Modelling self-organizing networks

Paweł Prałat

Department of Mathematics, Ryerson University, Toronto, ON

Cargese Fall School on Random Graphs
(September 2015)

Outline

Multidisciplinary research

Pure Mathematics:

- Graph Theory
- Random Structures and Algorithms
- Modelling

Applied Computer Science:

- ...

Social Science: for example,

- *Homophily, contagion and the decay of community structure in self-organizing networks* (PNAS paper!)
- *Social learning in a large, evolving network* (BlackBerry)

Multidisciplinary research

Applied Computer Science:

- *Utilizing big data for business-to-business matching and recommendation system* (ComLinked Corp., 2014-15)
- *A self-organizing dynamic network model increasing the efficiency of outdoor digital billboards* (KPM, 2014)
- *Exploiting Big Data for Customized Online News Recommendation System* (The Globe and Mail, 2014)
- *Personalized Mobile Recommender System* (BlackBerry, 2013-14)
- *Intelligent Rating System* (Mako, 2012-13)
- *Dynamic clustering and prediction of taxi service demand* (Winston, 2012)

Multidisciplinary research

Applied Computer Science (currently):

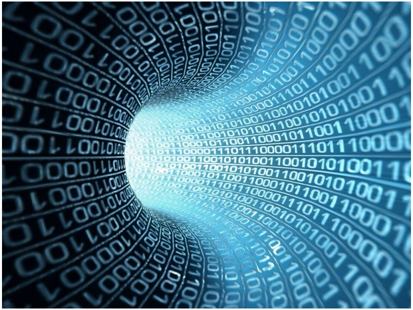
- *Web Visitor Engagement Measurement and Maximization* (The Globe and Mail, 2014-15)
- *Hypergraphs and their applications* (Tutte Institute for Mathematics and Computing)
- *Relationship Mapping Analytics for Fundraising and Sales Prospect Research* (Charter Press Ltd.)

Applied Computer Science (near future):

- *Network Modeling of Trust in Online Scientific Information Sources* (Bell Labs)
- ...

Outline

Big Data Era



Every human-technology interaction, or sensor network, generates new data points that can be viewed, based on the type of interaction, as a self-organizing network.

The web graph

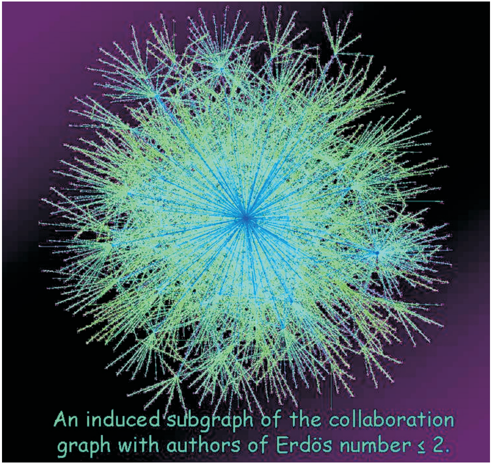
nodes: *web pages* edges: *hyperlinks*

Social networks

nodes: *people* edges: *social interaction (e.g. Facebook friendship)*

Social networks

nodes: *scientists* edges: *co-authorship*



Are these networks similar?

Are these networks similar?

Answer: Yes!

- large scale
- 'small world' property
  (e.g. low diameter of $O(\log n)$, high clustering coefficient)
- degree distribution
  (power-law, the number of nodes of degree k is proportional to $k^{-\gamma}$)
- bad expansion
- etc.

Why model self-organizing networks?

- uncover the generative mechanisms underlying self-organizing networks,
- models are a predictive tool,
- community detection,
- improving search engines (the web graph),
- spam and worm defense,
- nice mathematical challenges.

Why model self-organizing networks?

- uncover the generative mechanisms underlying self-organizing networks,
- models are a predictive tool,
- community detection,
- improving search engines (the web graph),
- spam and worm defense,
- nice mathematical challenges.

(For example, PA model justifies “rich get richer” principle.)

A good graph model should...

