CentraleSupélec

ESSEC BUSINESS SCHOOL

# Network Science Analytics

## Option Applied Math and M.Sc. in DSBA

Lecture 3A

Time-evolving graphs and network models

Fragkiskos Malliaros

Friday, February 2, 2018

# Announcements

- First assignment will be out tomorrow
  - Complete the assignment individually
  - Due: Feb 25, 23:00
- Project
  - Teams of 3 or 4 students (preferably 4)
  - Proposal due on Feb 10

# Acknowledgements


- The lecture is partially based on material by
  - Jure Leskovec, Stanford University
  - Aaron Clauset, CU Boulder
  - Manos Papagelis, York University
  - Gonzalo Mateos, University of Rochester
  - Albert-László Barabási, Northeastern University
  - Christos Faloutsos, CMU
  - Danai Koutra, University of Michigan
  - R. Zafarani, M. A. Abbasi, and H. Liu, Social Media Mining: An Introduction, Cambridge University Press, 2014. Free book and slides at http://socialmediamining.info/

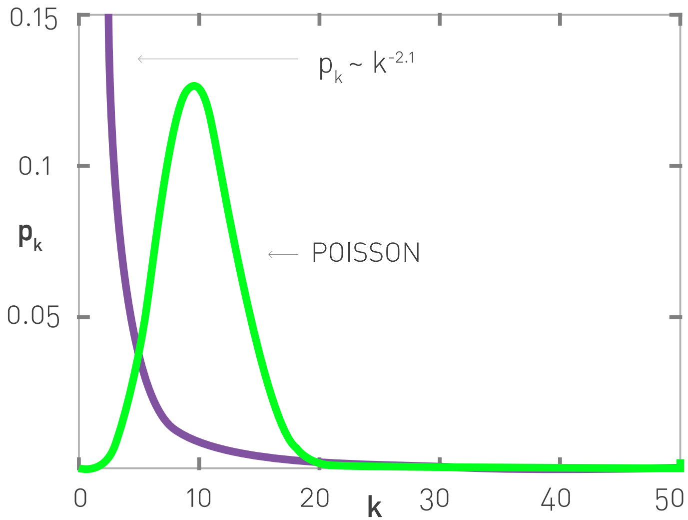
Thank you!

# Last Lecture

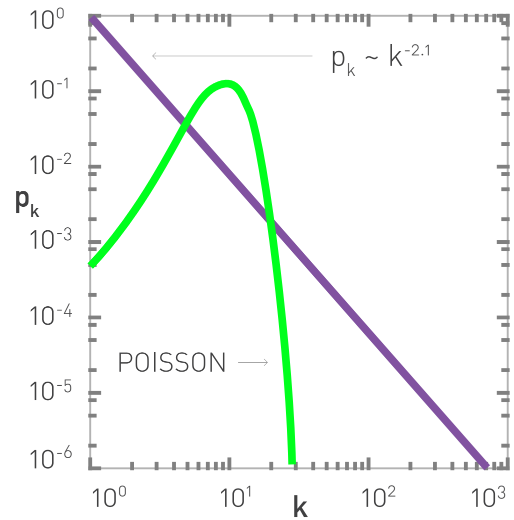
## Power-law degree distribution and the Preferential Attachment model

# Power-Law vs. Exponential Degree Distributions

(a)



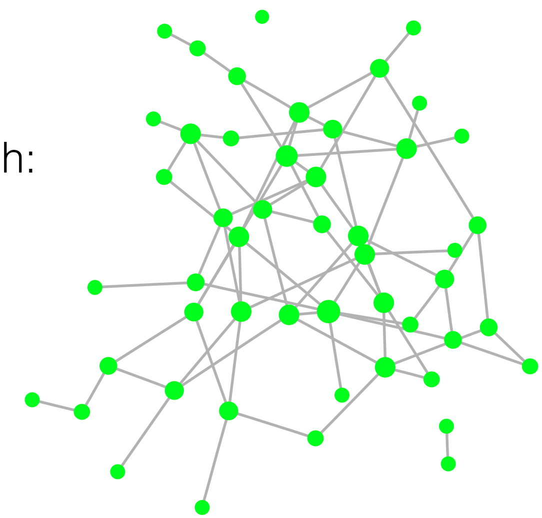

(b)



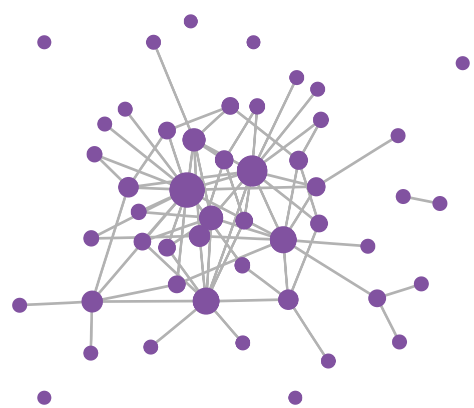

*$\langle k\rangle=11$*

(c)

Random graph:
*$\langle k\rangle = 3$*
*$n = 50$*



(d)



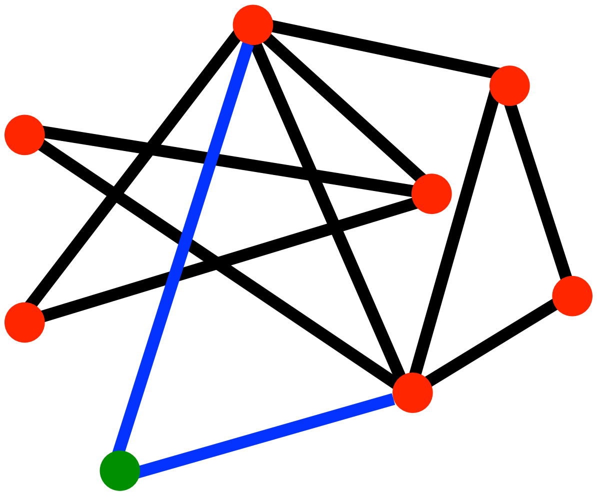
Power-law degree distr. graph:
*$\langle k\rangle = 3$*
*$\alpha = 2.1$*

# Model: Preferential attachment

- Preferential attachment
  - [Start with a small number $m_0$ of nodes]
  - Nodes arrive in order $1,2,\ldots,n$
  - At step $j$, let $k_i$ be the degree of node $i < j$
  - A new node $j$ arrives and creates $m$ links
  - Prob. of $j$ linking to a previous node $i$ is proportional to the degree $d_i$ of node $i$

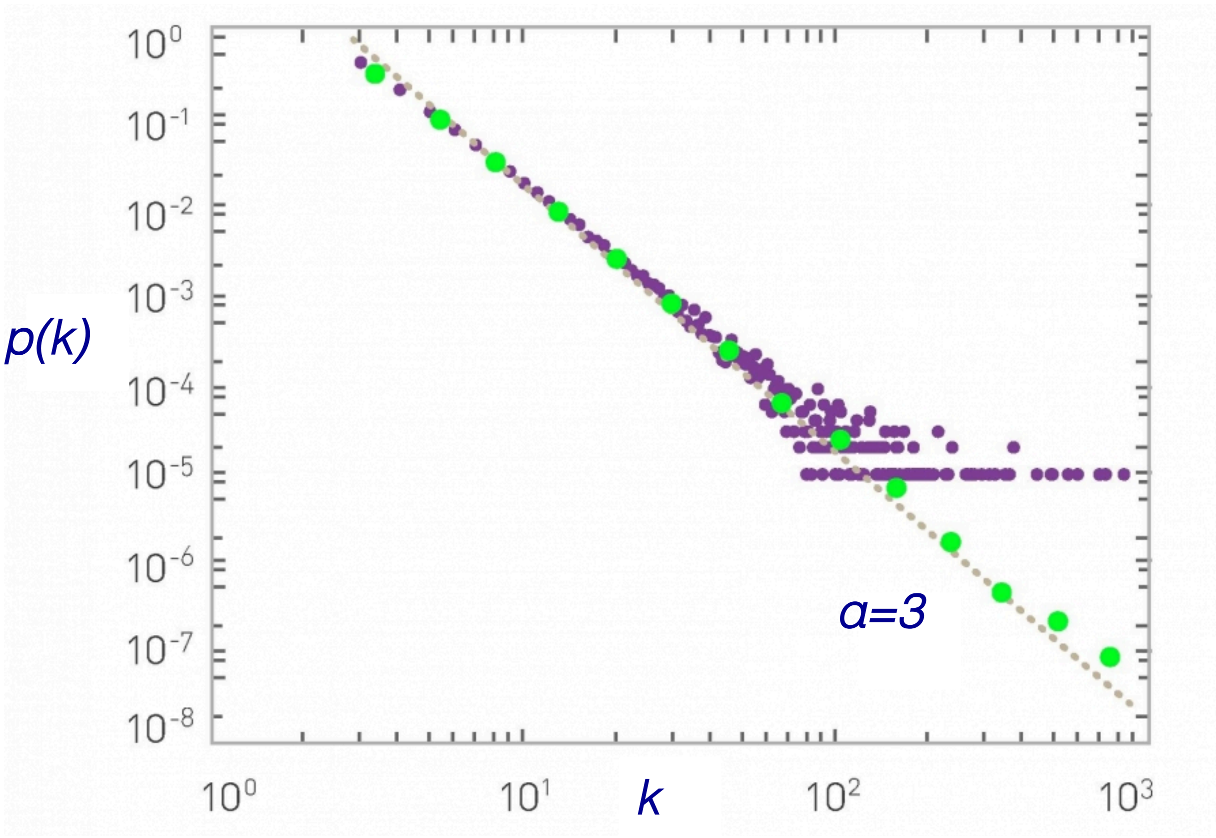
$$P(j \to i) = \frac{k_i}{\sum_v k_v}$$

All pre-existing nodes v



[de Solla Price '65, Albert and Barabasi '99, Mitzenmacher '03]

