preferential attachment

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2023/24

preferential *attachment*

- *generative models* reason about *network evolution*
- *cumulative advantage* process of *Price model* [Pri76]
- *preferential attachment* of *Barabási-Albert model* [BA99]

Pólya process Yule process Zipf's law Matthew effect
rich-get-richer proportional growth cumulative advantage

see preferential attachment model NetLogo demo

Derek de Solla Price

Albert-László Barabási

Réka Albert

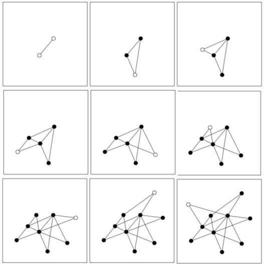
preferential $G(n, c, a)$ *model*

- $G(n, c, a)$ *cumulative advantage* model [Pri76]
- each new node i *forms* $k_i^{out} = c > 0$ *directed links*
- node j *receives link with probability* $\sim k_j^{in} + a = q_j + a > 0$

n, c, a given $\qquad p_q$ *unknown*

input parameters n, c, a
output *directed* graph G

```
1: G ← ≥ c isolated nodes
2: while not G has n nodes do
3:     add node i to G
4:     for c times do
5:         add link (i, j) with ~ q_j + a
6: return G
```



preferential $G(n, c, a)$ equation

— *master equation* for *in-degree distribution* $p_q(n)$

– $p_q(n)$ is in-degree distribution p_q *at time* n

$$\frac{q_i+a}{\sum_i q_i+a} = \frac{q_i+a}{n(c+a)} \qquad cnp_q(n)\frac{q+a}{n(c+a)} = \frac{c(q+a)}{c+a}p_q(n)$$

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

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

— *power-law in-degree distribution* $p_q \sim q^{-\gamma}$ with $\gamma > 2$

– p_q is in-degree distribution *in limit* $n \to \infty$

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t \qquad B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \simeq x^{-y}\Gamma(y)$$

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

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

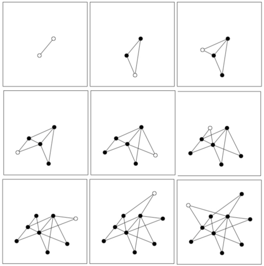
preferential $G(n,c)$ *model*

- $G(n,c)$ *preferential attachment* model [BA99]
- each new node i *forms* $c > 0$ *undirected links*
- node j *receives links with probability* $\sim k_i$

n, c given p_k *unknown*

input parameters n, c
output *undirected* graph G

$G \leftarrow c$ connected nodes
while not G has n nodes **do**
  add node i to G
  for c **times do**
    add link $\{i,j\}$ with $\sim k_j$
return G



preferential $G(n, c)$ *equation*

- *undirected* $G(n, c)$ is *directed* $G(n, c, c)$ for $k_i = q_i + c$
- *same master equation* for *in-degree distribution* p_q
  - p_q is in-degree distribution *in limit* $n \to \infty$

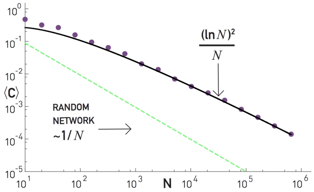
$$p_q = \frac{B(q+c, 2+c/c)}{B(c, 1+c/c)} = \frac{B(q+c, 3)}{B(c, 2)} \sim q^{-3}$$

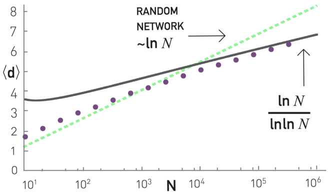
- *power-law degree distribution* $p_k \sim k^{-3}$
  - p_k is degree distribution *in limit* $n \to \infty$

$$p_k = \frac{B(k, 3)}{B(c, 2)} = \cdots = \frac{2c(c+1)}{k(k+1)(k+2)} \sim k^{-3}$$

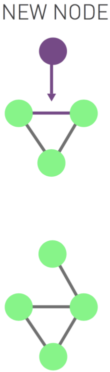
preferential $\neg$*small-world*

- *random graphs* are *"small-world"* as $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$
- *random graphs* are *not small-world* as $\langle C \rangle = \frac{\langle k \rangle}{n-1}$
- *scale-free networks* $\gamma = 3$ are *"small-world"* as $\langle d \rangle \sim \frac{\ln n}{\ln \ln n}$
- $G(n, c)$ *scale-free model* is *not small-world* as $\langle C \rangle \simeq \frac{(\ln n)^2}{n}$

