# TREC Evaluation 2008-2011

**Marc Lelarge**

**INRIA & ENS**

Rungis, March 2012.

# DISCRETE GEOMETRY: RANDOM GRAPHS

- Mathematical tools:

- Applications:

# DISCRETE GEOMETRY: RANDOM GRAPHS

- Mathematical tools:
  - Cavity method
  - Configuration model
- Applications:

# DISCRETE GEOMETRY: RANDOM GRAPHS

- Mathematical tools:
  - Cavity method
  - Configuration model
- Applications:
  - Pure Mathematics:
    - Random matrices
    - Combinatorial optimization
    - First passage percolation
  - Communication Networks:
    - Load balancing
  - Network economics:
    - Diffusions in social networks
    - Control of epidemics
    - Mean field games for security

# DISCRETE GEOMETRY: RANDOM GRAPHS

- Mathematical tools:
  - Cavity method
  - Configuration model
- Applications:
  - Pure Mathematics:
    - Random matrices (SODA, RSA, AoP)
    - Combinatorial optimization (AIHP, SIAMJComp, PTRF)
    - First passage percolation
  - Communication Networks:
    - Load balancing (SODA, SIGMETRICS)
  - Network economics:
    - Diffusions in social networks (GEB)
    - Control of epidemics (SIGMETRICS)
    - Mean field games for security (INFOCOM(3), SIGMETRICS)

# DISCRETE GEOMETRY: RANDOM GRAPHS

- Mathematical tools:
  - Cavity method
  - Configuration model
- Applications:
  - Pure Mathematics:
    - Random matrices C. Bordenave, J. Salez
    - Combinatorial optimization D. Aldous, C. Bordenave, J. Salez
    - First passage percolation M. Draief, H. Amini
  - Communication Networks:
    - Load balancing M. Leconte, L. Massoulié
  - Network economics:
    - Diffusions in social networks E. Coupechoux
    - Control of epidemics
    - Mean field games for security J. Bolot

# WEIGHTED DIAMETER

Graph $G = (V, E)$:

- Distance $\text{dist}(a, b) = \min_{\pi \in \Pi(a,b)} |\pi|$, the number of edges in $E$ in the shortest path connecting $a$ and $b$.

- Diameter of $G$ defined by:

$$\text{diam}(G) = \max\{\text{dist}(a, b),\ a, b \in V,\ \text{dist}(a, b) < \infty\}.$$

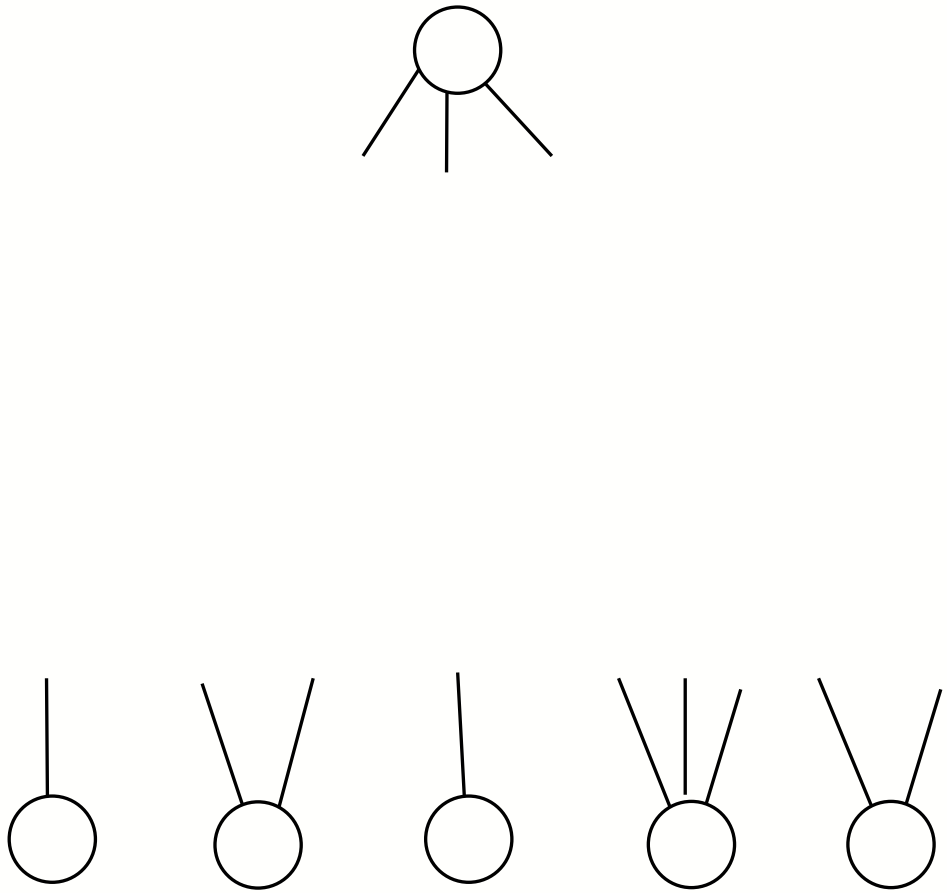
- Weight associated to each edge $e \in E$: $w_e$.

