Three Lectures on Networks

Aaron Clauset
@aaronclauset
Associate Professor of Computer Science
University of Colorado Boulder
External Faculty, Santa Fe Institute

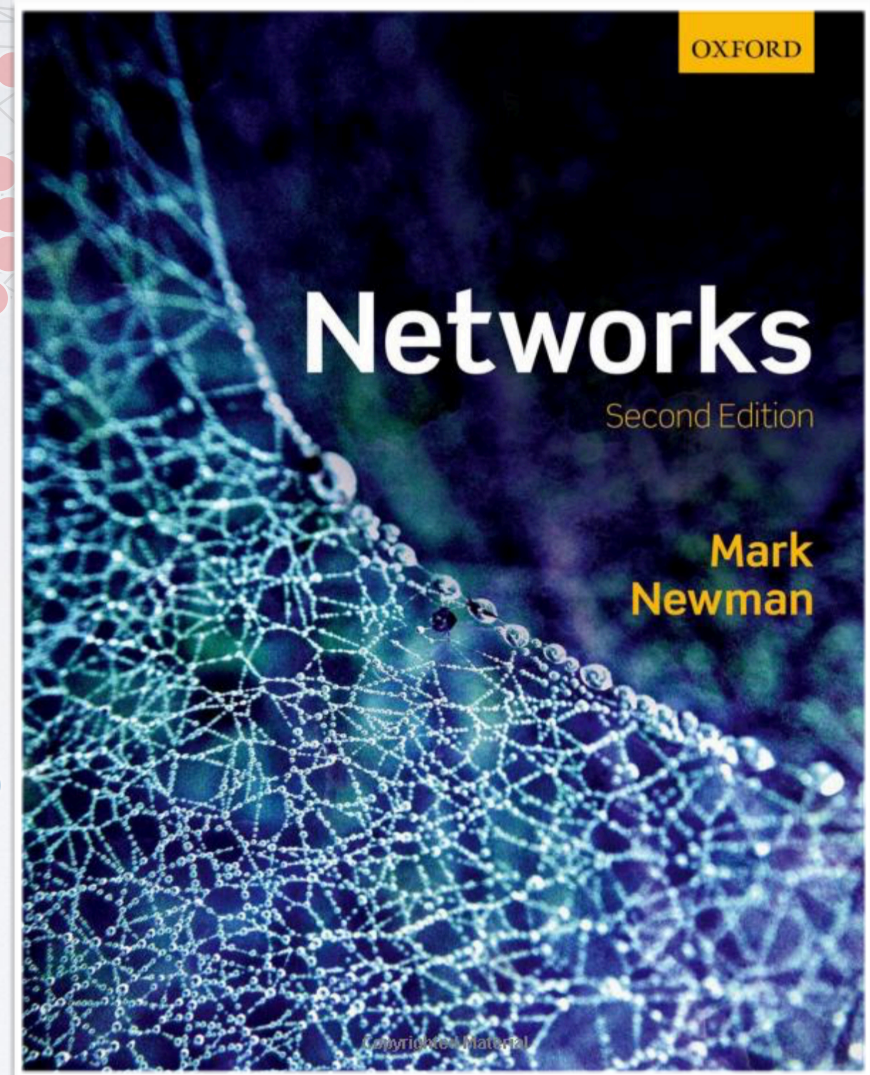
lecture 3: null models and statistical inference for network structure

Mark Newman

Professor of Physics
University of Michigan

External Faculty
Santa Fe Institute

http://www-personal.umich.edu/~mejn/



University of Colorado **Boulder**

Network Analysis and Modeling

Instructor: Aaron Clauset *or* Daniel B. Larremore

This graduate-level course will examine modern techniques for analyzing and modeling the structure and dynamics of complex networks. The focus will be on statistical algorithms and methods, and both lectures and assignments will emphasize model interpretability and understanding the processes that generate real data. Applications will be drawn from computational biology and computational social science. No biological or social science training is required. (Note: this is not a scientific computing course, but there will be plenty of computing for science.)

