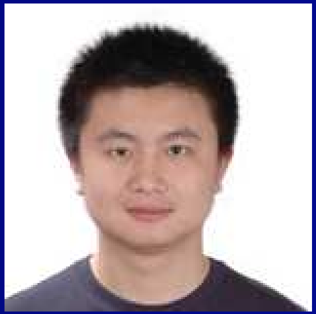
# Statistical Inference for Some Network Models

Aad van der Vaart
Mathematical Institute
Leiden University

Conference on Probability and Statistics in High Dimensions
A Scientific Tribute to Evarist Giné
CRM, Barcelona, June 2016

# Co-author

Fengnan Gao



Remco van der Hofstad

Rui Castro

## Preferential Attachment

Start: complete graph with nodes $1, 2$.

## Preferential Attachment

Start: complete graph with nodes $1, 2$.

Recursions for $n = 2, 3, \ldots$:

given graph with nodes $1, 2, \ldots, n$ with degrees $d_1^{(n)}, \ldots, d_n^{(n)}$,

connect node $n+1$ to node $k \in \{1, \ldots, n\}$ with probability $\propto f(d_k^{(n)})$.

## Preferential Attachment

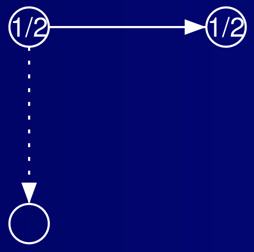
Start: complete graph with nodes $1, 2$.

Recursions for $n = 2, 3, \ldots$:

given graph with nodes $1, 2, \ldots, n$ with degrees $d_1^{(n)}, \ldots, d_n^{(n)}$,

connect node $n+1$ to node $k \in \{1, \ldots, n\}$ with probability $\propto f(d_k^{(n)})$.

f(k)=k



# Preferential Attachment

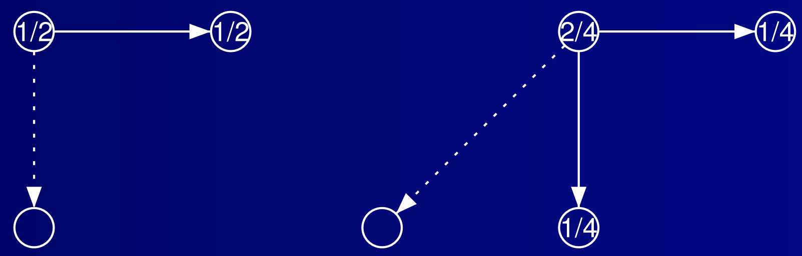
Start: complete graph with nodes $1, 2$.

Recursions for $n = 2, 3, \ldots$:

given graph with nodes $1, 2, \ldots, n$ with degrees $d_1^{(n)}, \ldots, d_n^{(n)}$,

connect node $n+1$ to node $k \in \{1, \ldots, n\}$ with probability $\propto f(d_k^{(n)})$.

f(k)=k



## Preferential Attachment

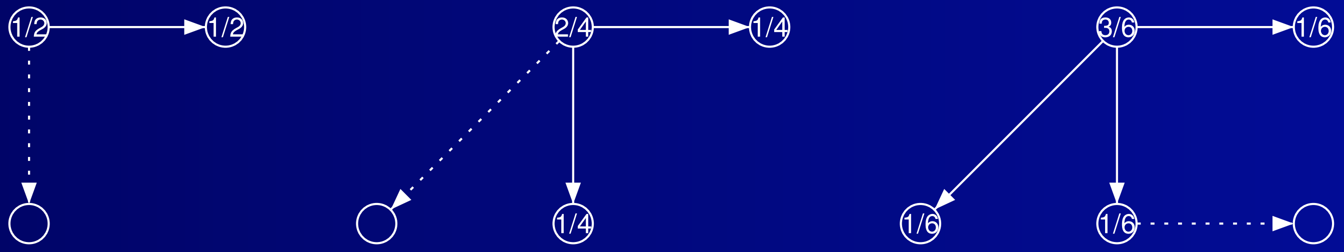
Start: complete graph with nodes $1, 2$.