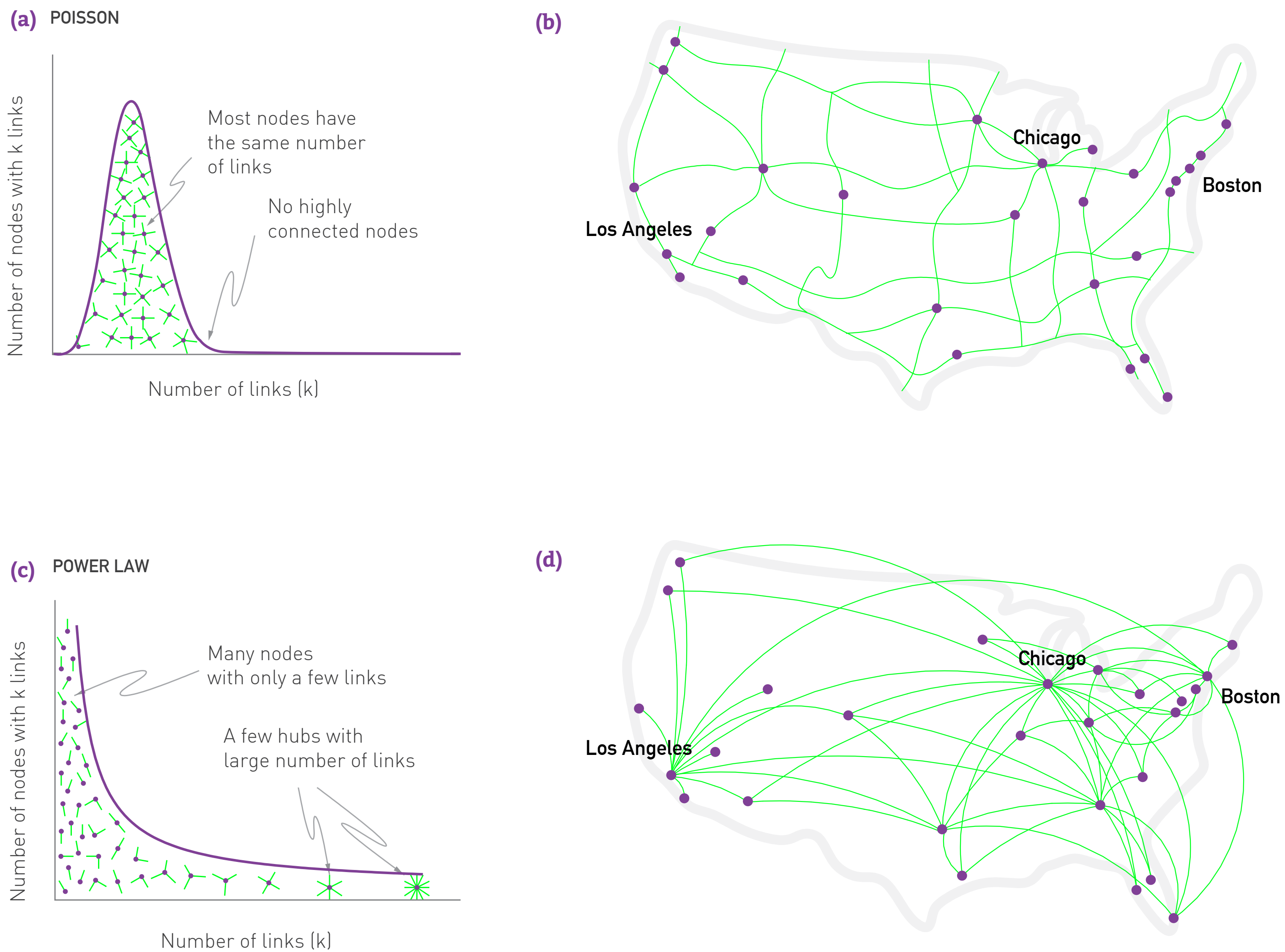
# Degree Distribution of BA Model




*n = 100,000*
*m = 3*

Power-law degree distribution: $p(k) \propto k^{-3}$

# Any Consequences of this Structural Property?

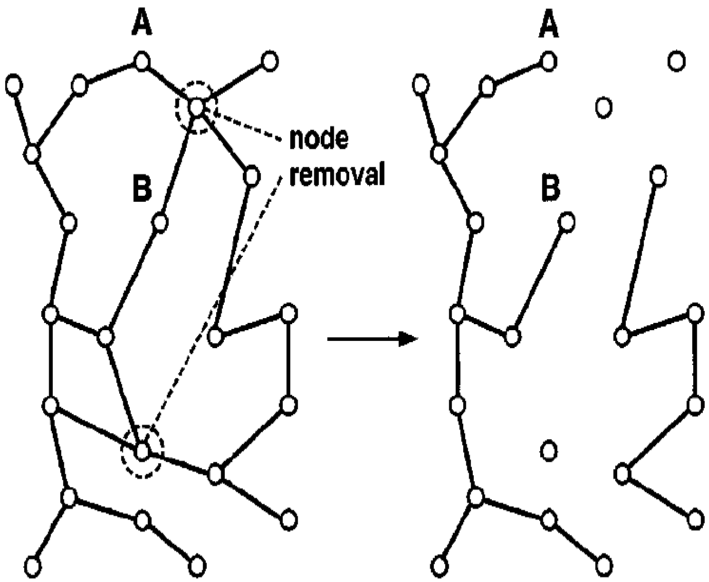


# Consequences of power-law degree distribution

## Error and attack tolerance of complex networks

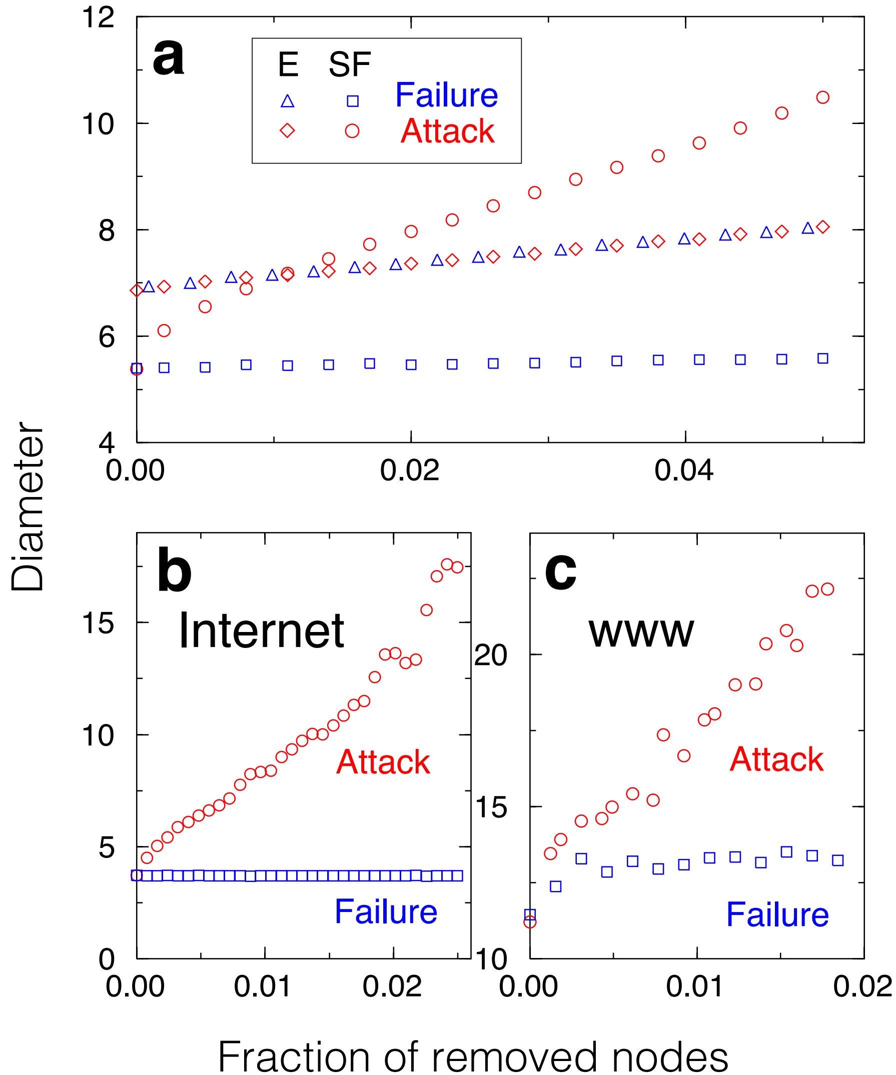
# Network Resilience

- How does network connectivity change as nodes get removed? [Albert et al. 00; Palmer et al. 01]
- Nodes can be removed
  - Random failure:
    - Remove nodes uniformly at random
  - Targeted attack:
    - Remove nodes in order of decreasing degree




This can have an important effect in the robustness of the internet as well as in the domain of epidemiology

# Error Tolerance




E: Random network (Exponential degree distr.)
SF: Scale-free network (Power-law degree distr.)

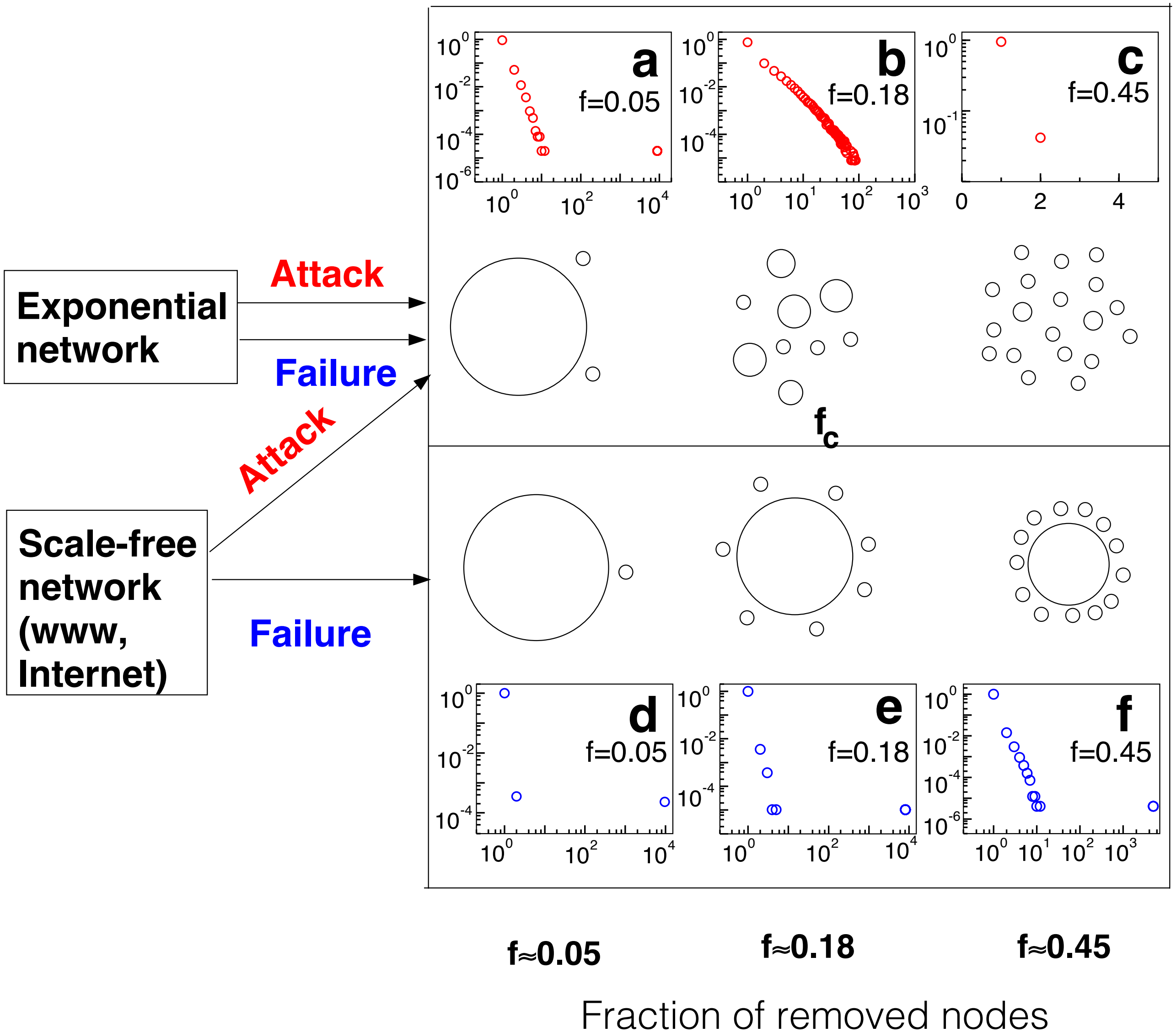
How does the diameter change if we remove a fraction of nodes

- Randomly (Failure)
- Targeted (Attack)

- Real networks are resilient to random failures
- But, they are vulnerable in targeted attacks to high degree nodes

Error and attack tolerance of complex networks. R. Albert, H. Jeong and A.-L. Barabási. Nature, 2000

# Response to Failures or Attacks




Cluster size distribution of SF network under attacks

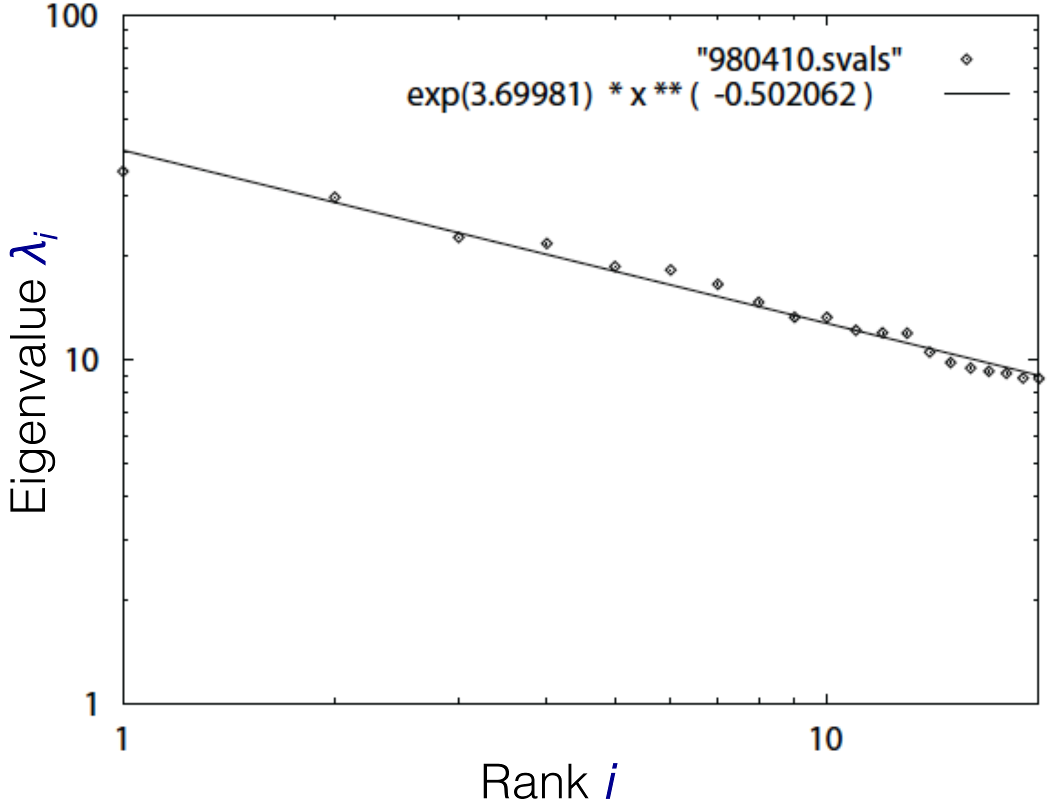
Cluster size distribution of SF network under failures