Full lectures notes online (~150 pages in PDF)

http://santafe.edu/~aaronc/courses/5352/

Software

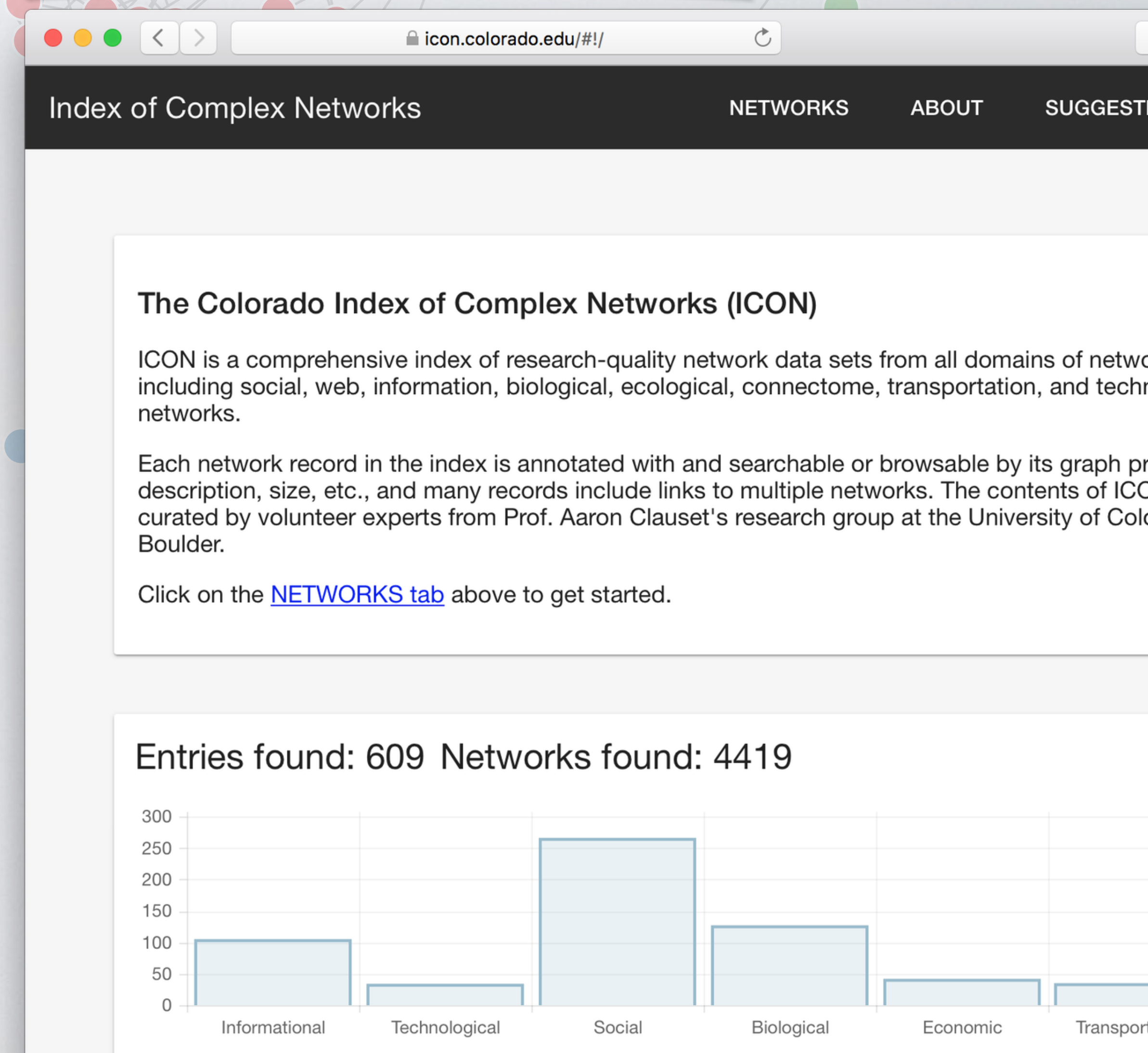
R
Python
Matlab
NetworkX [python]
graph-tool [python, c++]
GraphLab [python, c++]