- ...reproduce experimentally observed graph properties:
  - degree distribution follows a power law,
  - small average distance between nodes, (“small world”),
  - locally dense, globally sparse,
  - expansion properties (conductance),...
- ...include a credible model for agent behaviour guiding the formation of the link structure,
- ...agents should not need global knowledge of the network to determine their link environment.

A good graph model should...

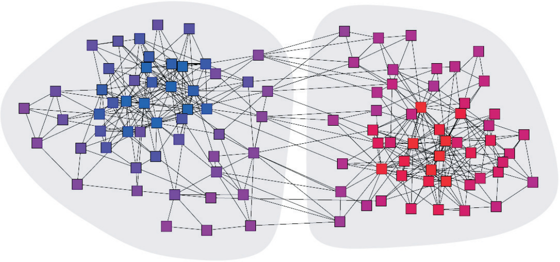
- ...reproduce experimentally observed graph properties:
  - degree distribution follows a power law,
  - small average distance between nodes, ("small world"),
  - locally dense, globally sparse,
  - expansion properties (conductance),...
- ...include a credible model for agent behaviour guiding the formation of the link structure,
- ...agents should not need global knowledge of the network to determine their link environment.

A good graph model should...

- ...reproduce experimentally observed graph properties:
  - degree distribution follows a power law,
  - small average distance between nodes, (“small world”),
  - locally dense, globally sparse,
  - expansion properties (conductance),...
- ...include a credible model for agent behaviour guiding the formation of the link structure,
- ...agents should not need global knowledge of the network to determine their link environment.

Common assumptions in the study of real-life networks

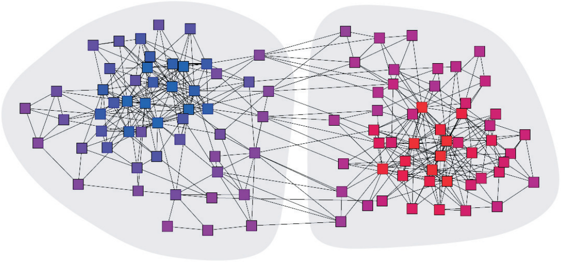
- Communities in a social network can be recognized as densely linked subgraphs.



- Web pages with many common neighbours contain related topics.
- Co-authors usually have similar research interests, etc.

Common assumptions in the study of real-life networks

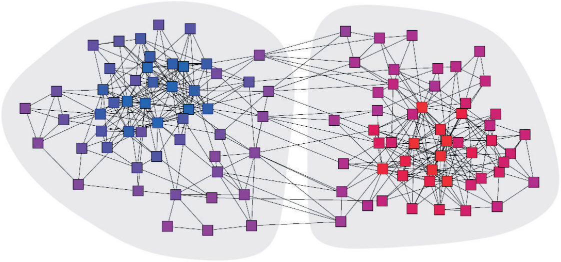
- Communities in a social network can be recognized as densely linked subgraphs.



- Web pages with many common neighbours contain related topics.
- Co-authors usually have similar research interests, etc.

Common assumptions in the study of real-life networks

- Communities in a social network can be recognized as densely linked subgraphs.



- Web pages with many common neighbours contain related topics.
- Co-authors usually have similar research interests, etc.

Underlying metric

Such assumptions, commonly used in experimental and heuristic treatments of real-life networks, imply that there is an a priori “community structure” or “relatedness measure” of the nodes, which is reflected by the link structure of the graph.

The network is a visible manifestation of an underlying hidden reality.

Spatial graph models

- Nodes correspond to points in a (high-dimensional) feature space.
- The metric distance between nodes is a measure of "closeness."
- The edge generation is influenced by the position and relative distance of the nodes.

This gives a basis for reverse engineering: given a graph, and assuming a spatial model, it is possible to estimate the distribution of nodes in the feature space from information contained in the graph structure.

Spatial graph models

- Nodes correspond to points in a (high-dimensional) feature space.
- The metric distance between nodes is a measure of “closeness.”
- The edge generation is influenced by the position and relative distance of the nodes.

This gives a basis for reverse engineering: given a graph, and assuming a spatial model, it is possible to estimate the distribution of nodes in the feature space from information contained in the graph structure.