# Overview of this Lecture

- More power-law distributions in real networks
- Properties of time-evolving networks
- Other network generative models

# Other power-law distributions in real graphs?

# Eigenvalue Distribution




Let $A$ be the adjacency matrix. Then:

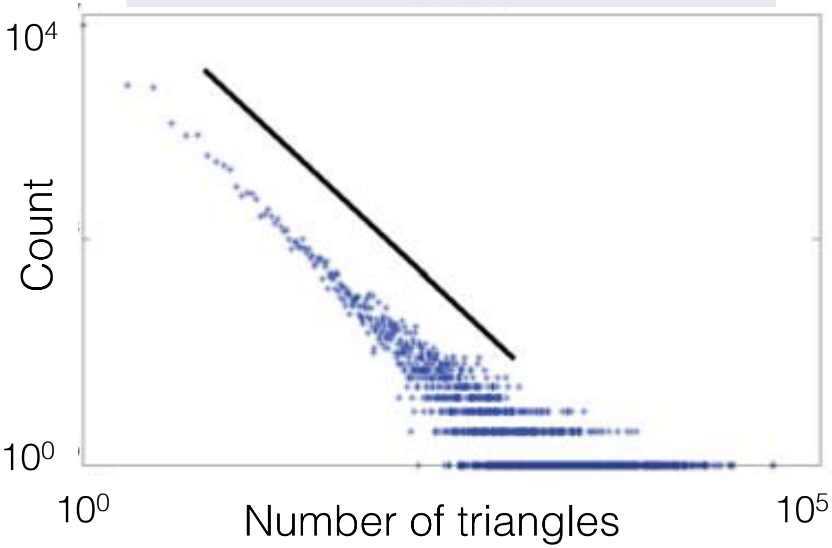
$$Au = \lambda u$$

[Faloutsos, Faloutsos, Faloutsos '99]

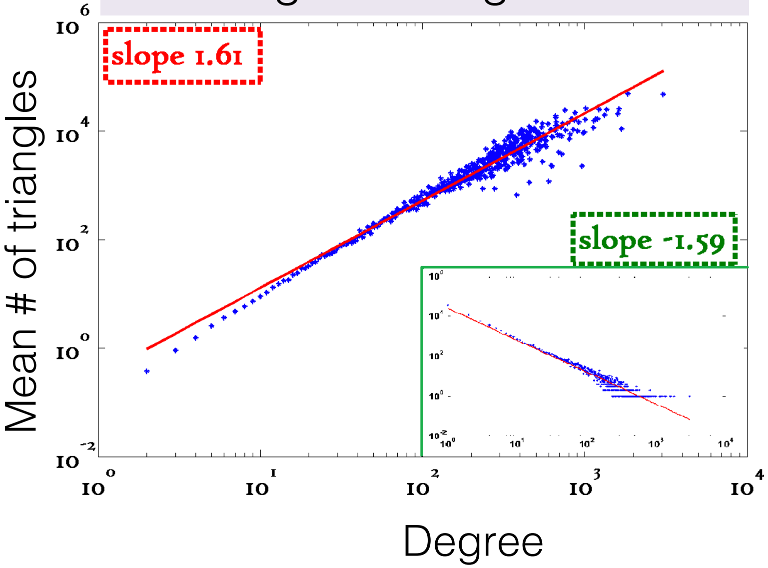
# Triangle Power-Law

Epinions graph

“triangle participation” law




“degree-triangle” law




[Tsourakakis ’08]

# Time-evolving networks

# Macroscopic Evolution

- How do networks evolve at the macro level?
  - What are global phenomena of network growth?

- Questions:
  - What is the relation between the number of nodes $n(t)$ and number of edges $e(t)$ over time $t$ ?
  - How does diameter change as the network grows?
  - How does degree distribution evolve as the network grows?

# Network Evolution (1/2)

- $n(t)$ : number of nodes at time $t$
- $e(t)$ : number of edges at time $t$

- Suppose that

$$n(t+1) = 2 * n(t)$$

- Q: what is happening with the number of edges?

$e(t+1) = ?$ Is it $2 * e(t)$ ?

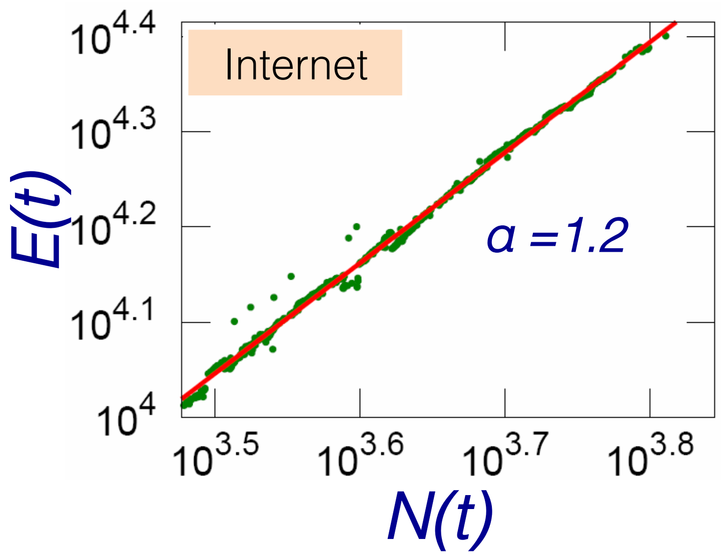
- A: More than doubled
  - But obeying the densification power-law

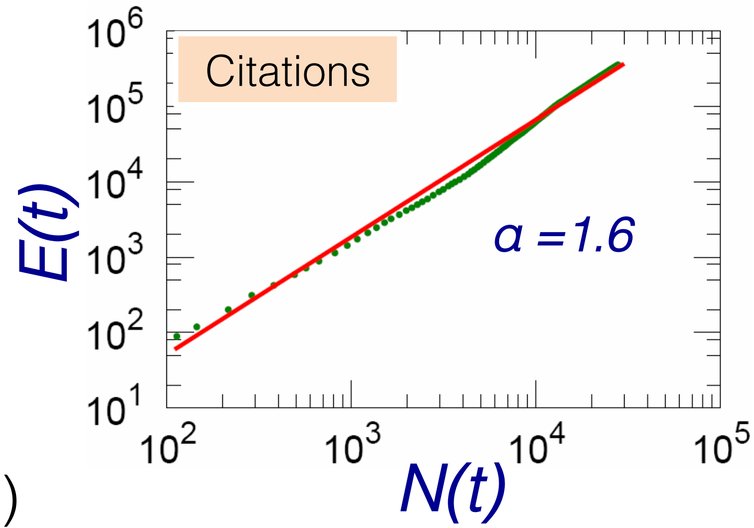
# Network Evolution (2/2)

- What is the relation between the number of nodes and edges over time?
- ~~First guess: constant average degree over time~~
- Networks are denser over time
- Densification Power Law

$$e(t) \propto n(t)^{\alpha}$$

$\alpha$ … densification exponent ( $1 \leq \alpha \leq 2$ )



# Densification Power-Law

- The number of edges grows faster than the number of nodes
  - The average degree is increasing

$$e(t) \propto n(t)^{\alpha}$$ or equivalently $$\frac{\log e(t)}{\log n(t)} = \text{const}$$

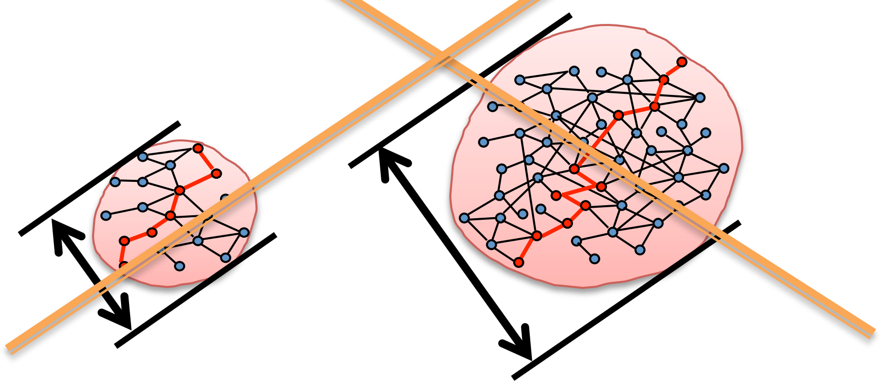
*a* is the densification exponent: *1 ≤ a ≤ 2*

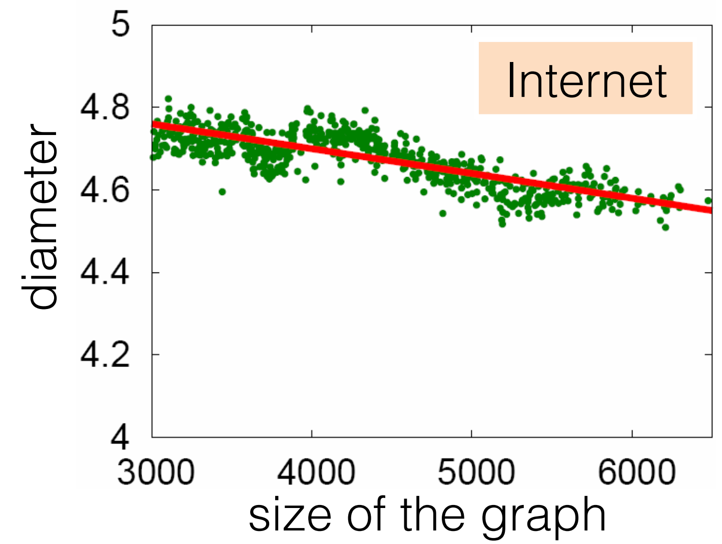
- *a=1* : linear growth – constant out-degree (traditionally assumed)
- *a=2* : quadratic growth – fully connected graph

[Leskovec et al. ’07]

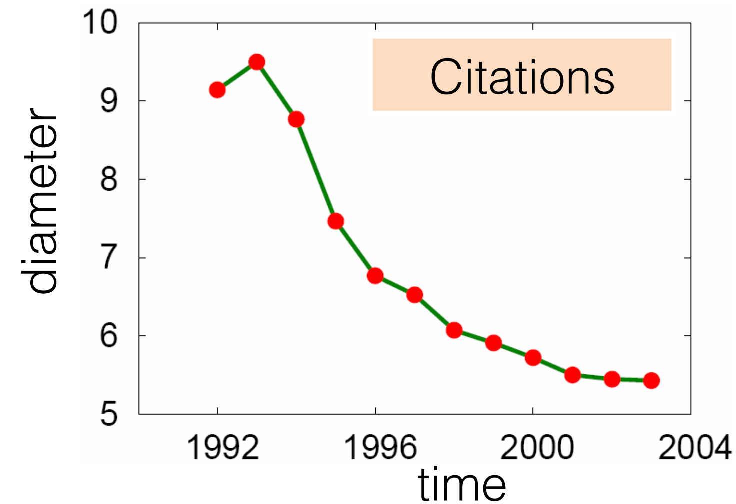
# Network Evolution - Diameter

- ~~Prior models and intuition say that the network diameter slowly grows (like *log n* )~~