Standalone editors

UCI-Net
NodeXL
Gephi
Pajek
Network Workbench
Cytoscape
yEd graph editor
Graphviz

Network data sets

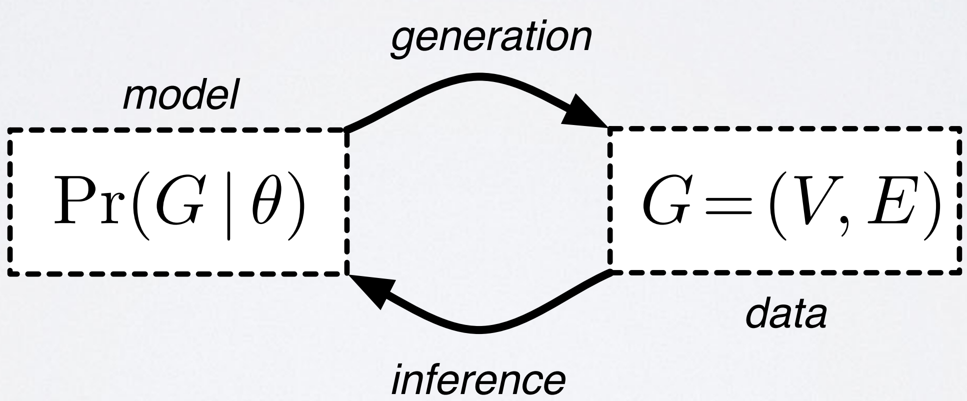
Colorado Index of Complex Networks



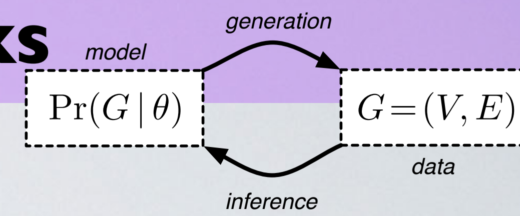
1. defining a network
2. describing a network
3. **null models and statistical inference for networks**

generative models for complex networks

- define a parametric probability distribution over networks $\Pr(G\,|\,\theta)$
- ***generation*** : given θ, draw G from this distribution
- ***inference*** : given G, choose θ that makes G likely



generative models for complex networks



general form

$$\Pr(G \mid \theta) = \prod_{ij} \Pr(A_{ij} \mid \theta)$$

edge generation function

assumptions about "structure" go into $\Pr(A_{ij} \mid \theta)$

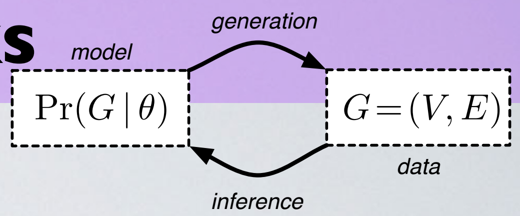
consistency $\lim_{n\to\infty} \Pr\left(\hat{\theta} \neq \theta\right) = 0$

requires that edges be conditionally independent*

3 main classes of these models

* Shalizi & Rinaldo, Annals of Statistics 41, 508–535 (2013)

generative models for complex networks



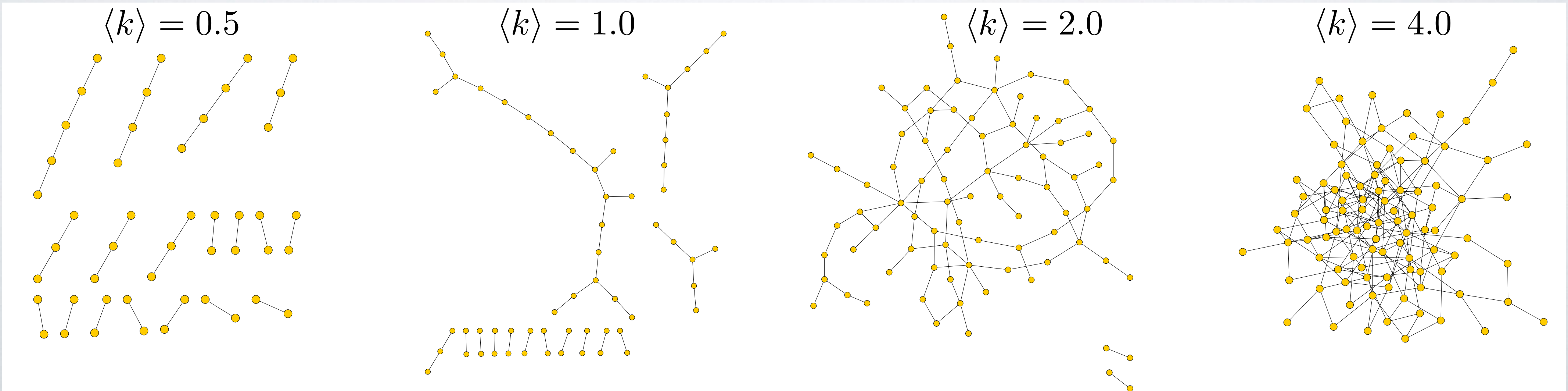
1. Random graph models (unstructured)