Outline

1. Introduction
2. Spatial Preferred Attachment (SPA) Model
3. Future work

Spatial Preferred Attachment (SPA) Model

- Nodes are points in *Euclidean space* (randomly and uniformly distributed).

  We let S be the unit hypercube in $\mathbb{R}^m$, equipped with the torus metric derived from any of the L_p norms. This means that for any two points x and y in S,

$$d(x,y) = \min \left\{ ||x - y + u||_p \ : \ u \in \{-1, 0, 1\}^m \right\}.$$

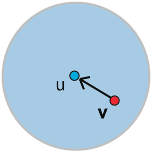
Spatial Preferred Attachment (SPA) Model

- Nodes are points in *Euclidean space* (randomly and uniformly distributed).
- Each node has a *“sphere of influence”* centered at the node. The size is determined by the *in-degree* of the node.

$$|S(v,t)| = \frac{A_1 \deg^{-}(v,t) + A_2}{t}$$

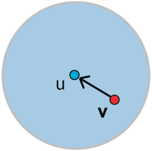
Spatial Preferred Attachment (SPA) Model

- Nodes are points in *Euclidean space* (randomly and uniformly distributed).
- Each node has a *"sphere of influence"* centered at the node. The size is determined by the *in-degree* of the node.
- A new node v can only link to an existing node u if v falls within the sphere of influence of u.



Spatial Preferred Attachment (SPA) Model

- Nodes are points in *Euclidean space* (randomly and uniformly distributed).
- Each node has a *“sphere of influence”* centered at the node. The size is determined by the *in-degree* of the node.
- A new node v can only link to an existing node u if v falls within the sphere of influence of u.
- If v falls into the sphere of influence u, it will link to u with probability p.



Spatial Preferred Attachment (SPA) Model

There are at least three features that distinguish the SPA model from previous models:

- A new node can choose its links purely based on *local* information.
- Since a new node links to each visible node independently, the out-degree is not a constant nor chosen according to a pre-determined distribution, but arises naturally from the model.
- The varying size of the influence regions allows for the occasional *long links*, edges between nodes that are spaced far apart. (This implies a certain "small world" property.)

Spatial Preferred Attachment (SPA) Model

There are at least three features that distinguish the SPA model from previous models:

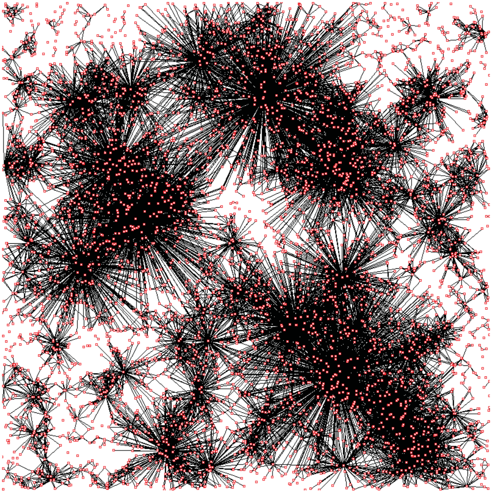
- A new node can choose its links purely based on *local* information.
- Since a new node links to each visible node independently, the out-degree is not a constant nor chosen according to a pre-determined distribution, but arises naturally from the model.
- The varying size of the influence regions allows for the occasional *long links*, edges between nodes that are spaced far apart. (This implies a certain "small world" property.)

Spatial Preferred Attachment (SPA) Model

There are at least three features that distinguish the SPA model from previous models:

- A new node can choose its links purely based on *local* information.
- Since a new node links to each visible node independently, the out-degree is not a constant nor chosen according to a pre-determined distribution, but arises naturally from the model.
- The varying size of the influence regions allows for the occasional *long links*, edges between nodes that are spaced far apart. (This implies a certain "small world" property.)

Spatial Preferred Attachment (SPA) Model



A simulation of the SPA model on the unit square with $t = 5,000$ and $p = 1$

Degree distribution

Power law with exponent $x = 1 + \frac{1}{p}$.