Recursions for $n = 2, 3, \ldots$:

given graph with nodes $1, 2, \ldots, n$ with degrees $d_1^{(n)}, \ldots, d_n^{(n)}$,

connect node $n+1$ to node $k \in \{1, \ldots, n\}$ with probability $\propto f(d_k^{(n)})$.

f(k)=k



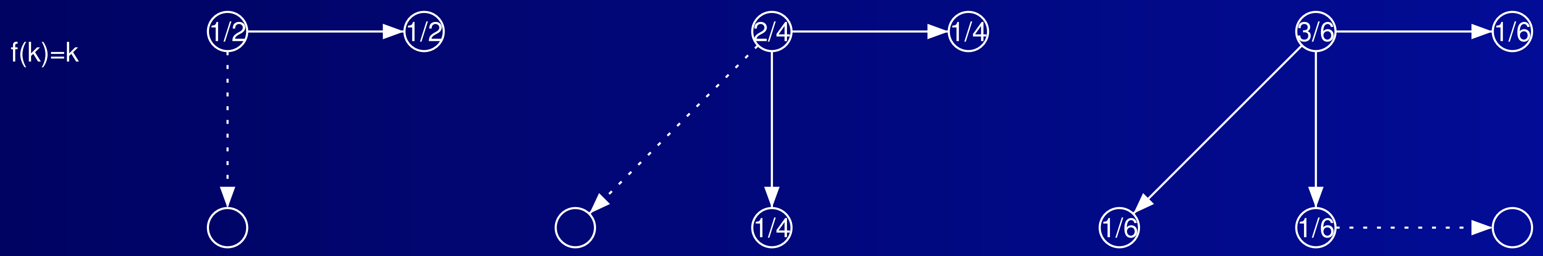
## Preferential Attachment

Start: complete graph with nodes $1, 2$.

Recursions for $n = 2, 3, \ldots$:

given graph with nodes $1, 2, \ldots, n$ with degrees $d_1^{(n)}, \ldots, d_n^{(n)}$,

connect node $n + 1$ to node $k \in \{1, \ldots, n\}$ with probability $\propto f(d_k^{(n)})$.

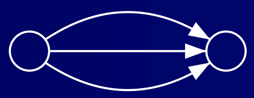



Estimate the attachment function $f\colon \mathbb{N} \to [0, \infty)$ from observed network

## Preferential Attachment with Random Initial Degree

For i.i.d. $m_1, m_2, \ldots$, connect node $n$ to $m_n$ existing nodes.

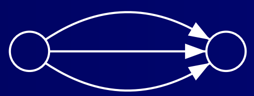
Start: graph with nodes $1, 2$ and $m_1$ edges.



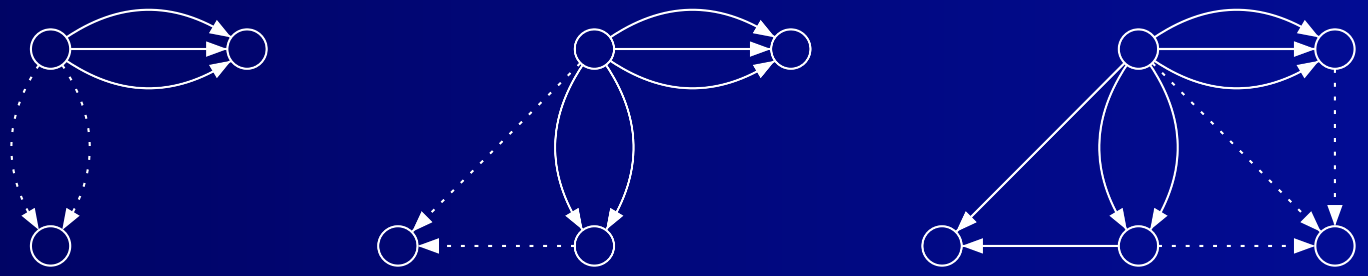
## Preferential Attachment with Random Initial Degree

For i.i.d. $m_1, m_2, \ldots$, connect node $n$ to $m_n$ existing nodes.

Start: graph with nodes $1, 2$ and $m_1$ edges.