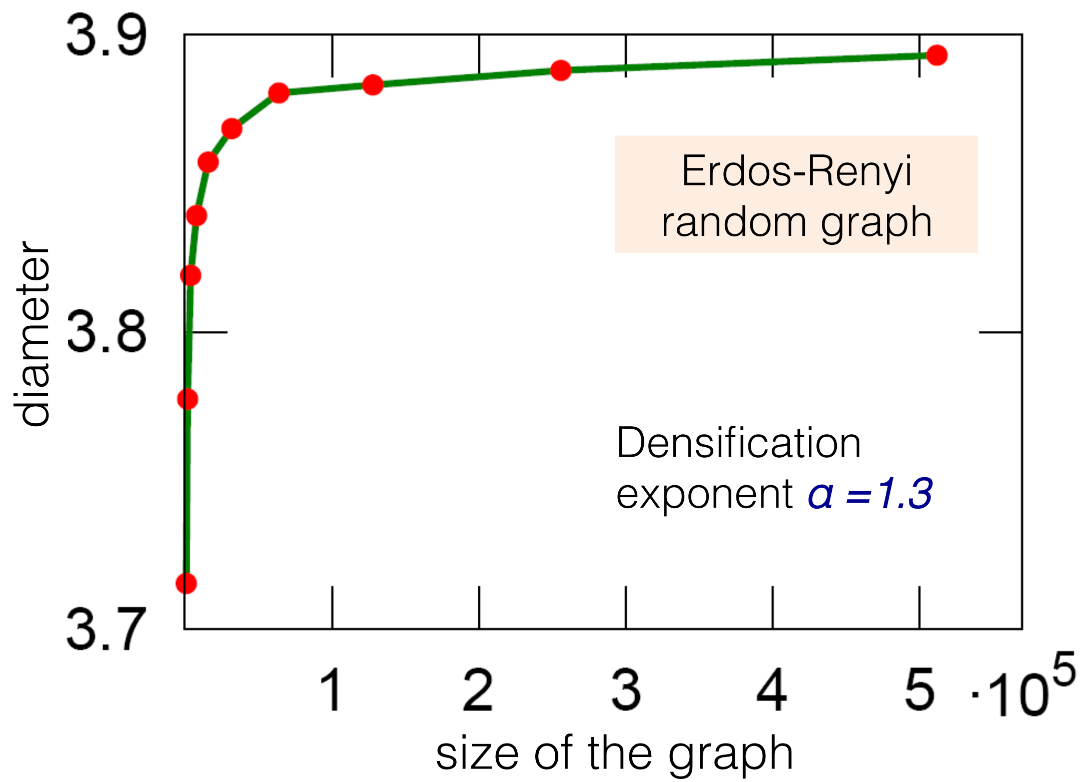

- Diameter shrinks over time
  - As the network grows the distances between the nodes slowly decrease




[Leskovec et al. '07]

# Diameter of a Densifying $G_{n,p}$

Is shrinking diameter just a consequence of densification?



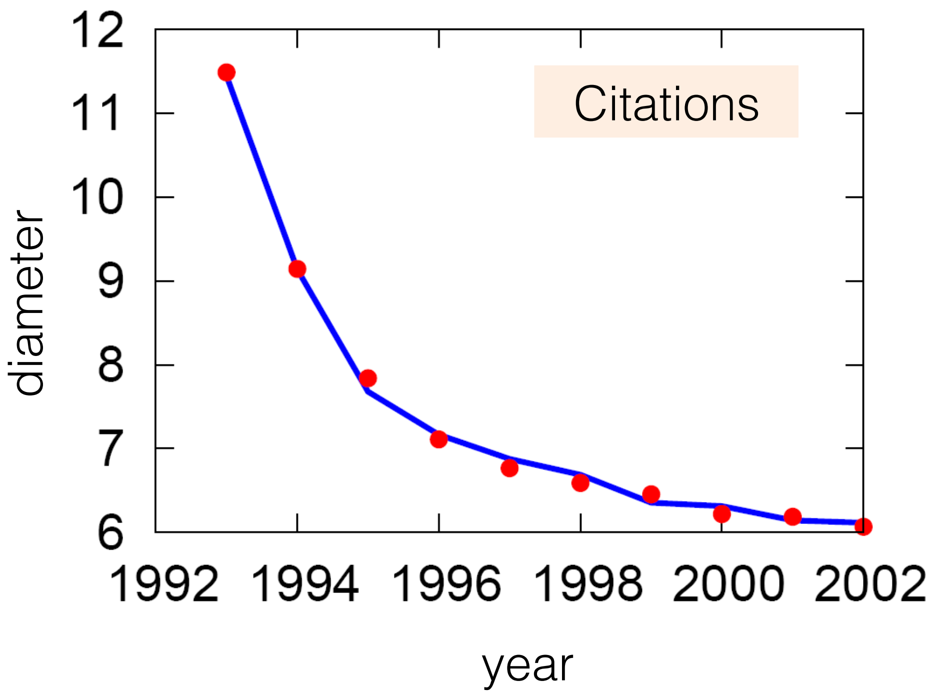

Densifying a random graph has increasing diameter
⇒ There is more to shrinking diameter than just densification!

# Diameter of a Rewired Network

**Is it the degree sequence?**

Compare the diameter of a:

- Real network (red)
- Random network with the same degree distribution (blue)

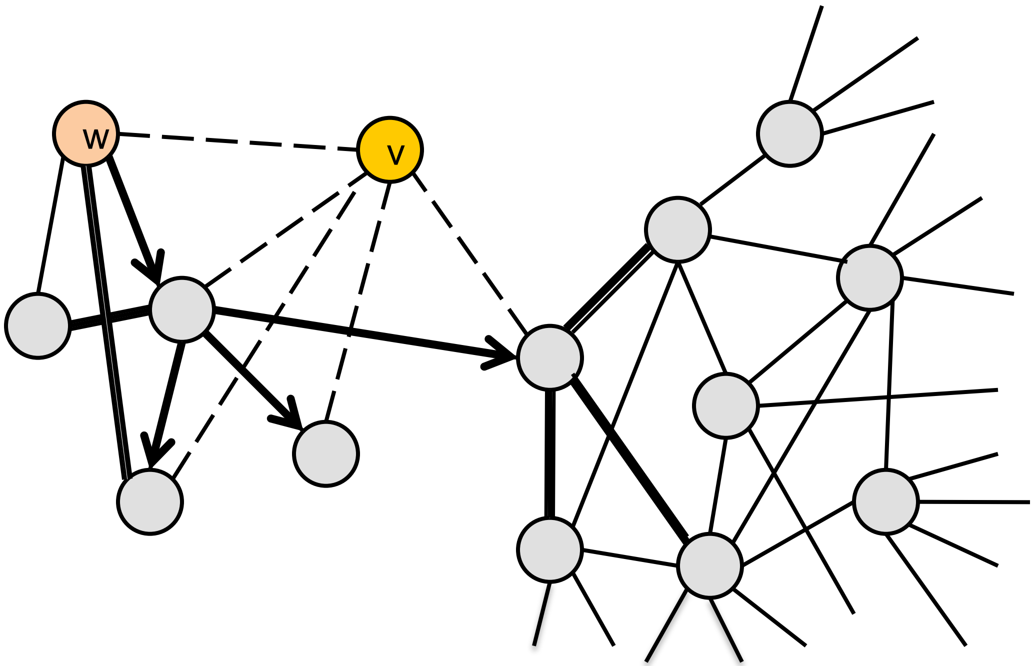



Densification + degree sequence gives shrinking diameter

# Forest fire model

# Forest Fire Model – Basic Idea

- Want to model graphs that become denser and have shrinking diameters
- Intuition:
  - How do we meet friends at a party?
  - How do we identify references when writing papers?




[Leskovec et al. '07]

# Forest Fire Model

- The Forest Fire model has 2 parameters
  - $p$ … forward burning probability
  - $r$ … backward burning probability

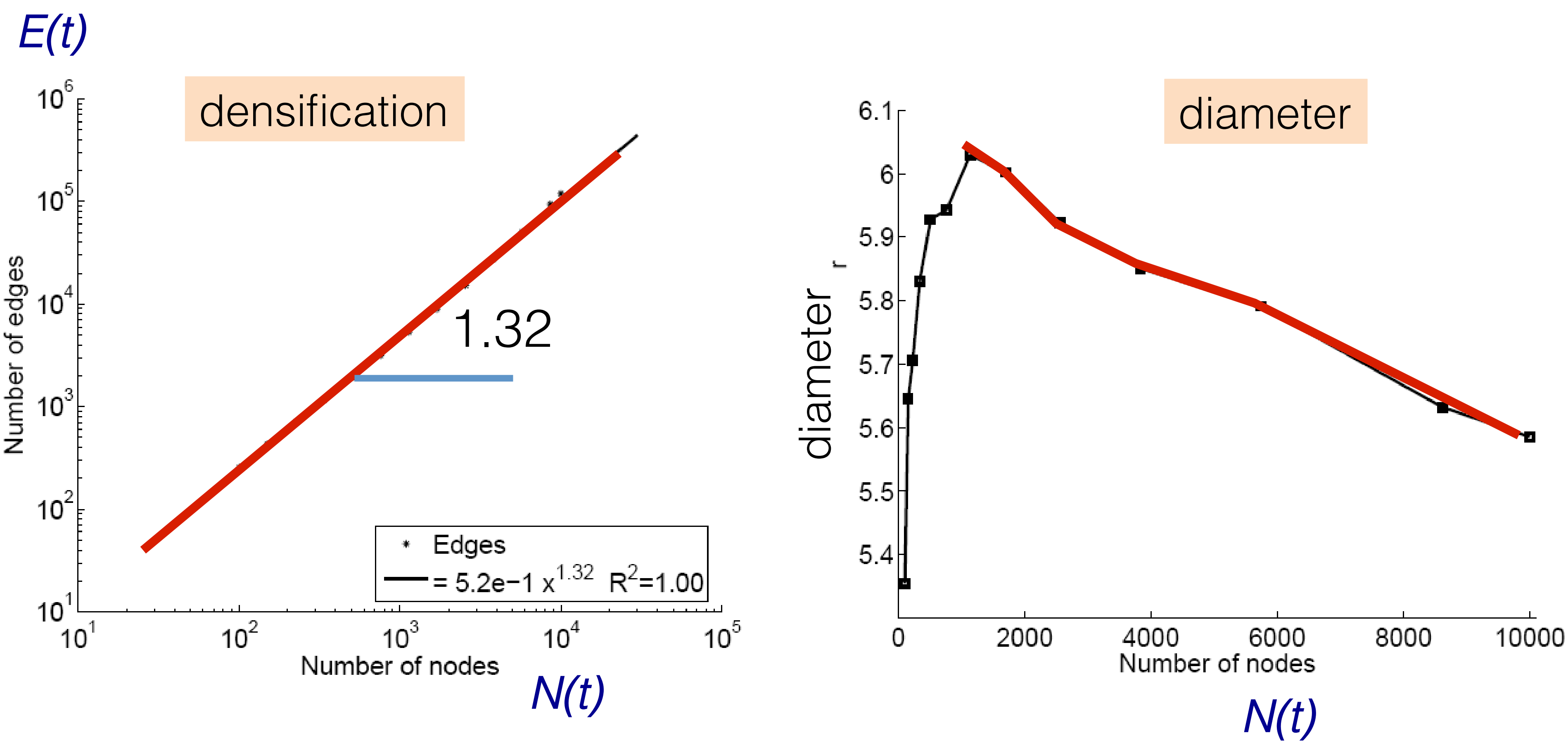
- The model: Directed Graph
  - Each turn a new node $v$ arrives
  - Uniformly at random chooses an "ambassador" $w$
  - Flip 2 geometric coins (based on $p$ and $r$ ) to determine the number of out- (x) and in-links (y) of $w$ to follow
    - $v$ chooses $x$ out-links and $y$ in-links of $w$ which are incident to unvisited nodes
  - Let $w_1$, $w_2$, …, $w_{x+y}$ be the chosen endpoints. Mark them as visited and apply the previous step recursively ("fire" spreads recursively until it dies)
  - New node $v$ links to all burned nodes

The "burning" of links starts at $w$, spreads at $w_1$, $w_2$, …, $w_k$ and proceeds recursively until it dies out

Geometric distribution:
$\Pr(X = k) = (1 - p)^{k-1} p$

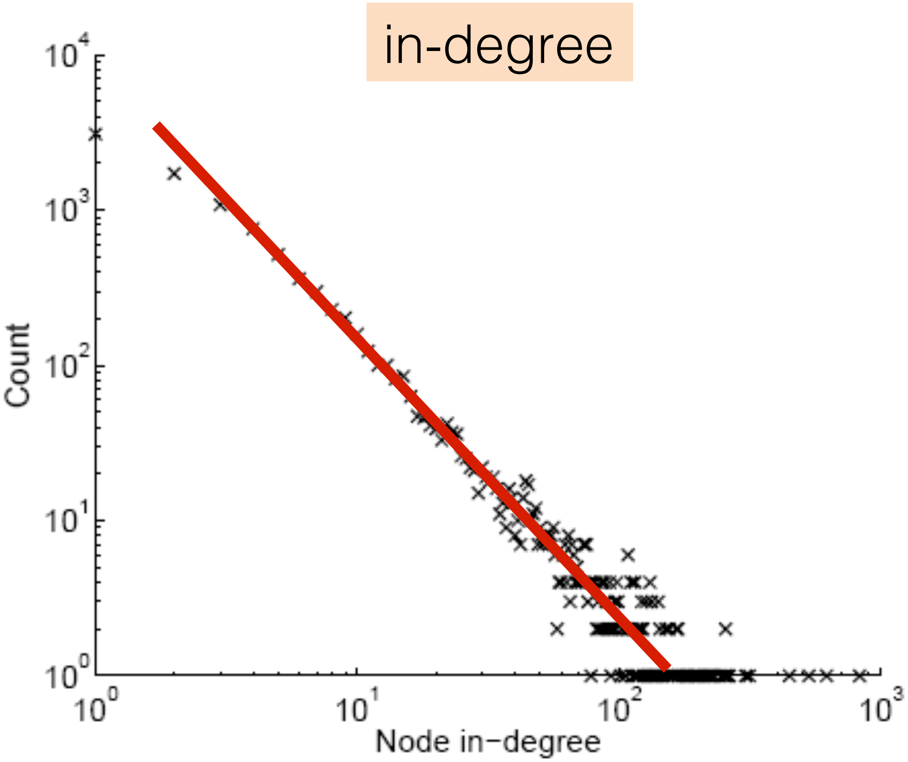
# Forest Fire Model – Properties (1/2)