Theorem (Aiello, Bonato, Cooper, Janssen, Prałat)

A.a.s.

$$N(0,t) = (1+o(1))\frac{t}{1+p},$$

and for all k satisfying $1 \le k \le \left(\frac{t}{\log^8 t}\right)^{\frac{p}{4p+2}}$,

$$N(k,t) = (1+o(1))\frac{p^k}{1+p+kp} t \prod_{j=0}^{k-1} \frac{j}{1+p+jp}.$$

(The differential equations method is used.)

A little taste of DEs method

Definition

A **martingale** is a sequence $X_0, X_1, \dots$ of random variables defined on the random process such that

$$\mathbb{E}(X_{n+1} \mid X_0, X_1, \dots, X_n) = X_n.$$

In most applications, the martingale satisfies the property that $\mathbb{E}(X_{n+1} \mid X_0, X_1, \dots, X_n) = \mathbb{E}(X_{n+1} \mid X_n) = X_n$.

Example

Toss a coin n times. Let S_n be the difference between the number of heads and the number of tails after n tosses.

A little taste of DEs method

Definition

A **martingale** is a sequence $X_0, X_1, \ldots$ of random variables defined on the random process such that

$$\mathbb{E}(X_{n+1} \mid X_0, X_1, \ldots, X_n) = X_n.$$

In most applications, the martingale satisfies the property that $\mathbb{E}(X_{n+1} \mid X_0, X_1, \ldots, X_n) = \mathbb{E}(X_{n+1} \mid X_n) = X_n$.

Example

Toss a coin n times. Let S_n be the difference between the number of heads and the number of tails after n tosses.

A little taste of DEs method

Theorem (Hoeffding-Azuma inequality)

Let $X_0, X_1, \ldots$ be a martingale. Suppose that there exist constants $c_k > 0$ such that

$$|X_k - X_{k-1}| \leq c_k$$

for each $k \leq n$. Then, for every $t > 0$,

$$\begin{aligned} \mathbb{P}(X_n \geq \mathbb{E}X_n + t) &\leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right), \\ \mathbb{P}(X_n \leq \mathbb{E}X_n - t) &\leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right). \end{aligned}$$

A little taste of DEs method

$$\mathbb{E}(N(0,t+1) - N(0,t) \mid N(0,t)) = 1 - \frac{N(0,t)pA_2}{t}$$

We first transform $N(0,t)$ into something close to a martingale. It provides some insight if we define real function $f(x)$ to model the behaviour of the scaled random variable $\frac{N(0,xn)}{n}$. If we presume that the changes in the function correspond to the expected changes of random variable, we obtain the following differential equation

$$f'(x) = 1 - f(x)\frac{pA_2}{x}$$

with the initial condition $f(0) = 0$.

A little taste of DEs method

The general solution of this equation can be put in the form

$$f(x)x^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2} = C.$$

Consider the following real-valued function

$$H(x,y) = yx^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2}.$$

(We expect $H(\mathbf{w}_t) = H(t, N(0,t))$ to be close to zero.)

$$\begin{aligned} \mathbb{E}(H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) \mid G_t) &= O(t^{pA_2-1}) \\ |H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t)| &= O(t^{pA_2}\log^2 n). \end{aligned}$$

Use generalized Azuma-Hoeffding inequality: a.a.s.

$$|H(\mathbf{w}_t) - H(\mathbf{w}_{t_0})| = O(n^{1/2+pA_2}\log^3 n).$$

A little taste of DEs method

The general solution of this equation can be put in the form

$$f(x)x^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2} = C.$$

Consider the following real-valued function

$$H(x,y) = yx^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2}.$$

(We expect $H(\mathbf{w}_t) = H(t, N(0,t))$ to be close to zero.)

$$\begin{aligned} \mathbb{E}(H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) \mid G_t) &= O(t^{pA_2-1}) \\ |H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t)| &= O(t^{pA_2} \log^2 n). \end{aligned}$$

Use generalized Azuma-Hoeffding inequality: a.a.s.

$$|H(\mathbf{w}_t) - H(\mathbf{w}_{t_0})| = O(n^{1/2+pA_2} \log^3 n).$$

A little taste of DEs method

The general solution of this equation can be put in the form

$$f(x)x^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2} = C.$$

Consider the following real-valued function

$$H(x,y) = yx^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2}.$$

(We expect $H(\mathbf{w}_t) = H(t, N(0,t))$ to be close to zero.)

