{q+1} + \ldots = \frac{u^q}{1-u}$$

$$(-1 < u < 1)$$

$$= \frac{1}{B(a,1+a/c)} \int_0^1 u^{q+a-1}(1-u)^{a/c} du$$

$$= \frac{B(q+a, 1+a/c)}{B(a, 1+a/c)}$$

cf.

$$p_q = \frac{B(q+a, 2+a/c)}{B(a, 1+a/c)} \approx (q+a)^{-\alpha} \approx q^{-\alpha}$$

$$\alpha = 2 + \frac{a}{c}$$

the model of Barabasi and Albert (1)

- a model of an undirected network
- the number of connections made by each vertex is exactly c (must be an integer)
- vertices and edges are never taken away
- k_i : the degree of vertex i
- the smallest degree in the network is always c
- the model is equivalent to a special case of Price's model

$$p_q = \frac{B(q+a, 2+a/c)}{B(a, 1+a/c)} \quad a = c \;\Rightarrow\; p_q = \frac{B(q+c, 3)}{B(c, 2)}$$

the model of Barabasi and Albert (2)

the total degree   in-degree   out-degree

- Price's model : $k_i = q_i + c$
- BA model : $p_q = \frac{B(q+c,3)}{B(c,2)}$ replace q+c by k

$$p_k = \begin{cases} \frac{B(k,3)}{B(c,2)} & for\_k \geq c \\ 0 & for\_k < c \end{cases}$$

- this can be simplified further $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$

$$B(k,3) = \frac{\Gamma(k)\Gamma(3)}{\Gamma(k+3)} = \frac{\Gamma(3)}{k(k+1)(k+2)} \qquad \frac{\Gamma(x+n)}{\Gamma(x)} = (x+n-1)(x+n-2)\ldots x$$

$$B(c,2) = \frac{\Gamma(2)}{c(c+1)}$$

$for\_k \geq c$

$$p_k = \frac{\Gamma(3)}{\Gamma(2)} \frac{c(c+1)}{k(k+1)(k+2)} = \frac{2c(c+1)}{k(k+1)(k+2)} \approx k^{-3}$$

In the limit where k becomes large $\alpha = 3$

degree distribution with R+igraph

synthetic network based on Barabasi-Albert model (n = 10000, c=3, undirected)

```
> library(igraph)
> ba <- barabasi.game(10000,m=3,directed =
FALSE)
> summary(ba)
```

summary of the network

```
Vertices: 10000
Edges: 29997
Directed: FALSE
No graph attributes.
No vertex attributes.
No edge attributes.
> no.clusters(ba)
```

the number of clusters

```
[1] 1
> average.path.length(ba)
[1] 3.170916
```

average path length

```
> transitivity(ba)
```

clustering coefficient

```
[1] 0.001810504
> mean(degree(ba))
```

average degree

```
[1] 5.9994
> max(degree(ba))
```

the maximum degree

```
[1] 4184
> min(degree(ba))
```

the minimum degree

```
[1] 3
> power.law.fit(degree(ba))
Loading required package: stats4
Call:
```

fits a power-law distribution

```
mle(minuslogl = mlogl, start = list(alpha = start))
Coefficients:
  alpha
3.155893
```

power-law distribution

```
> plot(degree.distribution(ba),log="xy")
```