preferential attachment

introduction to network analysis (ina)

Lovro Šubelj
University of Ljubljana
spring 2021/22
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 $k_i^{\text{out}} = c > 0$ directed links
- node $j$ receives link with probability $\sim k_j^{\text{in}} + a = q_j + a > 0$

$n, c, a$ given \hspace{1cm} p_q \text{ unknown}$

---

**Algorithm**

```
input parameters $n, c, a$
output directed graph $G$

1: $G \leftarrow \geq c$ 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 $\sim q_j + a$
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* \( n \)

\[
\frac{q_i+a}{\sum_i q_i+a} = \frac{q_i+a}{n(c+a)} \quad \text{cn}p_q(n) \quad \frac{q+a}{n(c+a)} = \frac{c(q+a)}{c+a} p_q(n)
\]

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

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

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

- \( p_q \) is in-degree distribution *in limit* \( n \to \infty \)

\[
\Gamma(x) = \int_0^\infty t^{x-1}e^{-t} \, dt \quad \text{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \sim x^{-y}\Gamma(y)
\]

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

\[
p_0 = \frac{1+a/c}{a+1+a/c}
\]
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 \( \sim k_i \)
  
\[
\begin{align*}
n, c \text{ given} & \quad p_k \text{ unknown}
\end{align*}
\]

\[
\begin{aligned}
\text{input} & \quad \text{parameters } n, c \\
\text{output} & \quad \text{undirected graph } G \\
1: & \quad G \leftarrow c \text{ connected nodes} \\
2: & \quad \text{while not } G \text{ has } n \text{ nodes do} \\
3: & \quad \quad \text{add node } i \text{ to } G \\
4: & \quad \quad \text{for } c \text{ times do} \\
5: & \quad \quad \quad \text{add link } \{i, j\} \text{ with } \sim k_j \\
6: & \quad \text{return } G
\end{aligned}
\]
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 $\langle d \rangle \sim \frac{\ln n}{\ln \langle k \rangle}$

— 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 $\langle C \rangle \sim \frac{(\ln n)^2}{n}$
preferential *models*

*link selection* [DM02]  
random *link copying* model [KKR+99]
preferential optimization
preferential history

Preferential attachment has emerged independently in many disciplines, helping explain the presence of power laws characterising various systems. In the context of networks preferential attachment was introduced in 1999 to explain the scale-free property.

György Pólya [1887-1985]
Pólya process
Mathematician

George Kinsley Zipf
Wealth distribution
Economist

Herbert Alexander Simon
Master equation
Political scientist

Robert Merton
Matthew effect
Sociologist

Albert-László Barabási & Réka Albert
Preferential attachment
Network scientists

"For everyone who has will be given more, and he will have an abundance."

Gospel of Matthew

Milestones

Publication Date

György Pólya [1887-1985]
Preferential attachment made its first appearance in 1923, in the celebrated urn model of the Hungarian mathematician György Pólya [2]. Hence, in mathematics preferential attachment is often called a Pólya process.

George Udny Yule [1871-1951]
Yule process
Statistician

Robert Gibrnat [1904-1980]
Proportional growth
Economist

Derek de Solla Price [1922-1983]
Cumulative advantage
Physicist

Robert Merton [1910-2003]
In sociology preferential attachment is often called the Matthew effect, named by Merton [8] after a passage in the Gospel of Matthew.

Introduce the term preferential attachment to explain the origin of scale-free networks [1].
preferential references

A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.

A.-L. Barabási.
*Network Science*.

Emergence of tempered preferential attachment from optimization.

S. N. Dorogovtsev and J. F. F. Mendes.
Evolution of networks.

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

David Easley and Jon Kleinberg.
*Networks, Crowds, and Markets: Reasoning About a Highly Connected World*.

Ernesto Estrada and Philip A. Knight.
*A First Course in Network Theory*.
Alex Fabrikant, E. Koutsoupias, and C. H. Papadimitriou.
Heuristically optimized trade-offs: A new paradigm for power laws in the Internet.

Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins.
The web as a graph: Measurements, models, and methods.

*Networks*.

Derek De Solla Price.
A general theory of bibliometric and other cumulative advantage processes.