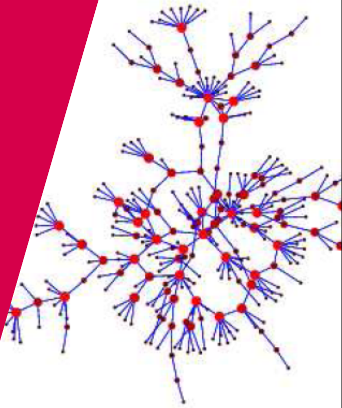
Distances in power-law random graphs

Sander Dommers

Supervisor: Remco van der Hofstad



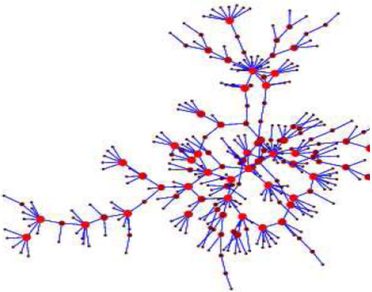
Technische Universiteit
Eindhoven
University of Technology

February 2, 2009

Introduction

There are many complex real-world networks, e.g.

- Social networks (friendships, business relationships, sexual contacts, ...)
- Information networks (World Wide Web, citations, ...)
- Technological networks (Internet, airline routes, ...)
- Biological networks (protein interactions, neural networks,...)

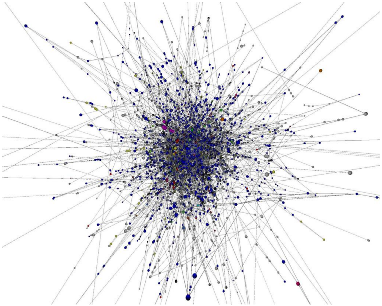


Sexual network Colorado Springs, USA (Potterat, et al., 2002)

Introduction

There are many complex real-world networks, e.g.

- Social networks (friendships, business relationships, sexual contacts, ...)
- Information networks (World Wide Web, citations, ...)
- Technological networks (Internet, airline routes, ...)
- Biological networks (protein interactions, neural networks,...)



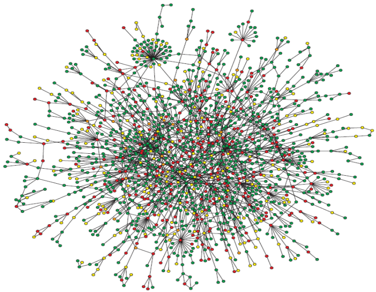
Small part of the Internet
http://www.fractalus.com/steve/stuff/ipmap/

Technische Universiteit Eindhoven University of Technology

Introduction

There are many complex real-world networks, e.g.

- Social networks (friendships, business relationships, sexual contacts, ...)
- Information networks (World Wide Web, citations, ...)
- Technological networks (Internet, airline routes, ...)
- Biological networks (protein interactions, neural networks,...)

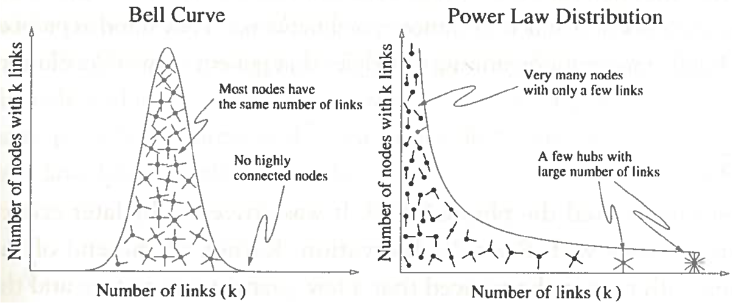


Yeast protein interaction network
(Jeong, et al., 2001)

Properties of complex networks

Power law behavior

Number of vertices with degree k is proportional to $k^{-\tau}$



Small worlds

Distances in the network are small

Erdős-Rényi random graph

- Start with n vertices
- Draw edges between each pair of vertices with probability p

Exponential tails, so no power-law distribution

Erdős-Rényi random graph

- Start with n vertices
- Draw edges between each pair of vertices with probability p