- Weighted distance $\text{dist}_w(a, b) = \min_{\pi \in \Pi(a,b)} \sum_{e \in \pi} w_e$.

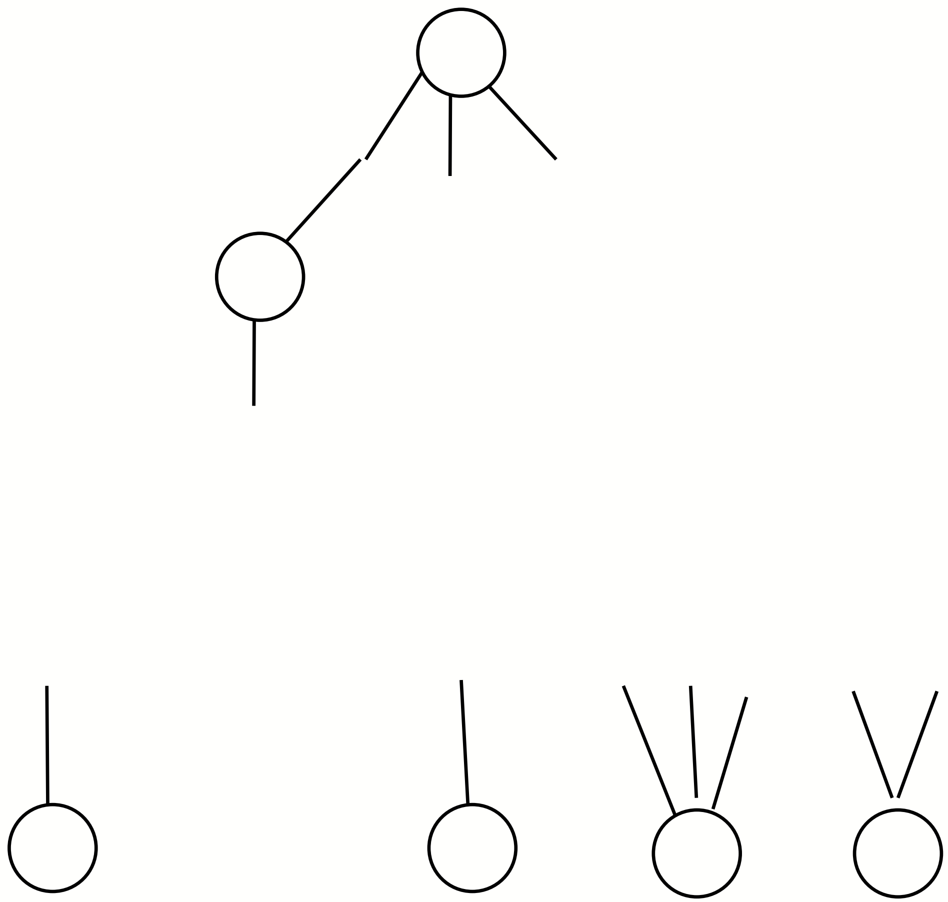
- Weighted diameter of $G$ defined by:

$$\text{diam}_w(G) = \max\{\text{dist}_w(a, b),\ a, b \in V,\ \text{dist}_w(a, b) < \infty\}\ .$$