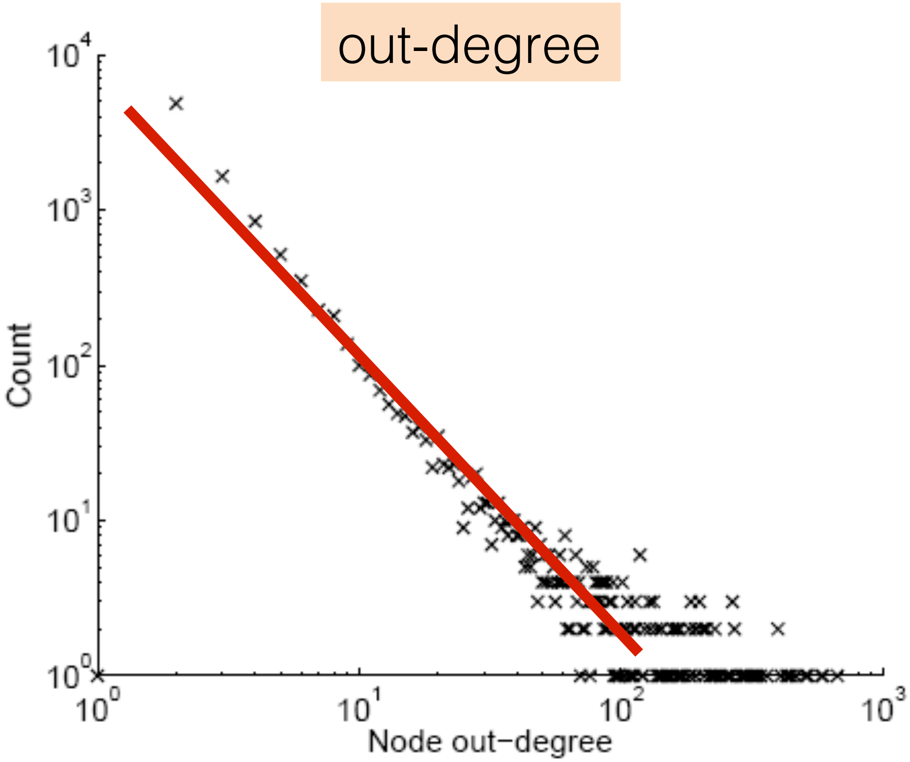
- Forest Fire generates graphs that **densify** and have **shrinking diameter**



# Forest Fire Model – Properties (2/2)

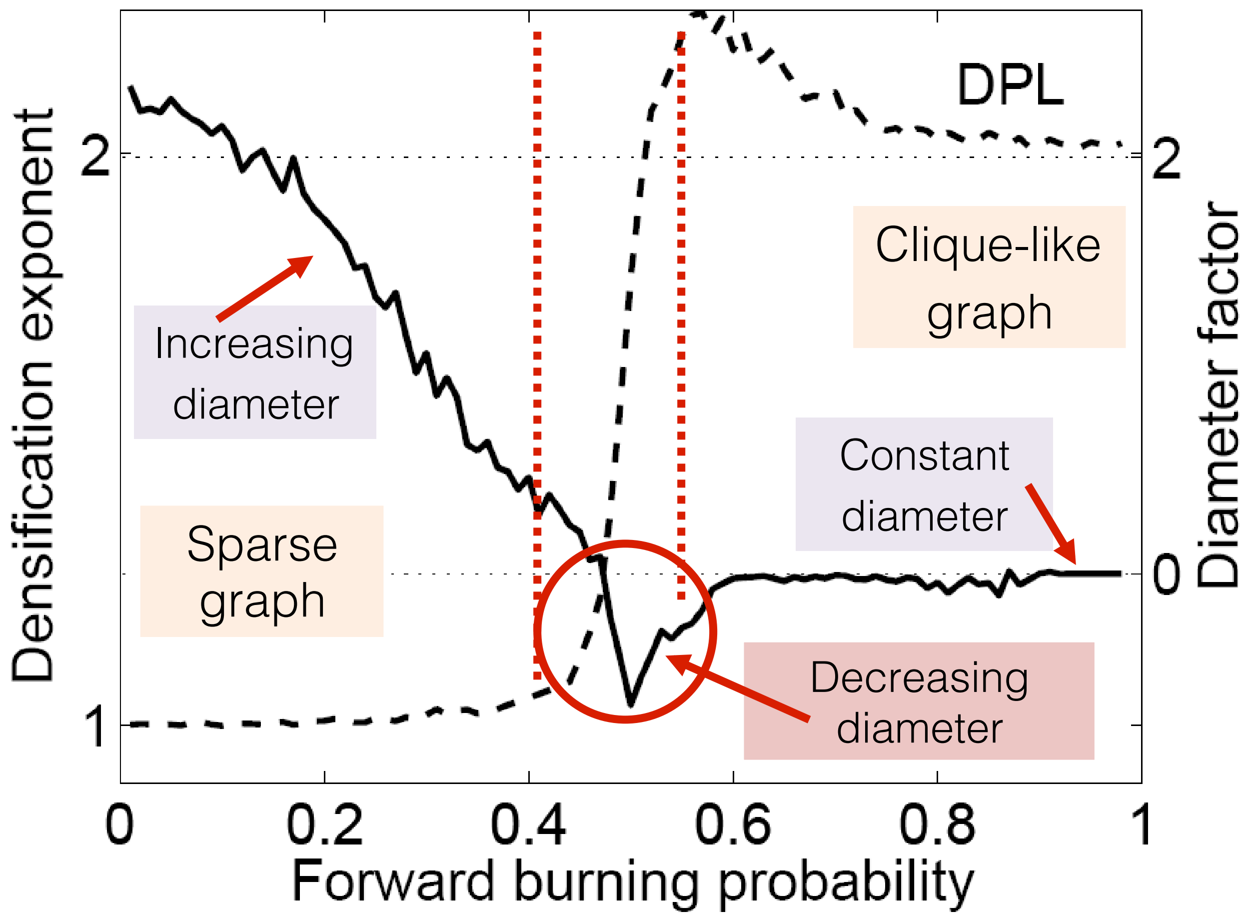
- Forest Fire also generates graphs with power-law degree distribution



# Forest Fire: Phase Transition

- Fix backward probability *r* and vary forward burning prob. *p*
- Notice a sharp transition between sparse and clique-like graphs
- The "sweet spot" is very narrow



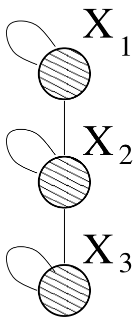
# Kronecker graphs

# What We Have Seen Until Now

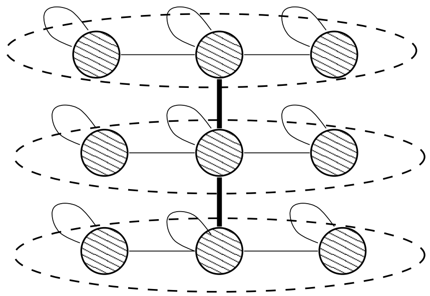
- What is the goal of modeling networks?
  - Discover structural properties of networks
    - Small-world, Clustering coefficient, Heavy-tailed degrees
  - Find a model that gives graphs with such properties
    - Erdős–Rényi, Watts-Strogatz, Barabasi-Albert models
- In this part
  - Can we have a model that attempts to reproduce all of these properties?
  - Can we fit the model to a network and accurately reproduce the network?

# Idea: Recursive Graph Generation

- How can we think of network structure recursively?
  - Intuition: Self-similarity
  - **Object is similar to a part of itself**: the whole has the same shape as one or more of the parts
- Mimic recursive graph/community growth:




Initial graph

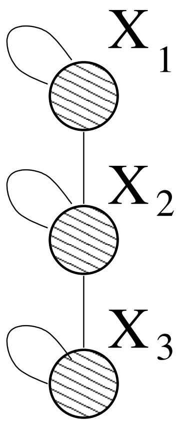


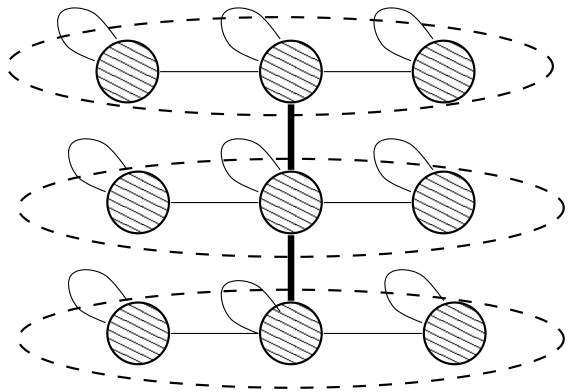
Recursive expansion

Kronecker graph is a way of generating self-similar matrices

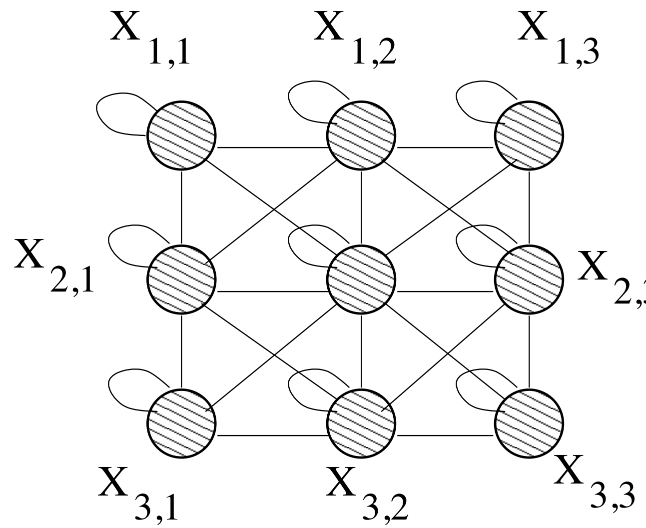
[Leskovec et al. '10]

# Kronecker Graph – The Idea



Intermediate stage




| | | |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 1 | 1 |
| 0 | 1 | 1 |

(3 x 3)

$K_1$

Initiator graph

(9 x 9)

| | | |
|---|---|---|
| $K_1$ | $K_1$ | 0 |
| $K_1$ | $K_1$ | $K_1$ |
| 0 | $K_1$ | $K_1$ |

$$K_2 = K_1 \otimes K_1$$

After the growth phase

# Kronecker Graph – Recursive Model

- Kronecker graphs
  - A recursive model of network structure

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

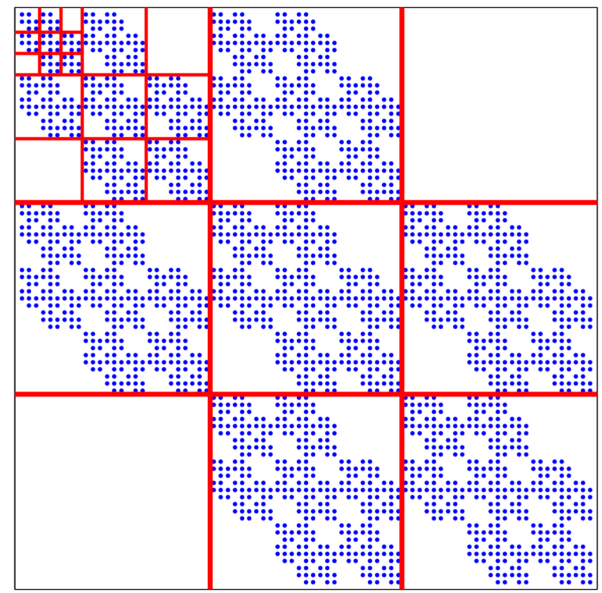
$K_1$

3 x 3

$$\begin{bmatrix} K_1 & K_1 & 0 \\ K_1 & K_1 & K_1 \\ 0 & K_1 & K_1 \end{bmatrix}$$