Exponential tails, so no power-law distribution

Configuration model

- Start with n vertices
- Let vertex i have D_i half-edges
  D_i has power-law distribution with exponent τ
- Connect the half-edges uniformly at random

Power-law because of Strong Law of Large Numbers

Preferential attachment (PA)

- Networks are growing
- Popular vertices are more likely to get more edges than unpopular vertices

Preferential attachment (PA)

- Networks are growing
- Popular vertices are more likely to get more edges than unpopular vertices

PA model

- Choose integer m, and parameter $\delta > -m$
- Start at time $t = 2$ with 2 vertices connected with $2m$ edges
- At each time step, say $t+1$, add a new vertex with m edges
- Connect its m edges independently according to

$$\mathbb{P}\left[\text{edge of } t+1 \longrightarrow i \middle| \mathrm{PA}_{m,\delta}(t)\right] = \frac{D_i(t)+\delta}{t(2m+\delta)}, \qquad \text{for } i \in [t].$$

Properties of PA model

Power law with exponent

$$\tau = 3 + \frac{\delta}{m}$$

So any exponent $\tau > 2$ possible

TU/e Technische Universiteit Eindhoven University of Technology

Properties of PA model

Power law with exponent

$$\tau = 3 + \frac{\delta}{m}$$

So any exponent $\tau > 2$ possible

The diameter of a graph at time t, $\mathrm{diam}(t)$,
is the largest distance present in the graph at time t

Technische Universiteit Eindhoven University of Technology

Properties of PA model

Power law with exponent

$$\tau = 3 + \frac{\delta}{m}$$

So any exponent $\tau > 2$ possible

The diameter of a graph at time t, $\mathrm{diam}(t)$,
is the largest distance present in the graph at time t

$\tau = 3$ (Bollobás and Riordan, 2004)
For $m \geq 2$ and $\delta = 0$, with high probability,

$$(1-\varepsilon)\frac{\log t}{\log \log t} \leq \mathrm{diam}(t) \leq (1+\varepsilon)\frac{\log t}{\log \log t}$$

for all $\varepsilon > 0$

Properties of PA model (continued)

$\tau > 3$ (Van der Hofstad and Hooghiemstra, 2008)
For $m \geq 2$ and $\delta > 0$, with high probability,

$$c_1 \log t \leq \mathrm{diam}(t) \leq c_2 \log t$$

for some $c_1, c_2 > 0$

$\tau \in (2, 3)$ (Van der Hofstad and Hooghiemstra, 2008)
For $m \geq 2$ and $\delta \in (-m, 0)$, with high probability,

$$\mathrm{diam}(t) \leq c \log \log t$$

for some $c > 0$

log log lowerbound for $\tau > 2$

Theorem

Consider preferential attachment with $m \geq 2$ *and* $\delta > -m$ *(i.e.* $\tau > 2$*)*

Let $k = \frac{\varepsilon}{\log m} \log \log t$*, with* $0 < \varepsilon < 1$

Then, with high probability,

$$diam(2t) \geq k$$

Proper k-exploration trees

Construction of k-exploration tree of a vertex i

- Start from vertex i
- Connect its m edges $\longrightarrow$ vertices at distance 1 from vertex i
- Connect the m edges of vertices at distance 1, $\longrightarrow \ldots$
- Continue in same fashion, up to distance k

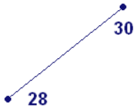
Example for $m = 2$, $k = 3$ and $i = 30$

Proper k-exploration trees

Construction of k-exploration tree of a vertex i

- Start from vertex i
- Connect its m edges $\longrightarrow$ vertices at distance 1 from vertex i
- Connect the m edges of vertices at distance 1, $\longrightarrow$ ...
- Continue in same fashion, up to distance k

Example for $m = 2$, $k = 3$ and $i = 30$

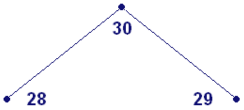


Proper k-exploration trees

Construction of k-exploration tree of a vertex i

- Start from vertex i
- Connect its m edges $\longrightarrow$ vertices at distance 1 from vertex i
- Connect the m edges of vertices at distance 1, $\longrightarrow \ldots$
- Continue in same fashion, up to distance k