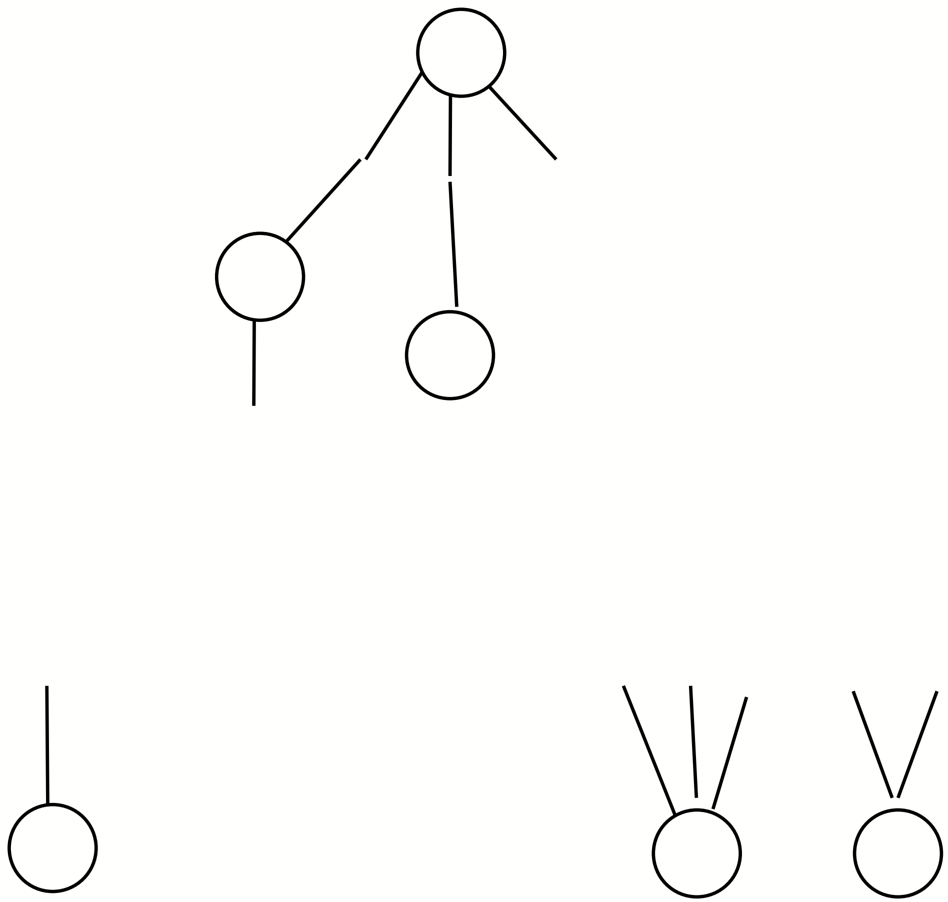
# EXPLORATION PROCESS



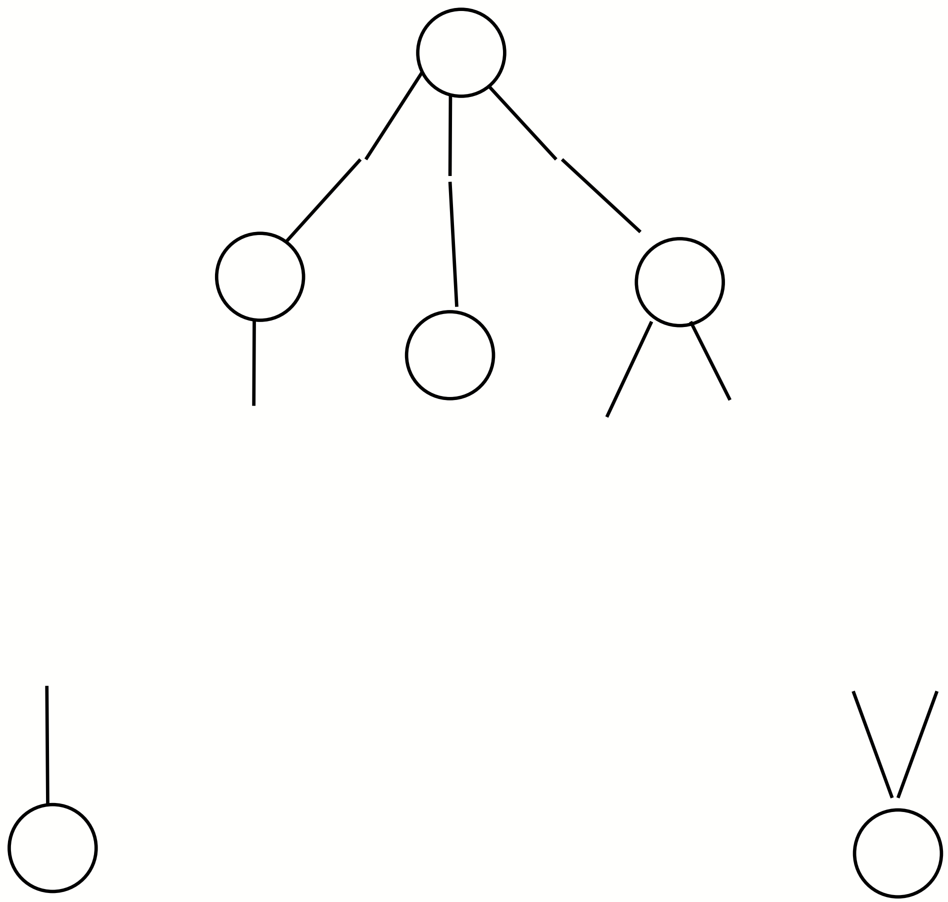
# EXPLORATION PROCESS



# EXPLORATION PROCESS



# EXPLORATION PROCESS



# BRANCHING PROCESS APPROXIMATION

The first individual has offspring distribution $\{p_k\}$.

The other individuals have offspring distribution $\{q_k\}$.

Let $\{q_k\}_{k=0}^{\infty}$ the size-biased probability mass function corresponding to $\{p_k\}$, by

$$q_k = \frac{(k+1)p_{k+1}}{\lambda} \text{ , and, } \nu = \sum_{k=0}^{\infty} k q_k \in (0, \infty).$$

The mean of the size of generation $k$ is $\lambda \nu^{k-1}$.

The condition $\nu > 1$ is equivalent to the existence of a giant component.

# TYPICAL GRAPH DISTANCE

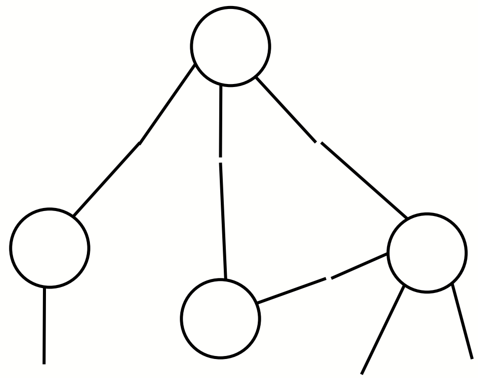