Recursions for $n = 2, 3, \ldots$:
for $i = 1, \ldots, m_n$,
given graph with nodes $1, 2, \ldots, n$ with current degrees $d_1^{(n,i-1)}, \ldots, d_n^{(n,i-1)}$,
connect node $n+1$ to node $k \in \{1, \ldots, n\}$ with probability $\propto f(d_k^{(n,i-1)})$.



## Degree distribution

$$p_k(n) := \frac{1}{n} \#\text{nodes } 1, 2, \ldots, n \text{ of degree } k.$$

## Degree distribution

$$p_k(n) := \frac{1}{n}\#\text{nodes } 1, 2, \ldots, n \text{ of degree } k.$$

THEOREM [Barabasi-Albert, 2000; Mori 2002; Rudas, Toth, Valko, 2007]

$$p_k(n) \to p_k := \frac{\alpha}{\alpha + f(k)} \prod_{j=1}^{k-1} \frac{f(j)}{\alpha + f(j)}, \qquad \text{a.s..}$$

($\alpha$ makes $(p_k)$ a probability distribution on $\mathbb{N}$.)

EXAMPLE

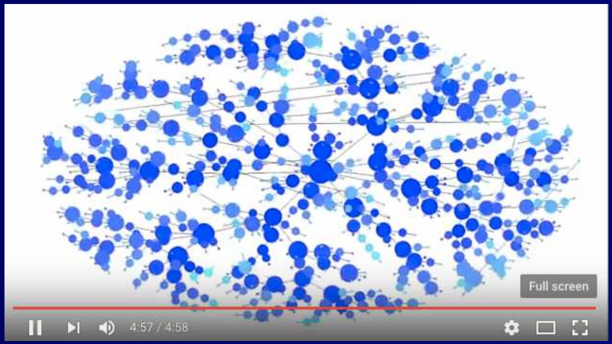
| | | |
|---|---|---|
| $f(k) = k$ | [Barabasi, Albert, 99] | $p_k = 4/(k(k+1)(k+2))$. |
| $f(k) = k + \delta$ | | $p_k \sim k^{-3-\delta}$. |
| $f(k) = k^\beta$, $\beta \in [1/2, 1)$ | [Krapivsky, Redner, 01] | $p_k = k^{c_1} e^{-c_2 k^{1-\beta}}$. |

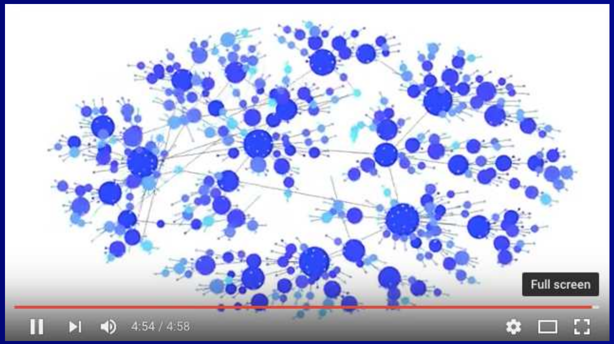
# Preferential Attachment with $f(k) = k$

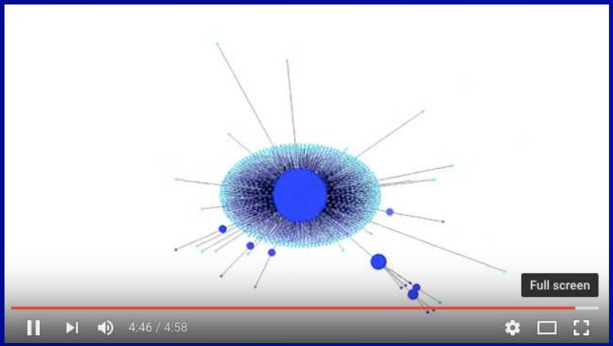
start movie

Movies by Matjaz Perc downloaded from Youtube

# Preferential Attachment with $f(k) = k^{0.25}$ or $f(k) = k$ or $f(k) = k^2$








From movies by Matjaz Perc downloaded from Youtube

## Empirical Estimator

$$\hat{f}_n(k) = \frac{p_{>k}(n)}{p_k(n)}$$

Motivation: $\#$ nodes of degree $> k$ at time $t$
$= \#$ of times up to time $t$ that the new node chose a node of degree $k$.