Example for $m = 2$, $k = 3$ and $i = 30$

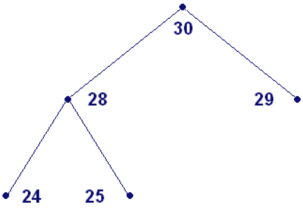


Proper k-exploration trees

Construction of k-exploration tree of a vertex i

- Start from vertex i
- Connect its m edges $\longrightarrow$ vertices at distance 1 from vertex i
- Connect the m edges of vertices at distance 1, $\longrightarrow \ldots$
- Continue in same fashion, up to distance k

Example for $m = 2$, $k = 3$ and $i = 30$

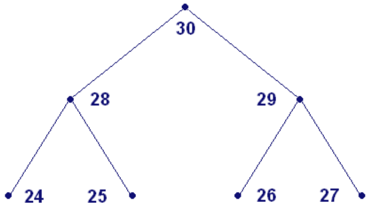


Proper k-exploration trees

Construction of k-exploration tree of a vertex i

- Start from vertex i
- Connect its m edges $\longrightarrow$ vertices at distance 1 from vertex i
- Connect the m edges of vertices at distance 1, $\longrightarrow \ldots$
- Continue in same fashion, up to distance k

Example for $m = 2$, $k = 3$ and $i = 30$

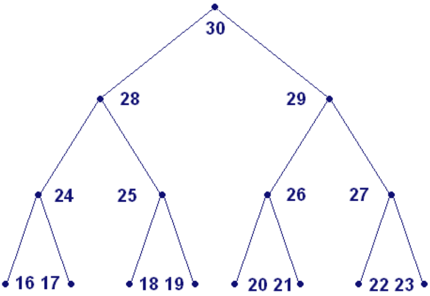


Proper k-exploration trees

Construction of k-exploration tree of a vertex i

- Start from vertex i
- Connect its m edges $\longrightarrow$ vertices at distance 1 from vertex i
- Connect the m edges of vertices at distance 1, $\longrightarrow \ldots$
- Continue in same fashion, up to distance k

Example for $m = 2$, $k = 3$ and $i = 30$

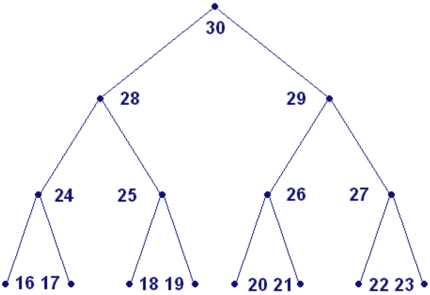


Proper k-exploration trees (continued)

A k-exploration tree is proper when

- The k-exploration tree has no collisions
- No other vertex connects to a vertex in the tree
- All vertices of the tree are in $[2t] \setminus [t]$

Example for $m = 2$, $k = 3$ and $i = 30$

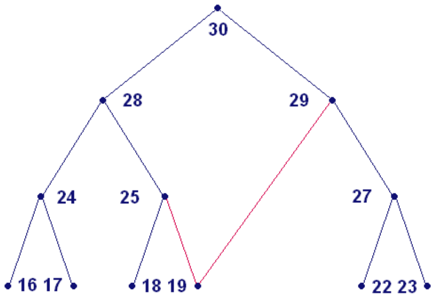


Proper k-exploration trees (continued)

A k-exploration tree is proper when

- The k-exploration tree has no collisions
- No other vertex connects to a vertex in the tree
- All vertices of the tree are in $[2t] \backslash [t]$

Example for $m = 2$, $k = 3$ and $i = 30$

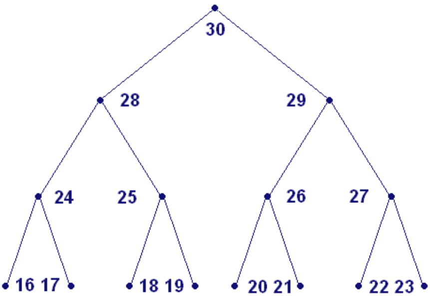


Proper k-exploration trees (continued)