**Theorem 1.** *For $a$ and $b$ chosen uniformly at random in the giant component of $G(n, (d_i)_1^n)$, we have*

$$\frac{\textit{dist}(a,b)}{\log n} \xrightarrow{p} \frac{1}{\log \nu}.$$

Van der Hofstad, Hooghiemstra, Van Mieghem 2005

# A SIMPLE HEURISTIC

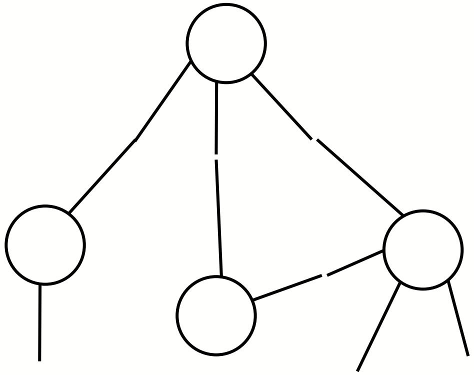
Let $Z_k^{(1)}$ be the number of free half-edges in the ball $B(a,k) = \{i, \text{dist}(1,i) \leq k\}$.
$Z_0^{(1)}$ is the degree of node $1$.



By the branching process approximation, $Z_k^{(1)}$ is close to $\lambda\nu^{k-1}$.

# A SIMPLE HEURISTIC

Let $Z_k^{(1)}$ be the number of free half-edges in the ball $B(a, k) = \{i, \text{dist}(1, i) \leq k\}$.
$Z_0^{(1)}$ is the degree of node $1$.