## Empirical Estimator

$$\hat{f}_n(k) = \frac{p_{>k}(n)}{p_k(n)}$$

Motivation: $\#$ nodes of degree $> k$ at time $t$
$= \#$ of times up to time $t$ that the new node chose a node of degree $k$.

So, for large $t$:

$$\begin{aligned} p_{>k}(n) &\approx \mathrm{P}(\text{node } t+1 \text{ connects to node of degree } k) \\ &\propto f(k)p_k(t) \approx f(k)p_k(n). \end{aligned}$$

## Empirical Estimator

$$\hat{f}_n(k) = \frac{p_{>k}(n)}{p_k(n)}$$

Motivation: $\#$ nodes of degree $> k$ at time $t$
$= \#$ of times up to time $t$ that the new node chose a node of degree $k$.

So, for large $t$:

$$p_{>k}(n) \approx \mathrm{P}(\text{node } t+1 \text{ connects to node of degree } k)$$
$$\propto f(k)p_k(t) \approx f(k)p_k(n).$$

THEOREM [Gao,vdV]

$\hat{f}_n(k) \to f(k)/\sum_j f(j)p_j$ a.s. as $n \to \infty$, for every fixed $k$.

Proof is based on LLN of supercritical branching processes by Jagers, 1975 and Nerman, 1981, along the lines of Rudas, Toth, Valko, 2008.

## Supercritical Branching

Individual $x$ born at time $\sigma_x$ has children at times of counting process $(\xi_x(t-\sigma_x)\colon t \geq \sigma_x)$.
For given numerical time-dependent characteristic $(\phi_x(t-\sigma_x)\colon t \geq \sigma_x)$:

$$Z_t^\phi := \sum_{x:\sigma_x \leq t} \phi_x(t-\sigma_x).$$

If $(\xi_x, \phi_x, \psi_x)$ are i.i.d. $\sim (\xi, \phi, \psi)$ and suitably integrable, then

$$\frac{Z_t^\phi}{Z_t^\psi} \to \frac{\int e^{-\alpha t} \mathrm{E}\phi(t)\, dt}{\int e^{-\alpha t} \mathrm{E}\psi(t)\, dt}, \qquad \text{a.s.,}$$

for $\alpha$ the "Malthusian parameter": $\int e^{-\alpha t}\, \mu(dt) = 1$, for $\mu(t) = \mathrm{E}\xi(t)$.

In fact $e^{-\alpha t} Z_t^\phi$ converges to a random limit.

EXAMPLE $\phi(t) = 1_{t \geq 0}$ gives $Z_t^\phi = \#(x\colon \sigma_x \leq t)$.

## Maximum Likelihood

Growing the graph is (nonstationary) Markov.

The log likelihood for observing the full evolution up to time $n$ is

$$f \mapsto \log \prod_{t=3}^{n} \frac{f(d_t)}{S_f(t)} = \sum_{k=1}^{\infty} \log f(k) N_{>k}(n) - \sum_{t=3}^{n} \log S_f(t),$$

where

$$d_t = \text{ degree of the node to which node } t+1 \text{ is attached,}$$

$$S_f(t) = tf(1) + \sum_{i=2}^{t-1} \big(f(d_i+1) - f(d_i)\big) = \sum_{k=1}^{\infty} f(k) t p_k(t).$$

## Maximum Likelihood

Growing the graph is (nonstationary) Markov.

The log likelihood for observing the full evolution up to time $n$ is

$$f \mapsto \log \prod_{t=3}^{n} \frac{f(d_t)}{S_f(t)} = \sum_{k=1}^{\infty} \log f(k) N_{>k}(n) - \sum_{t=3}^{n} \log S_f(t),$$

where

$$d_t = \text{ degree of the node to which node } t+1 \text{ is attached,}$$

$$S_f(t) = tf(1) + \sum_{i=2}^{t-1} \big( f(d_i + 1) - f(d_i) \big) = \sum_{k=1}^{\infty} f(k) t p_k(t).$$

The degree sequence $d_3, d_4, \ldots, d_n$ is sufficient.