$$\begin{aligned} \mathbb{E}(H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) \mid G_t) &= O(t^{pA_2-1}) \\ |H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t)| &= O(t^{pA_2} \log^2 n). \end{aligned}$$

Use generalized Azuma-Hoeffding inequality: a.a.s.

$$|H(\mathbf{w}_t) - H(\mathbf{w}_{t_0})| = O(n^{1/2+pA_2} \log^3 n).$$

A little taste of DEs method

The general solution of this equation can be put in the form

$$f(x)x^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2} = C.$$

Consider the following real-valued function

$$H(x,y) = yx^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2}.$$

(We expect $H(\mathbf{w}_t) = H(t, N(0,t))$ to be close to zero.)

$$\begin{aligned} \mathbb{E}(H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) \mid G_t) &= O(t^{pA_2-1}) \\ |H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t)| &= O(t^{pA_2} \log^2 n). \end{aligned}$$

Use generalized Azuma-Hoeffding inequality: a.a.s.

$$|H(\mathbf{w}_t) - H(\mathbf{w}_{t_0})| = O(n^{1/2+pA_2} \log^3 n).$$

Degree distribution

Out-degree: An important difference between the SPA model and many other models is that the out-degree is not a parameter of the model, but is the result of a stochastic process.

Theorem (Aiello, Bonato, Cooper, Janssen, Prałat)

A.a.s.

$$\max_{0\le i\le t} \deg^+(v_i,t) \ge (1+o(1))p\frac{\log t}{\log\log t}.$$

However, a.a.s. all nodes have out-degree $O(\log^2 t)$.

Theorem (Aiello, Bonato, Cooper, Janssen, Prałat)

A.a.s. $\deg^+(v_t,t) = O(\log^2 t)$.

Degree distribution

Out-degree: An important difference between the SPA model and many other models is that the out-degree is not a parameter of the model, but is the result of a stochastic process.

Theorem (Aiello, Bonato, Cooper, Janssen, Prałat)

A.a.s.

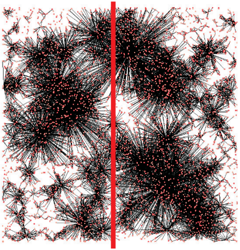
$$\max_{0 \le i \le t} \deg^+(v_i, t) \ge (1 + o(1)) p \frac{\log t}{\log \log t}.$$

However, a.a.s. all nodes have out-degree $O(\log^2 t)$.

Theorem (Aiello, Bonato, Cooper, Janssen, Prałat)

A.a.s. $\deg^+(v_t, t) = O(\log^2 t)$.

Sparse cuts

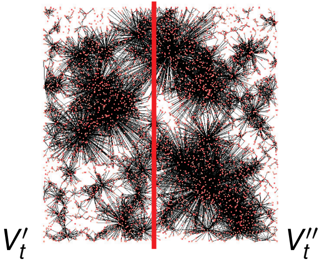


Let us partition the vertex set V_t as follows:

$$V'_t = \left\{ x = (x_1, x_2, \ldots, x_m) \in V_t : x_1 < \frac{1}{2} \right\}$$

and $V''_t = V_t \setminus V'_t$.

Sparse cuts



Theorem (Cooper, Frieze, Prałat)

A.a.s. the following holds $|V'_t| = (1+o(1))t/2$, $|V''_t| = (1+o(1))t/2$, *and*

$$|E(V'_t, V''_t)| = O(t^{\max\{1-1/m, pA_1\}} \log^5 t) = o(t).$$

Diameter

Let $l(v_i, v_j)$ denote the length of the shortest directed path from v_j to v_i if such a path exists, and let $l(v_i, v_j) = 0$ otherwise.

The directed diameter of a graph G_t is defined as

$$D(G_t) = \max_{1 \le i < j \le t} l(v_i, v_j).$$

Diameter

Let $l(v_i, v_j)$ denote the length of the shortest directed path from v_j to v_i if such a path exists, and let $l(v_i, v_j) = 0$ otherwise.