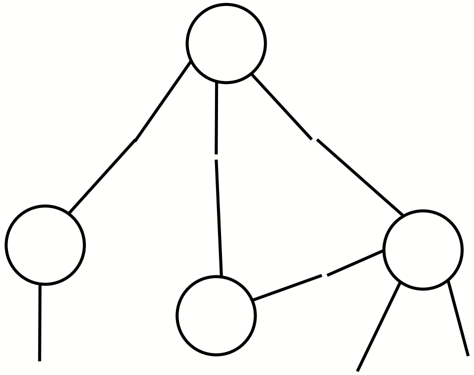


By the branching process approximation, $Z_k^{(1)}$ is close to $\lambda \nu^{k-1}$.

A free half-edge of $Z_k^{(1)}$ is attached to a free half-edge of $Z_k^{(2)}$ with positive probability if $Z_k^{(1)} Z_k^{(2)}$ is of order the total number of free half-edges left after $k$ exploration steps.

# A SIMPLE HEURISTIC

Let $Z_k^{(1)}$ be the number of free half-edges in the ball $B(a,k) = \{i, \text{dist}(1,i) \leq k\}$.
$Z_0^{(1)}$ is the degree of node $1$.



By the branching process approximation, $Z_k^{(1)}$ is close to $\lambda\nu^{k-1}$.

A free half-edge of $Z_k^{(1)}$ is attached to a free half-edge of $Z_k^{(2)}$ with positive probability if $Z_k^{(1)} Z_k^{(2)}$ is of order the total number of free half-edges left after $k$ exploration steps.

Take $k \approx \frac{1}{2}\frac{\log n}{\log \nu}$, then $Z_k^{(1)} \approx Z_k^{(2)} \approx \sqrt{n}$ and the number of free half-edges is $\approx n - 2\sqrt{n} \approx n$.

The typical distance between $1$ and $2$ is $\approx 2k = \frac{\log n}{\log \nu}$.

# WEIGHTED DIAMETER

**Theorem 2.** *Consider a random graph $G(n,(d_i)_1^n)$ with i.i.d. exponential $1$ weights on its edges, then*

$$\frac{\textit{diam}_w(G(n,(d_i)_1^n))}{\log n} \xrightarrow{p} \frac{1}{\nu-1} + \frac{2}{d_{\min}}\mathbf{1}_{(d_{\min}\geq 3)} + \frac{\mathbf{1}_{(d_{\min}=2)}}{1-q_1} + \frac{2}{1-\beta_*}\mathbf{1}_{(d_{\min}=1)}.$$

Ding, Han Kim, Lubetzky, Peres 2010 (random regular graphs)

Amini, Draief, L. 2011

Key ideas:

- Coupling with a continuous time branching process
- Large deviations for split times

# WEIGHTED DIAMETER

**Theorem 3.** *Consider a random graph $G(n,(d_i)_1^n)$ with i.i.d. exponential $1$ weights on its edges, then*

$$\frac{\text{diam}_w(G(n,(d_i)_1^n))}{\log n} \xrightarrow{p} \frac{1}{\nu-1} + \frac{2}{d_{\min}}\mathbf{1}_{(d_{\min}\geq 3)} + \frac{\mathbf{1}_{(d_{\min}=2)}}{1-q_1} + \frac{2}{1-\beta_*}\mathbf{1}_{(d_{\min}=1)}.$$

Ding, Han Kim, Lubetzky, Peres 2010 (random regular graphs)

Amini, Draief, L. 2011

Key ideas:

- Coupling with a continuous time branching process
- Large deviations for split times

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{\log n}\log\mathbb{P}\left(\infty > T_n^k \geq \left(x+\frac{1}{\nu-1}\right)\log n\right) &= -xg(\xi_{\min},k) \\
\mathbb{P}\left(T_n^k < \left(\frac{1-x}{\nu-1}\right)\log n\right) &= o\left(n^C e^{-n^x}\right).
\end{aligned}$$

# Game-theoretic contagion model

| Situation | Payoff (for both users) |
|---|---|
| | $q$ |
| | $1 - q > q$ |
| | $0$ |

Game on a network of interconnected players Blume 1995, Morris 2000

- Total payoff = sum of payoffs from all your neighbors
- Switch from to $\Leftrightarrow \frac{|\text{Neighbors using Skype}|}{|\text{Neighbors}|} > q.$