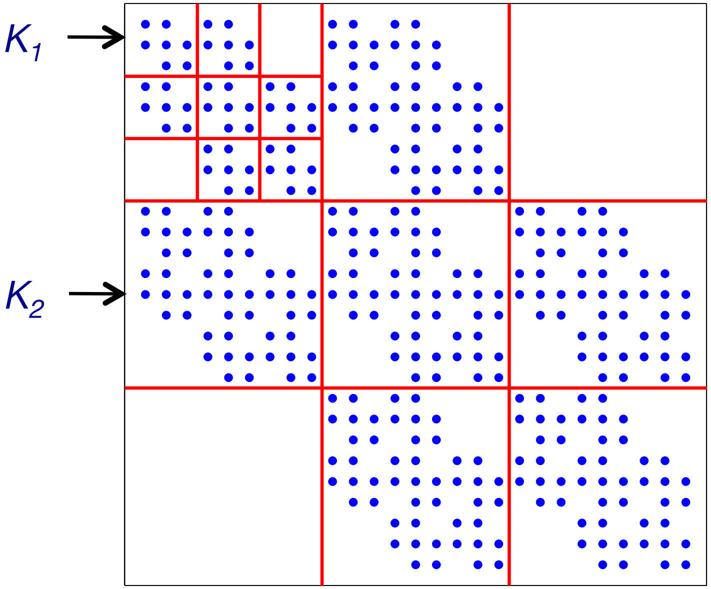
$K_2 = K_1 \otimes K_1$

9 x 9

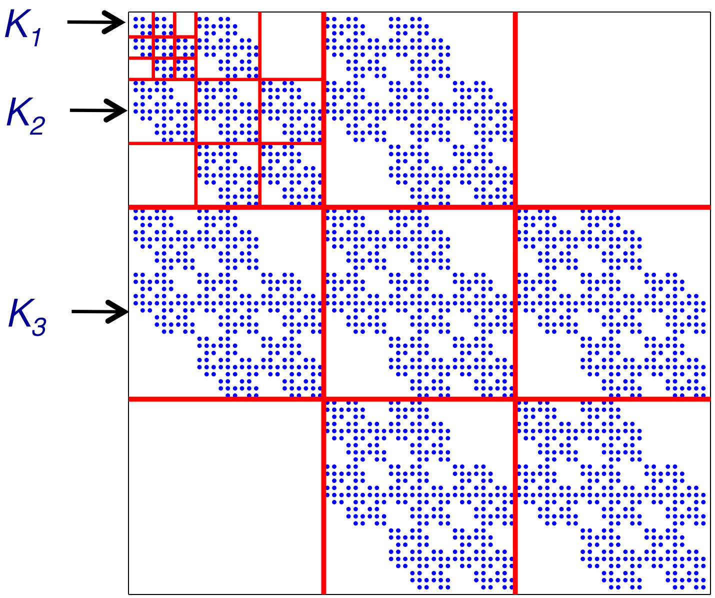


81 x 81 adjacency matrix

# Kronecker Graph - Example




$$K_3 = K_2 \otimes K_1$$




$$K_4 = K_3 \otimes K_1$$

# Kronecker Product: Definition

- Kronecker product of matrices *A* and *B* is given by

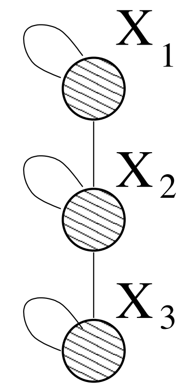
$$\underset{}{\mathbf{C}} = \underset{N\,x\,M}{\mathbf{A}} \otimes \underset{K\,x\,L}{\mathbf{B}} = \underbrace{\begin{pmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,m}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}\mathbf{B} & a_{n,2}\mathbf{B} & \cdots & a_{n,m}\mathbf{B} \end{pmatrix}}_{N*K\,x\,M*L}$$

- Define a Kronecker product of two graphs as a Kronecker product of their adjacency matrices

# Kronecker Graphs

- Kronecker graph: a growing sequence of graphs by iterating the Kronecker product

| | | |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 1 | 1 |
| 0 | 1 | 1 |

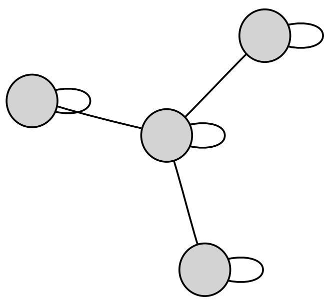



$K_1$

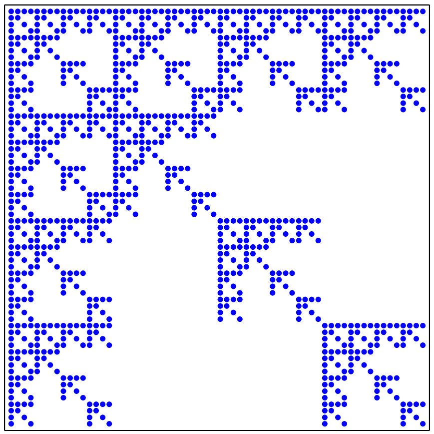
$$K_m = \underbrace{K_1 \otimes K_1 \otimes \ldots \otimes K_1}_{m \text{ times}} = K_{m-1} \otimes K_1$$

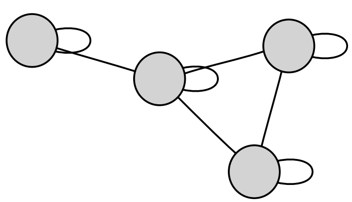
- Note: Once can easily use multiple initiator matrices

# Kronecker Initiator Matrices

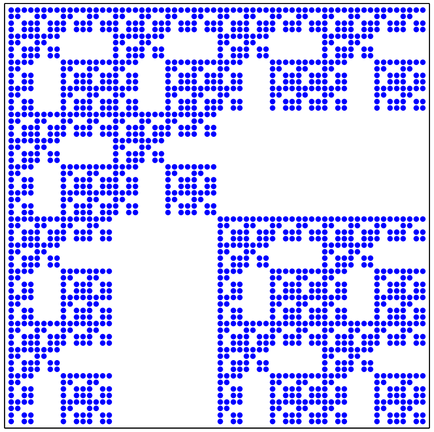


| | | | |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 1 | 0 | 0 | 1 |





| | | | |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 0 | 1 | 1 |



Initiator $K_1$

$K_1$ adjacency matrix

$K_3$ adjacency matrix

# Kronecker Graphs: First Fun Fact

$$K_1^{[m]} = K_m = \underbrace{K_1 \otimes K_1 \otimes \ldots \otimes K_1}_{m \text{ times}} = K_{m-1} \otimes K_1$$

$$K_1 \quad \begin{array}{|c|c|c|} \hline 1 & 1 & 0 \\ \hline 1 & 1 & 1 \\ \hline 0 & 1 & 1 \\ \hline \end{array}$$

- For $K_1$ on $n_1$ nodes and $e_1$ edges, $K_m$ ($m^{th}$ Kronecker power of $K_1$) has:
  - $n_m = n_1^m$ nodes
  - $e_m = e_1^m$ edges

- So, we get the densification power-law $e(t) \propto n(t)^\alpha$

What is $\alpha$?

$$\alpha = \frac{\log e(t)}{\log n(t)} = \frac{\log e_1^t}{\log n_1^t} = \frac{\log e_1}{\log n_1}$$

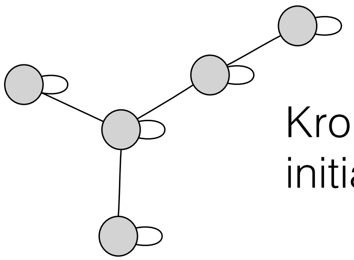
Does not depend on m
(# of iterations)

Since $e_t > n_t$,
then $\alpha > 1$

# Properties of Kronecker Graphs

- Properties of deterministic Kronecker graphs (can be proved!)
  - Properties of static networks
    - Power-law like degree distribution
    - Power-law eigenvalue distribution
    - Constant Diameter
  - Properties of evolving networks
    - Densification Power-law (just proved)
    - Shrinking/Stabilizing Diameter (for Stochastic Kronecker graphs)

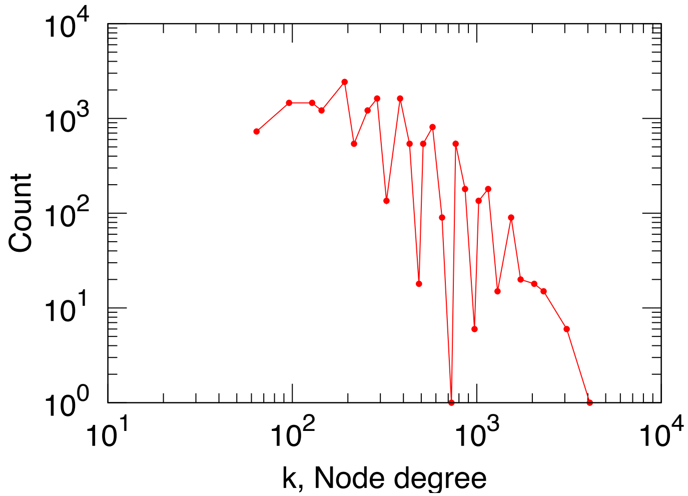
# The Staircase Effect



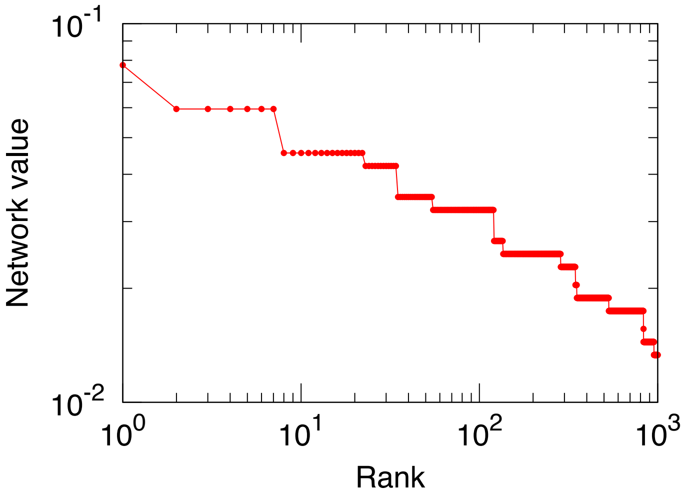
Kronecker initiator $K_1$

Binary adjacency matrix:

- Individual values have large multiplicities
- Staircase effect




Degree distribution of $K_6$




Principal eigenvector distribution of $K_6$

# Stochastic Kronecker Graphs (1/2)

- Create an $n_1 \times n_1$ probability matrix $\Theta_1$
- Compute the $k^{\text{th}}$ Kronecker power $\Theta_k$
- For each entry $p_{uv}$ of $\Theta_k$, include an edge $(u,v)$ in $K_k$ with probability $p_{uv}$

| | |
|---|---|
| 0.5 | 0.2 |
| 0.1 | 0.3 |

$\Theta_1$

Kronecker multiplication

Probability of edge $p_{uv}$

| | | | |
|---|---|---|---|
| 0.25 | 0.10 | 0.10 | 0.04 |
| 0.05 | 0.15 | 0.02 | 0.06 |
| 0.05 | 0.02 | 0.15 | 0.06 |
| 0.01 | 0.03 | 0.03 | 0.09 |

$\Theta_2 = \Theta_1 \otimes \Theta_1$

Instance matrix $K_2$

Flip biased coins

# Stochastic Kronecker Graphs (2/2)

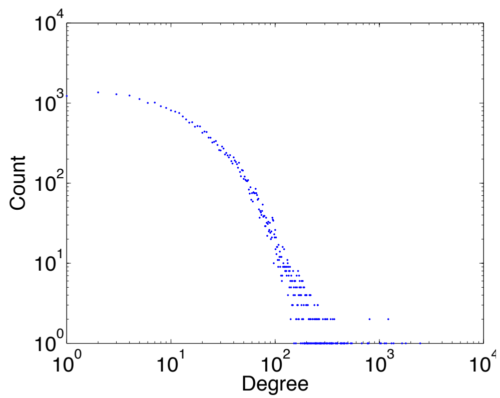
What is known about Stochastic Kronecker Graphs?