The directed diameter of a graph G_t is defined as

$$D(G_t) = \max_{1 \le i < j \le t} l(v_i, v_j).$$

Theorem (Cooper, Frieze, Prałat)

There exists absolute constant c_1 such that a.a.s.

$$D(G_t) \le c_1 \log t.$$

Diameter

Theorem (Cooper, Frieze, Prałat)

There exists absolute constant c_1 such that a.a.s.

$$D(G_t) \leq c_1 \log t.$$

Theorem (Cooper, Frieze, Prałat)

There exists absolute constant c_2 such that a.a.s.

$$D(G_t) \geq \frac{c_2 \log t}{\log \log t}.$$

(The lower bound requires the additional assumption that $A_1 < 3A_2$, and it is showed for dimension 2 only. However, it can be easily generalized.)

Estimating distnaces

The distance between u and v can be estimated from the graph properties ($cn(u, v, n)$, $\deg^-(u)$ and $\deg^-(v)$).

Theorem (Janssen, Prałat, Wilson)

Theorem 3.1. *Let $\omega = \omega(n)$ be any function tending to infinity together with n. The following holds a.a.s. Let v_k and v_ℓ be vertices such that*

$$k = \deg(v_k, n) \geq \deg(v_\ell, n) = \ell \geq \omega^2 \log n$$

in a graph generated by the SPA model. Let $d = d(v_k, v_\ell)$ be the distance between v_k and v_ℓ in the metric space. Finally, let $T = f^{-1}(\ell/(\omega \log n))$. Then,

Case 1. *If $d \geq \varepsilon(\omega \log n/T)^{1/m}$ for some $\varepsilon > 0$, then*

$$cn(v_\ell, v_k, n) = O(\omega \log n).$$

Case 2. *If $k \geq (1+\varepsilon)\ell$ for some $\varepsilon > 0$ and*

$$d \leq \left(\frac{A_1 k + A_2}{c_m n}\right)^{1/m} - \left(\frac{A_1 \ell + A_2}{c_m n}\right)^{1/m} = \Theta\left(\left(\frac{k}{n}\right)^{1/m}\right), \tag{5}$$

then

$$cn(v_\ell, v_k, n) = (1 + o(1))p\ell.$$

If $k = (1+o(1))\ell$ and $d \ll (k/n)^{1/m} = (1+o(1))(\ell/n)^{1/m}$, then $cn(v_\ell, v_k, n) = (1+o(1))p\ell$ as well.

Case 3. *If $k \geq (1+\varepsilon)\ell$ for some $\varepsilon > 0$ and*

$$\left(\frac{A_1 k + A_2}{c_m n}\right)^{1/m} - \left(\frac{A_1 \ell + A_2}{c_m n}\right)^{1/m} < d \ll (\omega \log n/T)^{1/m}, \tag{6}$$

then

$$cn(v_\ell, v_k, n) = C i_k^{-\frac{(pA_1)^2}{1-pA_1}} i_\ell^{-pA_1} d^{-\frac{mpA_1}{1-pA_1}} \left(1 + O\left(\left(\frac{i_k}{i_\ell}\right)^{pA_1/m}\right)\right), \tag{7}$$

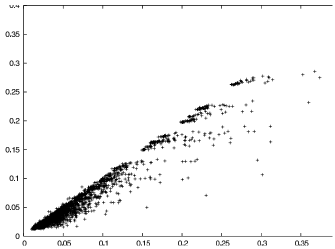
where $i_k = f^{-1}(k)$ and $i_\ell = f^{-1}(\ell)$ and $C = pA_1^{-1} A_2^{\frac{1}{1-pA_1}} c_m^{-\frac{pA_1}{1-pA_1}}$.

If $k = (1+o(1))\ell$ and $\varepsilon(k/n)^{1/m} < d \ll (\omega \log n/T)^{1/m}$ for some $\varepsilon > 0$, then

$$cn(v_\ell, v_k, n) = \Theta\left(i_k^{-\frac{(pA_1)^2}{1-pA_1}} i_\ell^{-pA_1} d^{-\frac{mpA_1}{1-pA_1}}\right).$$

Estimating distnaces

The distance between u and v can be estimated from the graph properties ($cn(u, v, n)$, $\deg^-(u)$ and $\deg^-(v)$).