## Contagion in networks

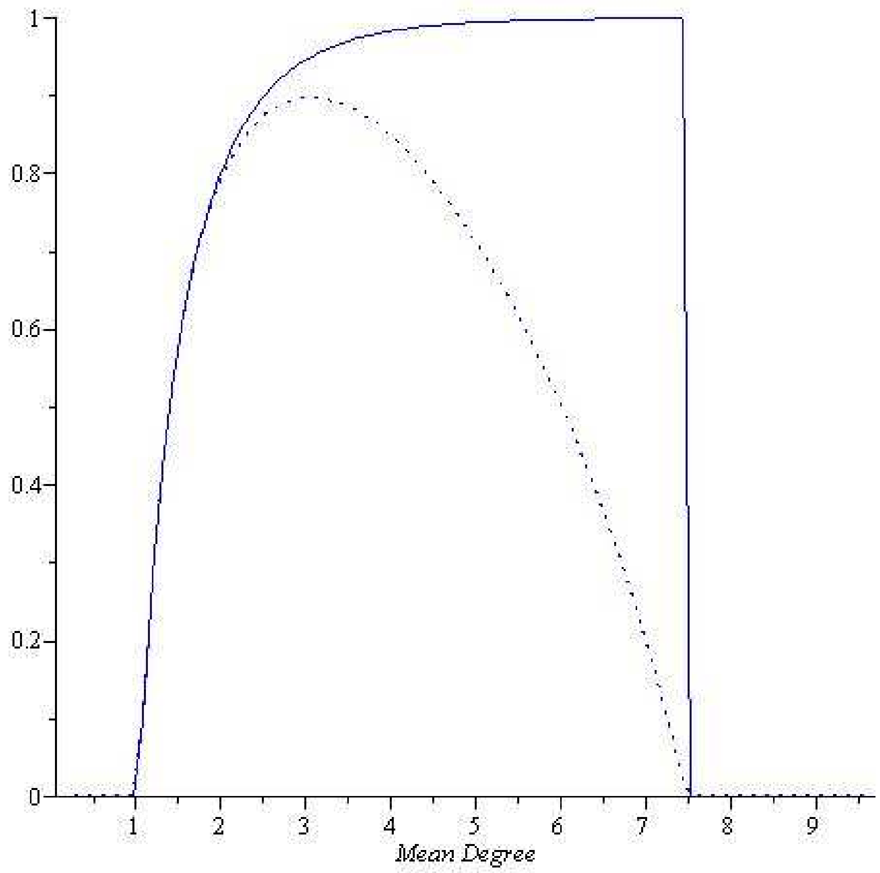
Parameter $q$ varies:

$$q \text{ small} \quad \Rightarrow \quad \boxed{\text{CASCADE}}$$

$$q \text{ higher} \quad \Rightarrow \quad \text{NO cascade}$$

$$\text{Contagion threshold } q_c^{(G)} := \sup \left\{ q \;\middle|\; \text{CASCADE in } G \text{ for parameter } q \right\}$$