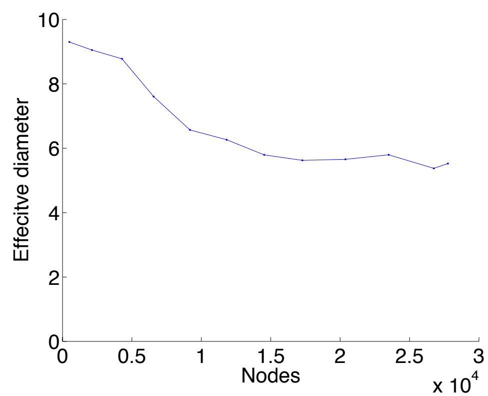
- Undirected Kronecker graph model with: $\Theta_1 = \begin{array}{|c|c|} \hline a & b \\ \hline b & c \\ \hline \end{array}$
  - Connected, if:
    - *$b+c > 1$*
  - Giant connected component of size *$\Theta(n)$*, if:
    - *$(a+b)(b+c) > 1$*
  - Constant diameter, if:
    - *$b+c > 1$*

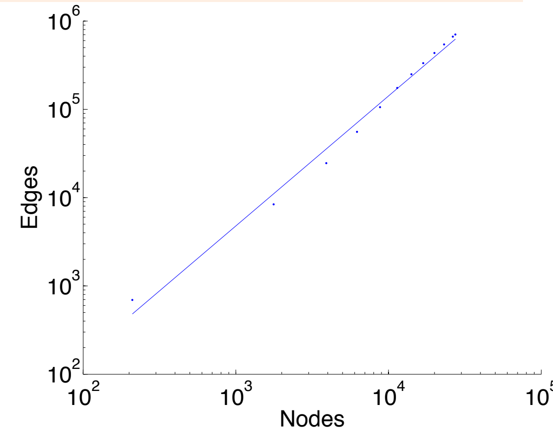
[Mahdian and Xu '07]

# Experimental Results

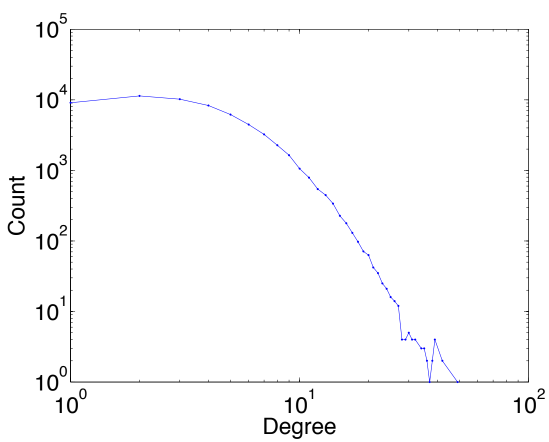
## Citation network



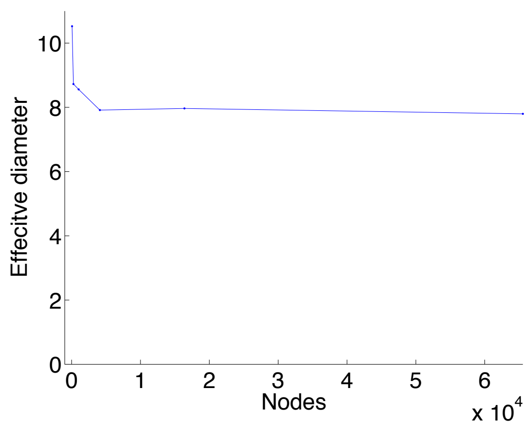



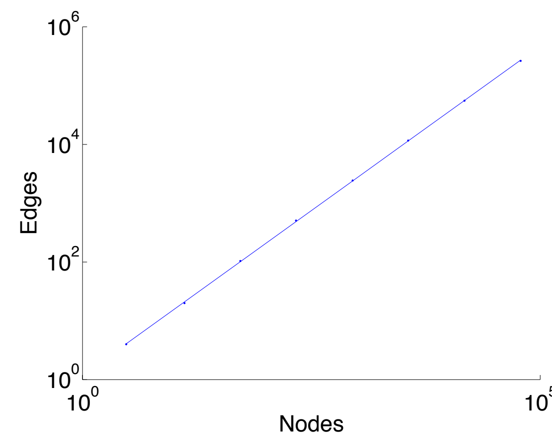

## Stochastic Kronecker Graph




Degree distribution




Shrinking diameter




Densification

# Summary of SKG

- Stochastic Kronecker Graph (SKG)
  - Start from an initiator matrix $K_1$=[a, b; c d] (2x2 or of different size)
  - Repeat Kronecker product $m$ times
  - Get the adjacency matrix of the graph from the probability matrix of the $m^{th}$ iteration
- Simple models that reproduces most of the key properties of real networks
- How do we select $K_1$ ?

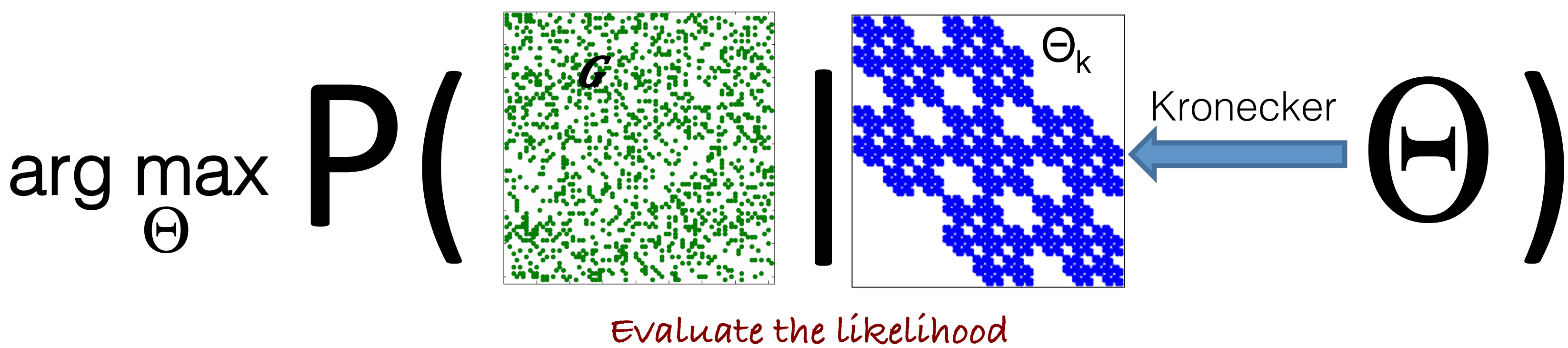
How do we estimate $K_1$ from the data?

# Kronecker Graphs: Estimation

$$\Theta = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

How to estimate Θ from a graph *G* ?

- KronFit algorithm: Maximum likelihood estimation
  - Given a real graph *G*
  - Find the Stochastic Kronecker initiator matrix Θ which

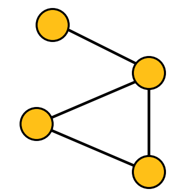



- To solve this, we need to
  - Efficiently calculate P(*G* | Θ)
  - Then, maximize over Θ (e.g., using gradient descent)

[Leskovec et al. '10]

# KronFit: Likelihood $P(G|\Theta)$

Given $G$ and $\Theta$ we calculate likelihood that $\Theta$ generated $G$ : $P(G|\Theta)$



G

| | |
|---|---|
| 0.5 | 0.2 |
| 0.1 | 0.3 |

$\Theta$

| | | | |
|---|---|---|---|
| 0.25 | 0.10 | 0.10 | 0.04 |
| 0.05 | 0.15 | 0.02 | 0.06 |
| 0.05 | 0.02 | 0.15 | 0.06 |
| 0.01 | 0.03 | 0.03 | 0.09 |

$\Theta_k$

| | | | |
|---|---|---|---|
| 1 | 0 | 1 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |

$G$

$P(G|\Theta)$

$$P(G|\Theta) = \prod_{(u,v)\in G} \Theta_k[u,v] \prod_{(u,v)\notin G} (1-\Theta_k[u,v])$$

Likelihood of edges in the graph    Likelihood of edges not in the graph

# Experiments on Real Networks

$$\Theta = \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}$$

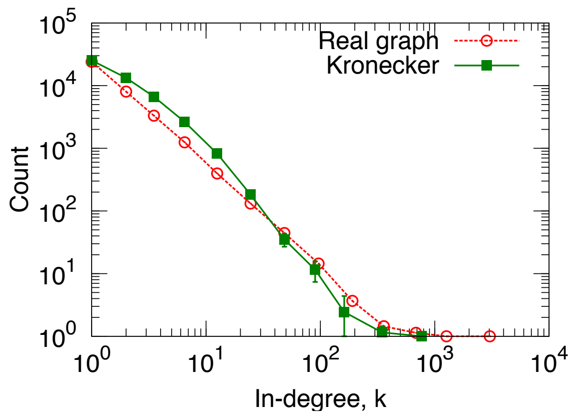
- Experimental setup
  - Given real graph *G*
  - Estimate parameters Θ
  - Generate synthetic graph *K* using Θ
  - Compare properties of graphs *G* and *K*

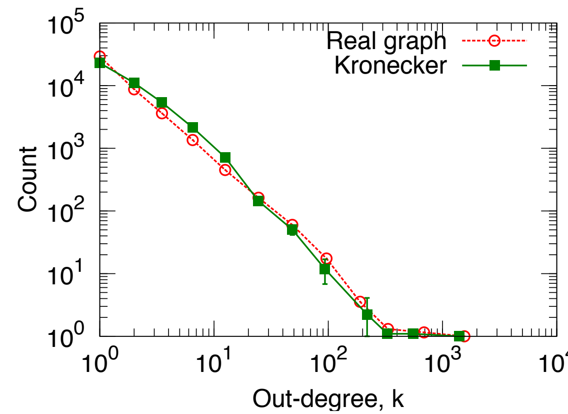
- Note
  - We do not fit the graph properties themselves
  - We fit the likelihood and then compare the properties

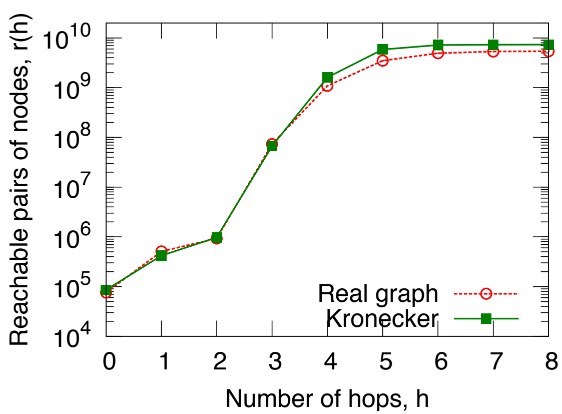
# Estimation: Epinions (n=76K, m=510K)

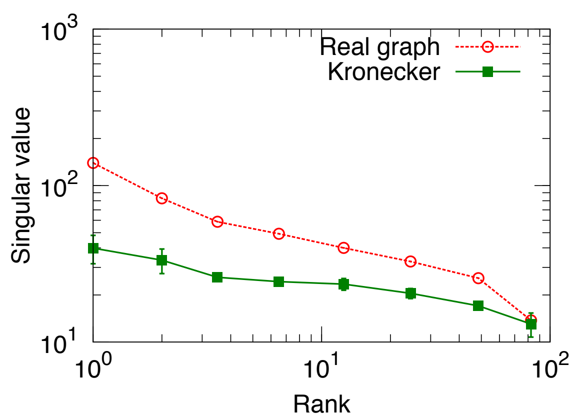
- Real and Kronecker are very close:

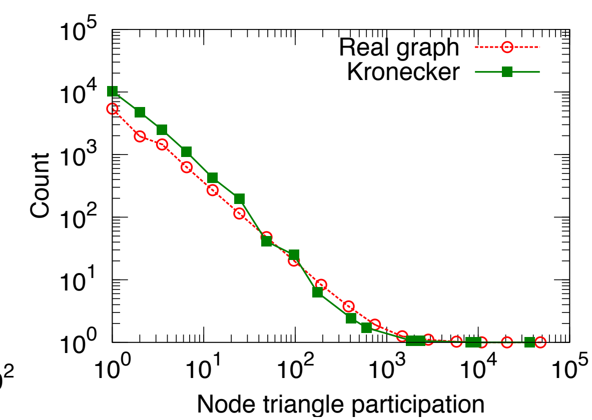
$$\Theta_1 = \begin{array}{|c|c|}\hline 0.99 & 0.54 \\ \hline 0.49 & 0.13 \\ \hline \end{array}$$