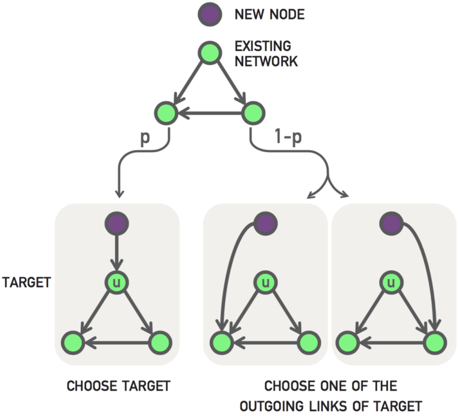




preferential *models*

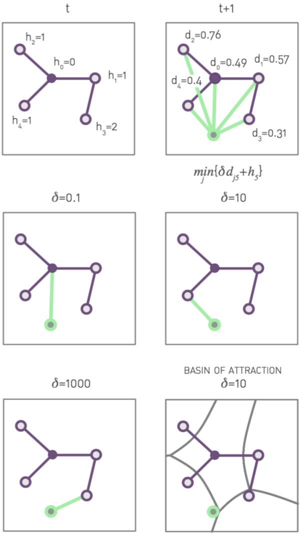


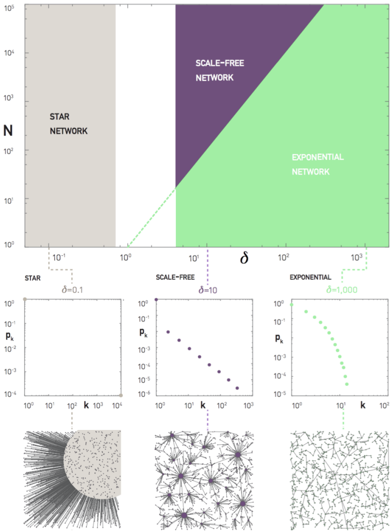
link selection [DM02]



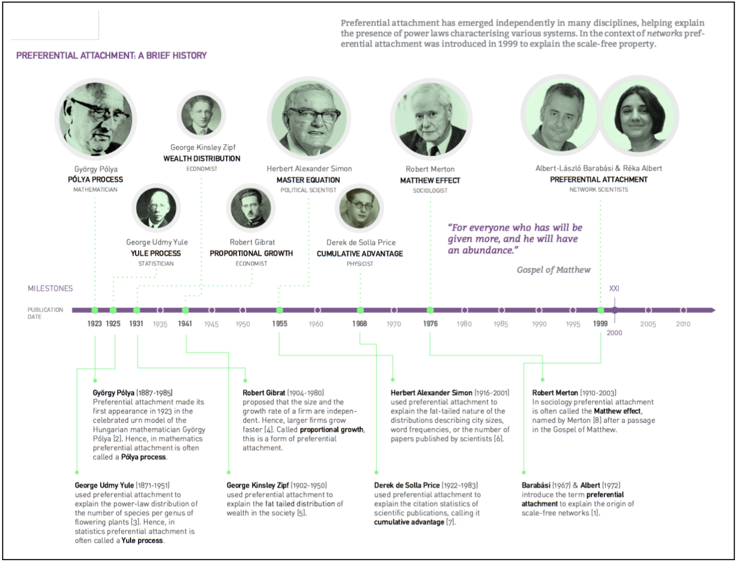
random *link copying* model [KKR+99]

preferential *optimization*





preferential *history*



preferential *references*

A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286(5439):509–512, 1999.

A.-L. Barabási.
Network Science.
Cambridge University Press, Cambridge, 2016.

Raissa M. D'Souza, Christian Borgs, Jennifer T. Chayes, Noam Berger, and Robert D. Kleinberg.
Emergence of tempered preferential attachment from optimization.
P. Natl. Acad. Sci. USA, 104(15):6112–6117, 2007.

S. N. Dorogovtsev and J. F. F. Mendes.
Evolution of networks.
Adv. Phys., 51(4):1079–1187, 2002.

Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.
Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition.
Cambridge University Press, Cambridge, 2011.

David Easley and Jon Kleinberg.
Networks, Crowds, and Markets: Reasoning About a Highly Connected World.
Cambridge University Press, Cambridge, 2010.

Ernesto Estrada and Philip A. Knight.
A First Course in Network Theory.
Oxford University Press, 2015.

preferential *references*

Alex Fabrikant, E. Koutsoupias, and C. H. Papadimitriou.
Heuristically optimized trade-offs: A new paradigm for power laws in the Internet.
In *Proceedings of the International Colloquium on Automata, Languages and Programming*, pages 110–122, Malaga, Spain, 2002.

Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins.
The web as a graph: Measurements, models, and methods.
In *Proceedings of the International Conference on Computing and Combinatorics*, pages 1–17, Tokyo, Japan, 1999.

Mark E. J. Newman.
Networks.
Oxford University Press, Oxford, 2nd edition, 2018.

Derek De Solla Price.
A general theory of bibliometric and other cumulative advantage processes.
J. Am. Soc. Inf. Sci., 27(5):292–306, 1976.