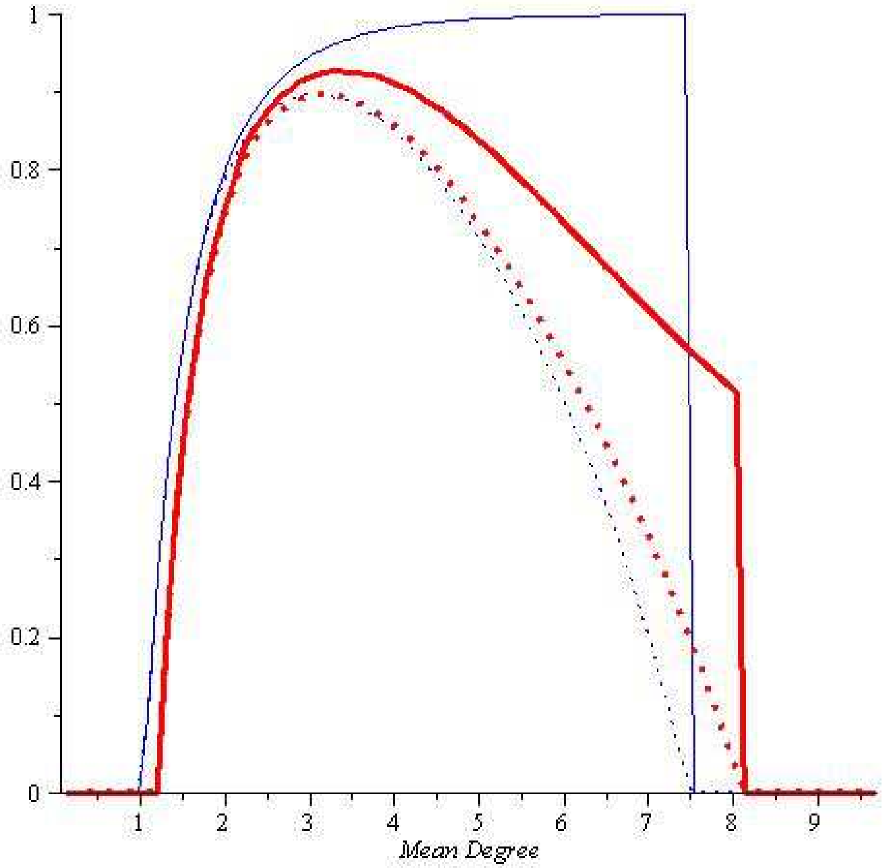
# Contagion in random networks: impact of connectivity




· · · Pivotal players in the random graph

— Cascade size in the random graph

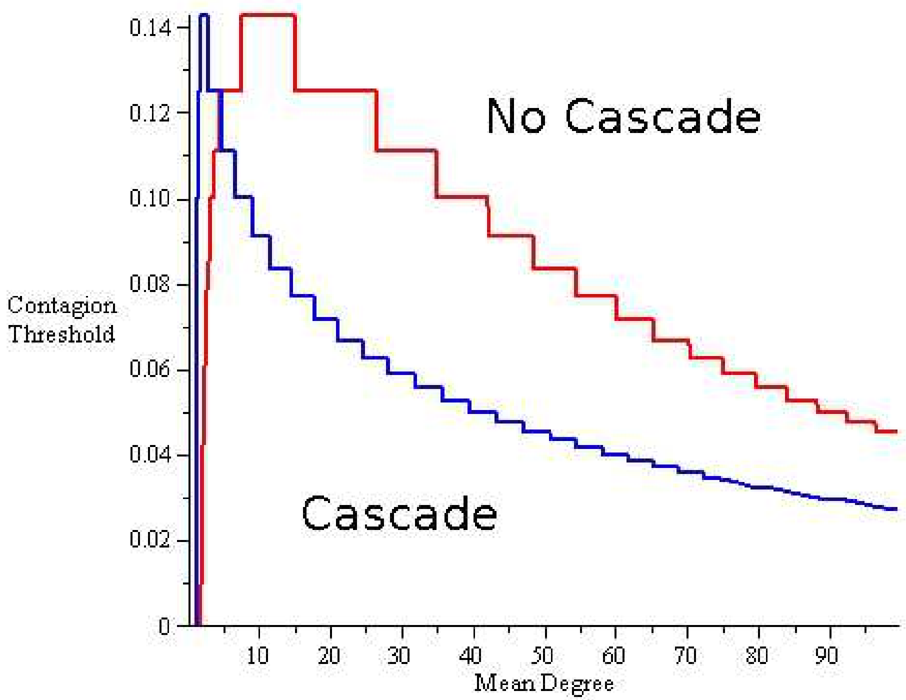
# Contagion in random networks with clustering




joint work with E. Coupechoux

# Contagion in random networks with clustering

Graphs with the SAME asymptotic degree distribution: $\tilde{p}_k \propto k^{-\tau} e^{-k/50}$




— Graph with clustering

— Graph with no clustering

joint work with E. Coupechoux

# FUTURE

- Mathematical tools:
  - Cavity method
  - Configuration model
- Applications:
  - Pure Mathematics:
    - Random matrices
    - Combinatorial optimization

    $\Big\} \longrightarrow$ Statistical Inference
    - First passage percolation $\longrightarrow$ stochastic order, Bartek
  - Communication Networks:
    - Load balancing
    - Queueing, Loss Networks $\longrightarrow$ Ana
  - Network economics:
    - Diffusions in social networks
    - Control of epidemics
    - Mean field games for security