A k-exploration tree is proper when

- The k-exploration tree has no collisions
- No other vertex connects to a vertex in the tree
- All vertices of the tree are in $[2t] \backslash [t]$

Example for $m = 2$, $k = 3$ and $i = 30$

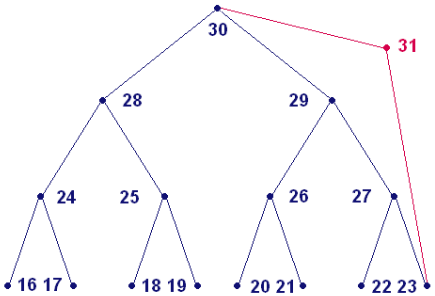


Proper k-exploration trees (continued)

A k-exploration tree is proper when

- The k-exploration tree has no collisions
- No other vertex connects to a vertex in the tree
- All vertices of the tree are in $[2t] \backslash [t]$

Example for $m = 2$, $k = 3$ and $i = 30$



Number of proper k-exploration trees

Let

$$Z_k(2t) = \#\ \text{proper } k\text{-exploration trees at time } 2t$$

Number of proper k-exploration trees

Let

$$Z_k(2t) = \#\text{ proper } k\text{-exploration trees at time } 2t$$

Then

$$\mathbb{P}\left[\mathrm{diam}(2t) < k\right] \leq \mathbb{P}\left[Z_k(2t) = 0\right]$$

Number of proper k-exploration trees

Let

$$Z_k(2t) = \#\text{ proper } k\text{-exploration trees at time } 2t$$

Then

$$\mathbb{P}\left[\text{diam}(2t) < k\right] \leq \mathbb{P}\left[Z_k(2t) = 0\right]$$

Using Chebychev inequality gives

$$\mathbb{P}\left[Z_k(2t) = 0\right] \leq \frac{\mathbb{V}ar\left[Z_k(2t)\right]}{\mathbb{E}\left[Z_k(2t)\right]^2} \xrightarrow{?} 0$$

Number of proper k-exploration trees

Let

$$Z_k(2t) = \#\text{ proper } k\text{-exploration trees at time } 2t$$

Then

$$\mathbb{P}[\mathrm{diam}(2t) < k] \leq \mathbb{P}[Z_k(2t) = 0]$$

Using Chebychev inequality gives

$$\mathbb{P}[Z_k(2t) = 0] \leq \frac{\mathbb{V}ar[Z_k(2t)]}{\mathbb{E}[Z_k(2t)]^2} \xrightarrow{?} 0$$

We need

$$\mathbb{E}[Z_k(2t)] \longrightarrow \infty \qquad \text{and} \quad \mathbb{V}ar[Z_k(2t)] \ll \mathbb{E}[Z_k(2t)]^2$$

$\mathbb{E}\left[Z_k(2t)\right] \longrightarrow \infty$

Let

$$\mathcal{B}_{\mathcal{T}} = \text{“no other vertex connects to a vertex in the tree } \mathcal{T}\text{”}$$

Then

$$\begin{aligned}
\mathbb{E}\left[Z_k(2t)\right] &= \mathbb{E}\left[\sum_{\mathcal{T} \text{ possible}} \mathbb{I}\{\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t) \text{ and } \mathcal{T} \text{ is proper}\}\right] \\
&= \sum_{\mathcal{T} \text{ possible}} \mathbb{P}\left[\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t) \text{ and } \mathcal{T} \text{ is proper}\right] \\
&= \sum_{\mathcal{T} \text{ possible}} \mathbb{P}\left[\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)\right] \cdot \mathbb{P}\left[\mathcal{B}_{\mathcal{T}} | \mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)\right]
\end{aligned}$$