## Maximum Likelihood in the Affine Case $f_\delta(k) = k + \delta$

$$S_{f_\delta}(t) = \sum_{k=1}^{\infty} (k+\delta) t p_k(t) = 2t + t\delta$$

The log likelihood for observing the full evolution up to time $n$ is

$$\delta \mapsto \log \prod_{t=3}^{n} \frac{f_\delta(d_t)}{S_{f_\delta}(t)} = \sum_{k=1}^{\infty} \log(k+\delta) N_{>k}(n) - \sum_{t=3}^{n} \log(2t + t\delta),$$

## Maximum Likelihood in the Affine Case $f_\delta(k) = k + \delta$

$$S_{f_\delta}(t) = \sum_{k=1}^{\infty} (k+\delta) t p_k(t) = 2t + t\delta$$

The log likelihood for observing the full evolution up to time $n$ is

$$\delta \mapsto \log \prod_{t=3}^{n} \frac{f_\delta(d_t)}{S_{f_\delta}(t)} = \sum_{k=1}^{\infty} \log(k+\delta) N_{>k}(n) - \sum_{t=3}^{n} \log(2t + t\delta),$$

Observation of the graph at time $n$ is sufficient for the full evolution.

## Maximum Likelihood in the Affine Case $f_\delta(k) = k + \delta$

$$S_{f_\delta}(t) = \sum_{k=1}^{\infty}(k+\delta)tp_k(t) = 2t + t\delta$$

The log likelihood for observing the full evolution up to time $n$ is

$$\delta \mapsto \log \prod_{t=3}^{n} \frac{f_\delta(d_t)}{S_{f_\delta}(t)} = \sum_{k=1}^{\infty} \log(k+\delta) N_{>k}(n) - \sum_{t=3}^{n} \log(2t + t\delta),$$

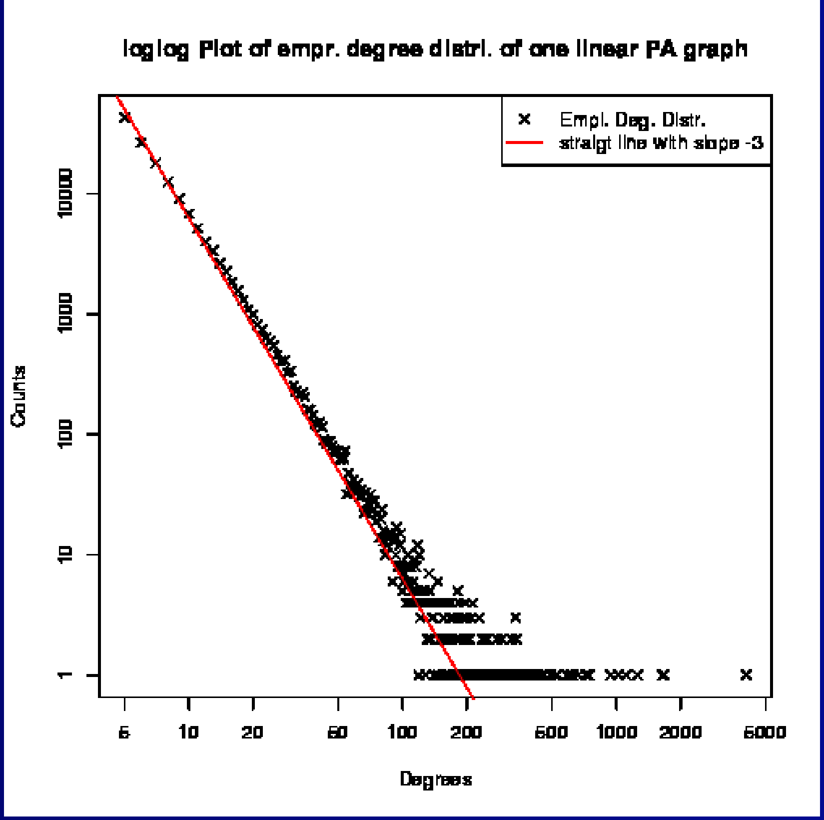
THEOREM [Gao, vdV]

The model is locally asymptotically normal in parameter $\delta$ and

$$\sqrt{n}(\hat{\delta}_n - \delta) \rightsquigarrow N(0, i_\delta^{-1}), \qquad i_\delta = \sum_{k=1}^{\infty} \frac{\mu(k+\delta)p_{\delta,k}}{(k+\delta)^2(2\mu+\delta)} - \frac{\mu}{(2\mu+\delta)^2}.$$

$\mu$ = mean initial degree distribution

Proof uses the martingale central limit theorem.