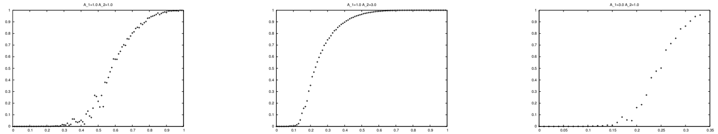
Actual distance vs. estimated distance from simulated data

Outline

Giant component

Conjecture (Cooper, Frieze, Prałat)

$p_3 := (2A_1 + 2A_2)^{-1}$ is the threshold for the giant component.



(a) $A_1 = 1, A_2 = 1$ (b) $A_1 = 1, A_2 = 3$ (c) $A_1 = 3, A_2 = 1$

Conjecture

The clustering coefficient of a vertex of degree k is of order $1/k$.

Common directions

- Adapt the model to specific types of real-world networks
- Find the right parameters for power law exponent etc.
- Validate the model by comparing graph properties
- ‘Social learning in evolving networks’ — design a model with vertices moving

Common directions

- Adapt the model to specific types of real-world networks
- Find the right parameters for power law exponent etc.
- Validate the model by comparing graph properties
- ‘Social learning in evolving networks’ — design a model with vertices moving

Common directions

- Adapt the model to specific types of real-world networks
- Find the right parameters for power law exponent etc.
- Validate the model by comparing graph properties
- ‘Social learning in evolving networks’ — design a model with vertices moving

Common directions

- Adapt the model to specific types of real-world networks
- Find the right parameters for power law exponent etc.
- Validate the model by comparing graph properties
- ‘Social learning in evolving networks’ — design a model with vertices moving

Spatial Preferred Attachment (SPA) Model

- Generalize the model:
  - Node and edge deletion
  - Adding edges to existing nodes
  - Updating the out-links of a node
  - Shifting coordinates ("learning process")
- Undirected graphs
- Non-uniform distribution of points

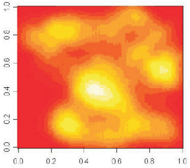
- Use the model to estimate the underlying geometry of the nodes.

Spatial Preferred Attachment (SPA) Model

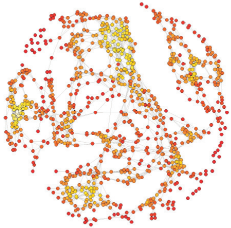
- Generalize the model:
  - Node and edge deletion
  - Adding edges to existing nodes
  - Updating the out-links of a node
  - Shifting coordinates ("learning process")
- Undirected graphs
- Non-uniform distribution of points
- Use the model to estimate the underlying geometry of the nodes.

Spatial Preferred Attachment (SPA) Model

- Generalize the model:
  - Node and edge deletion
  - Adding edges to existing nodes
  - Updating the out-links of a node
  - Shifting coordinates ("learning process")
- Undirected graphs
- Non-uniform distribution of points



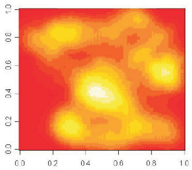
$\Rightarrow$



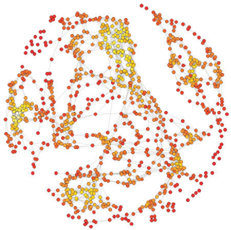
- Use the model to estimate the underlying geometry of the nodes.

Spatial Preferred Attachment (SPA) Model

- Generalize the model:
  - Node and edge deletion
  - Adding edges to existing nodes
  - Updating the out-links of a node
  - Shifting coordinates ("learning process")
- Undirected graphs
- Non-uniform distribution of points



$\Rightarrow$



- Use the model to estimate the underlying geometry of the nodes.

Story 1: Social Learning (BlackBerry)

Consider two homophily hypotheses:

- the likelihood of tie formation between two actors increases with greater similarities in the actors' tastes
- the likelihood of tie deletion between two actors increases with greater differences in the actors' tastes

The role of social influence—third main hypothesis:

- actors tend to adopt the tastes of others they share direct connections with

Story 2: GEO-P model and domination number