Technische Universiteit Eindhoven University of Technology

$\mathbb{E}[Z_k(2t)] \longrightarrow \infty$ (continued)

$$\mathbb{E}[Z_k(2t)] = \sum_{\mathcal{T} \text{ possible}} \mathbb{P}[\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)] \cdot \mathbb{P}[\mathcal{B}_{\mathcal{T}} | \mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)]$$

$$\mathbb{E}[Z_k(2t)] = \sum_{\mathcal{T} \text{ possible}} \mathbb{P}[\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)] \cdot \mathbb{P}[\mathcal{B}_{\mathcal{T}} | \mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)]$$

$$\gtrsim \left(\frac{1}{t}\right)^{|\mathcal{T}|-1}$$

$\mathbb{E}[Z_k(2t)] \longrightarrow \infty$ (continued)

$$\mathbb{E}[Z_k(2t)] = \sum_{\mathcal{T} \text{ possible}} \mathbb{P}[\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)] \cdot \mathbb{P}[\mathcal{B}_{\mathcal{T}} | \mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)]$$

$$\gtrsim \left(\frac{1}{t}\right)^{|\mathcal{T}|-1} \left(1 - \frac{|\mathcal{T}|}{t}\right)^{mt}$$

$\mathbb{E}[Z_k(2t)] \longrightarrow \infty$ (continued)

$$\mathbb{E}[Z_k(2t)] = \sum_{\mathcal{T} \text{ possible}} \mathbb{P}[\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)] \cdot \mathbb{P}[\mathcal{B}_{\mathcal{T}} | \mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)]$$

$$\gtrsim \left(\frac{t}{|\mathcal{T}|}\right)^{|\mathcal{T}|} \left(\frac{1}{t}\right)^{|\mathcal{T}|-1} \left(1 - \frac{|\mathcal{T}|}{t}\right)^{mt}$$

$\mathbb{E}\left[Z_k(2t)\right] \longrightarrow \infty$ (continued)

$$\mathbb{E}\left[Z_k(2t)\right] = \sum_{\mathcal{T}\text{ possible}} \mathbb{P}\left[\mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)\right] \cdot \mathbb{P}\left[\mathcal{B}_{\mathcal{T}} | \mathcal{T} \subseteq \mathrm{PA}_{m,\delta}(2t)\right]$$

$$\gtrsim \left(\frac{t}{|\mathcal{T}|}\right)^{|\mathcal{T}|} \left(\frac{1}{t}\right)^{|\mathcal{T}|-1} \left(1 - \frac{|\mathcal{T}|}{t}\right)^{mt}$$

$$\sim t\left(\frac{1}{(\log t)^{\varepsilon}}\right)^{(\log t)^{\varepsilon}} e^{-m(\log t)^{\varepsilon}}$$

$$\longrightarrow \infty$$

$\mathbb{V}ar\,[Z_k(2t)] \ll \mathbb{E}\,[Z_k(2t)]^2$

Study covariance of two different possible trees being formed in the graph and being proper

Take exact look at probabilities involved

Conclusion

$$\mathbb{V}ar\,[Z_k(2t)] \leq c\,\frac{(\log t)^2}{t}\,\mathbb{E}\,[Z_k(2t)]^2 + \mathbb{E}\,[Z_k(2t)] \ll \mathbb{E}\,[Z_k(2t)]^2$$

Technische Universiteit Eindhoven University of Technology

log log lowerbound for $\tau > 2$

Theorem

Consider preferential attachment with $m \geq 2$ *and* $\delta > -m$ *(i.e.* $\tau > 2$*)*

Let $k = \frac{\varepsilon}{\log m} \log \log t$*, with* $0 < \varepsilon < 1$

Then, with high probability,

$$diam(2t) \geq k$$

Proof

$$\begin{aligned} \mathbb{P}[\mathrm{diam}(2t) \geq k] &\geq 1 - \mathbb{P}\left[Z_k(2t) = 0\right] \\ &\geq 1 - \frac{\mathbb{V}ar\left[Z_k(2t)\right]}{\mathbb{E}\left[Z_k(2t)\right]^2} \\ &\longrightarrow 1 \end{aligned}$$

Conclusion

Small world property holds in preferential attachment model

Future research

- Better bounds on distances
- Distances in other extended PA models
- Processes on random graphs