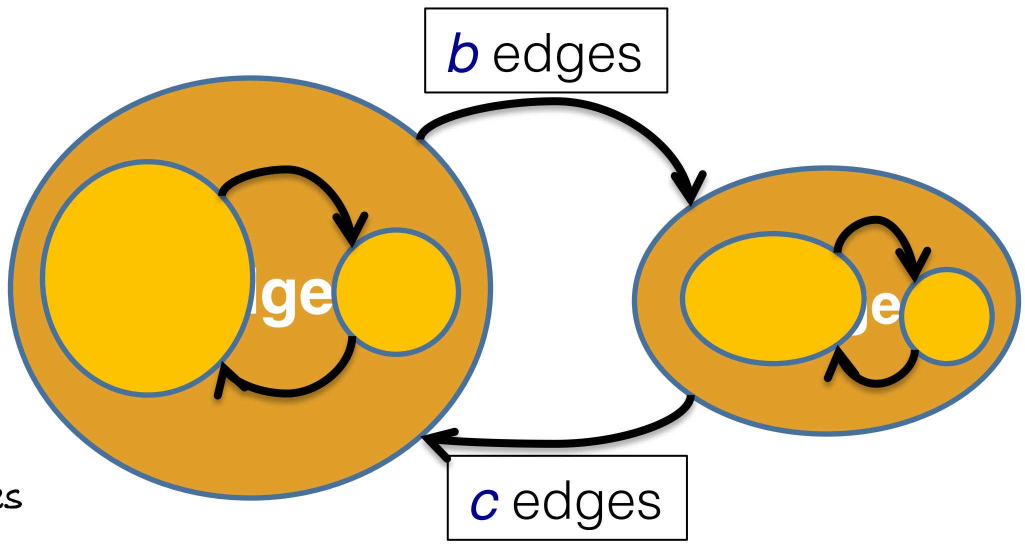




[Leskovec et al. '10]

# Kronecker and Network Structure

- What do the estimated parameters tell us about the network structure?

$$\Theta = \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}$$

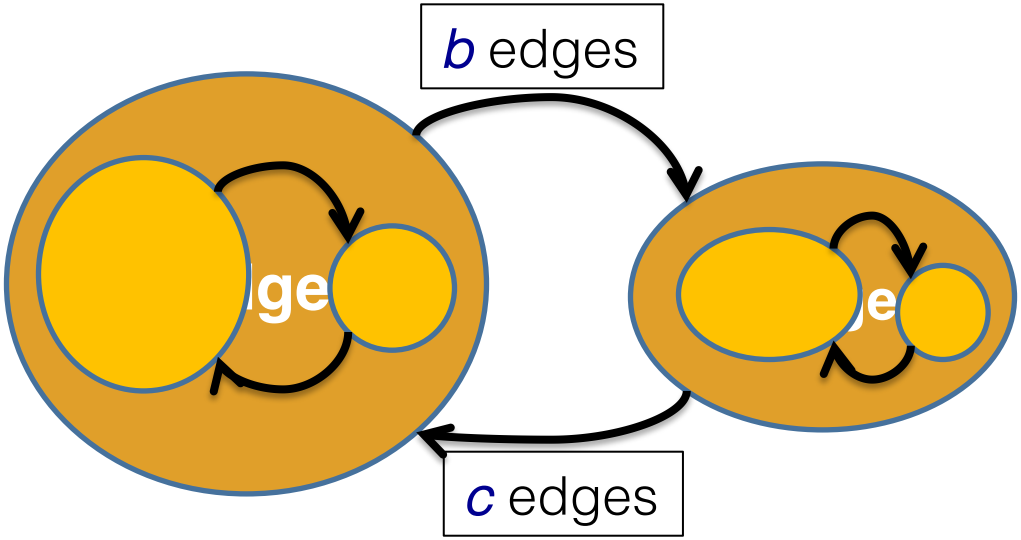


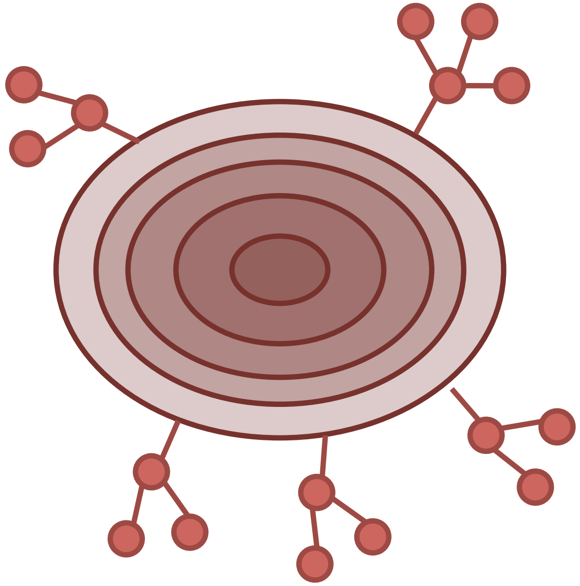

- Diagonal entries: proportion of edges inside each of the group
- Off-diagonal entries: fraction of edges connecting the groups

# Kronecker and Network Structure

- What do estimated parameters tell us about the network structure?

$$\Theta = \begin{array}{|c|c|}\hline 0.9 & 0.5 \\ \hline 0.5 & 0.1 \\ \hline \end{array}$$

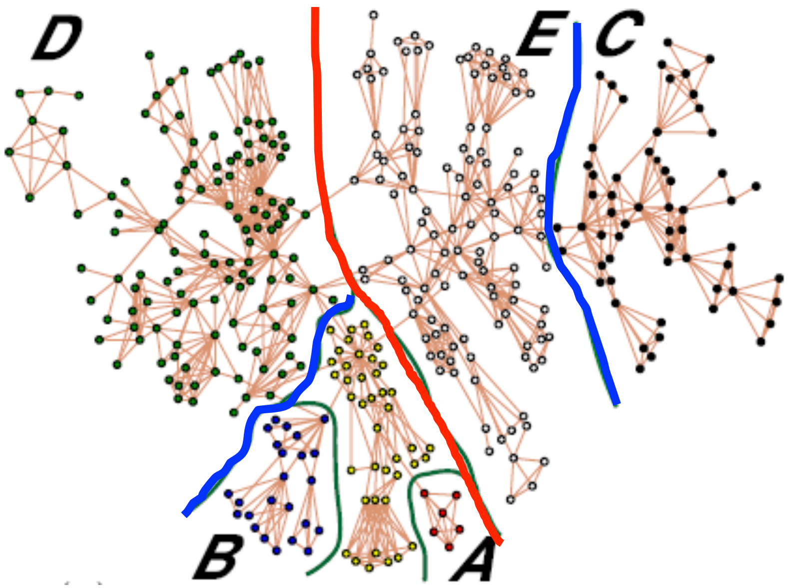


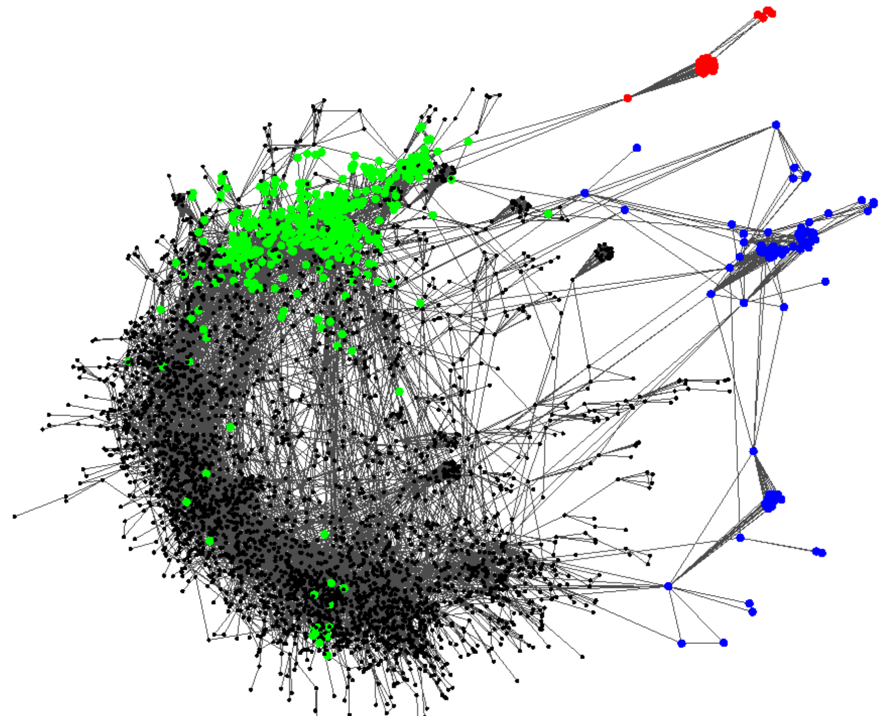



Nested Core-periphery

# Small vs. Large Networks

Small and large networks are very different:



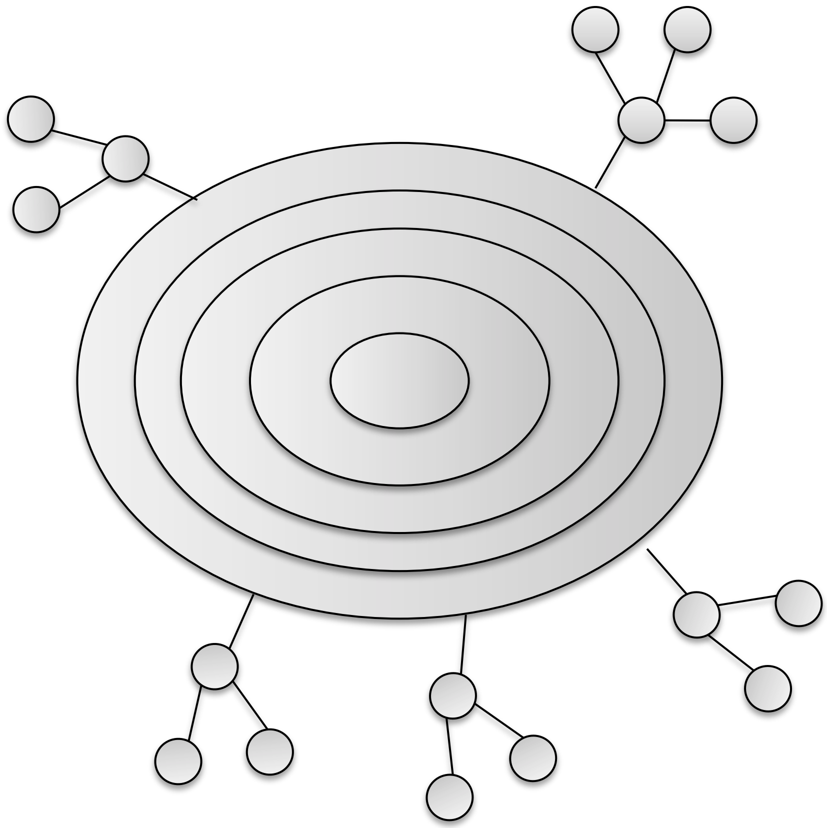
$$\Theta = \begin{bmatrix} 0.99 & 0.17 \\ 0.17 & 0.82 \end{bmatrix}$$

$$\begin{array}{ccc} & \text{core} & \\ \Theta = & \begin{bmatrix} 0.99 & 0.54 \\ 0.49 & 0.13 \end{bmatrix} & \\ & & \text{periphery} \end{array}$$

# Implications (1/2)

Large scale network structure

- Nested Core-periphery
  - Recursive onion-like structure of the network where each layer decomposes into a core and periphery



Communities-within-communities
(More on this topic soon: community detection)

# Implications (2/2)

$$\Theta = \begin{array}{|c|c|} \hline 0.99 & 0.55 \\ \hline 0.55 & 0.15 \\ \hline \end{array}$$

- Remember the SKG theorems:
  - Connected, if $b+c>1$:
    - $0.55+0.15 > 1$ No!
  - Giant component, if $(a+b)\cdot(b+c)>1$:
    - $(0.99+0.55)\cdot(0.55+0.15) > 1$ Yes!
- Real graphs are in the parameter region analogous to the giant component of an extremely sparse $G_{n,p}$

# Why Care about Modeling?

- Translate a physical problem to a mathematical one
  - Models can be used to prove properties – although they might not be the best ones
  - Better understanding of the data
- Anomaly detection
  - Find the distribution that a quantity should follow, and spot outliers (points that deviate from it)
- Answer what-if scenarios
  - What if somebody has twice as many friends? Do twice as many people adopt a product?

# Next Part

- Centrality criteria in graphs
- Link analysis algorithms

# Thank You!