$\log p_k(n)$ (vertical) versus $\log k$ for single realization with $n = 150000$ and $m = 5$.

$$p_k \sim k^{-3-\delta}, \qquad k \to \infty.$$

## Maximum Likelihood in the General Parametric Case $f_\theta$, for $\theta \in \mathbb{R}^d$

The log likelihood for observing the full evolution up to time $n$ is

$$\theta \mapsto \log \prod_{t=3}^{n} \frac{f_\theta(d_t)}{S_{f_\theta}(t)} = \sum_{k=1}^{\infty} \log f_\theta(k) N_{>k}(n) - \sum_{t=3}^{n} \log S_{f_\theta}(t),$$

THEOREM

Under general conditions on $f_\theta$ the model of observing the full evolution up to time $n$ is locally asymptotically normal with respect to $\theta$ and

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow N(0, i_\theta^{-1}), \qquad i_\theta = \sum_{k=1}^{\infty} \frac{\dot{f}_\theta}{f_\theta}(k) p_{\theta,>k} - \frac{\sum_{k=1}^{\infty} \dot{f}_\theta(k) p_{\theta,k}}{\sum_{k=1}^{\infty} f_\theta(k) p_{\theta,k}}.$$

Proof uses the martingale central limit theorem and the LLN for supercritical branching processes.

## Maximum Likelihood in the General Parametric Case $f_\theta$, for $\theta \in \mathbb{R}^d$

The log likelihood for observing the full evolution up to time $n$ is

$$\theta \mapsto \log \prod_{t=3}^{n} \frac{f_\theta(d_t)}{S_{f_\theta}(t)} = \sum_{k=1}^{\infty} \log f_\theta(k) N_{>k}(n) - \sum_{t=3}^{n} \log S_{f_\theta}(t),$$

THEOREM

Under general conditions on $f_\theta$ the model of observing the full evolution up to time $n$ is locally asymptotically normal with respect to $\theta$ and

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow N(0, i_\theta^{-1}), \qquad i_\theta = \sum_{k=1}^{\infty} \frac{\dot{f}_\theta}{f_\theta}(k) p_{\theta,>k} - \frac{\sum_{k=1}^{\infty} \dot{f}_\theta(k) p_{\theta,k}}{\sum_{k=1}^{\infty} f_\theta(k) p_{\theta,k}}.$$

Proof uses the martingale central limit theorem and the LLN for supercritical branching processes.

EXAMPLE ?? $f_\theta(k) = (k+\delta)^\beta$, for $\theta = (\delta, \beta)$.

## Some Open Questions

Is observing the graph at time $n$ asymptotically sufficient for observing the full evolution up to time $n$?

## Some Open Questions

Is observing the graph at time $n$ asymptotically sufficient for observing the full evolution up to time $n$?

Does the maximum likelihood estimator based on observing the graph at time $n$ behave the same as the maximum likelihood estimator based on the full evolution?

## Some Open Questions

Is observing the graph at time $n$ asymptotically sufficient for observing the full evolution up to time $n$?

Does the maximum likelihood estimator based on observing the graph at time $n$ behave the same as the maximum likelihood estimator based on the full evolution?

Is the empirical estimator asymptotically normal?

## Some Open Questions

Is observing the graph at time $n$ asymptotically sufficient for observing the full evolution up to time $n$?

Does the maximum likelihood estimator based on observing the graph at time $n$ behave the same as the maximum likelihood estimator based on the full evolution?

Is the empirical estimator asymptotically normal?

Can we estimate $f$ under nonparametric shape constraints?