- edge density (Erdös-Rényi)

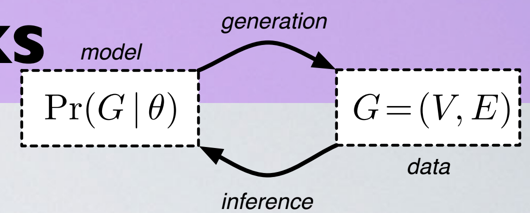
  edges are iid $\Pr(A_{ij}) = p$
  "*homogeneous*" *random graphs*

- degree-based (Chung-Lu & configuration)

  edges independent, conditioned on degree $\Pr(A_{ij}) \propto k_i k_j$
  "*heterogeneous*" *random graphs*

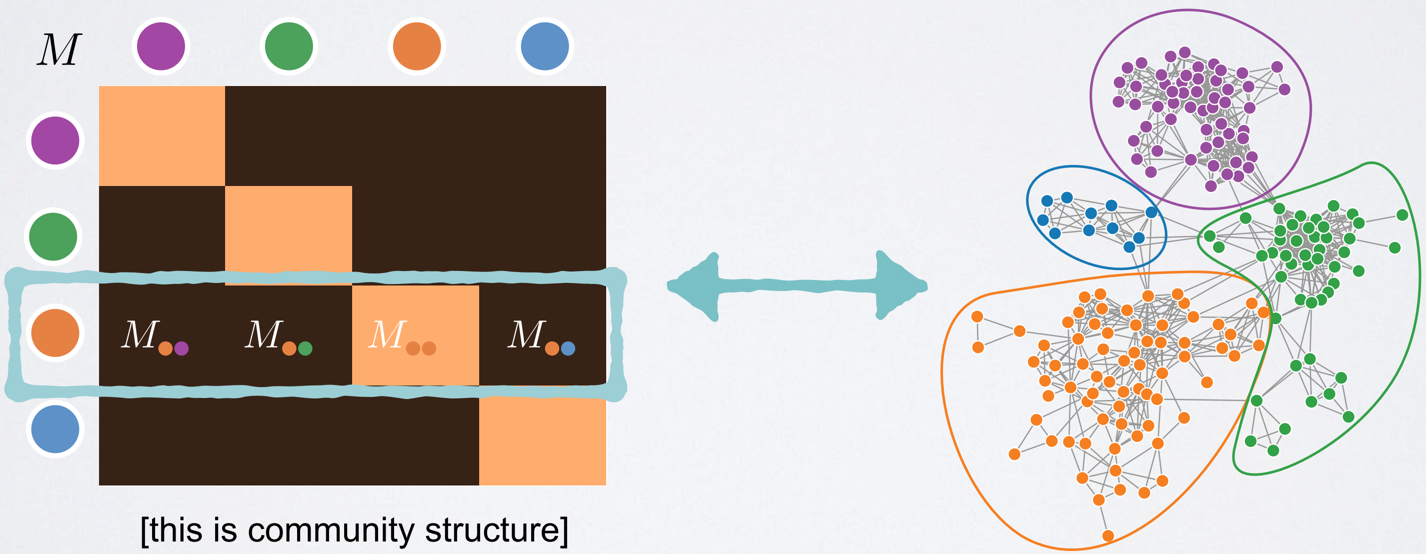


generative models for complex networks



2. Stochastic block models (community structure)

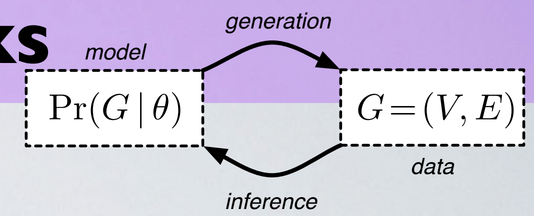
- k groups of nodes: $\Pr(A_{ij} \mid M, z)$ depends only on the types z_i, z_j of the pair i, j
- M is a mixing matrix : $\Pr(i \to j) = M_{z_i, z_j}$



[this is community structure]

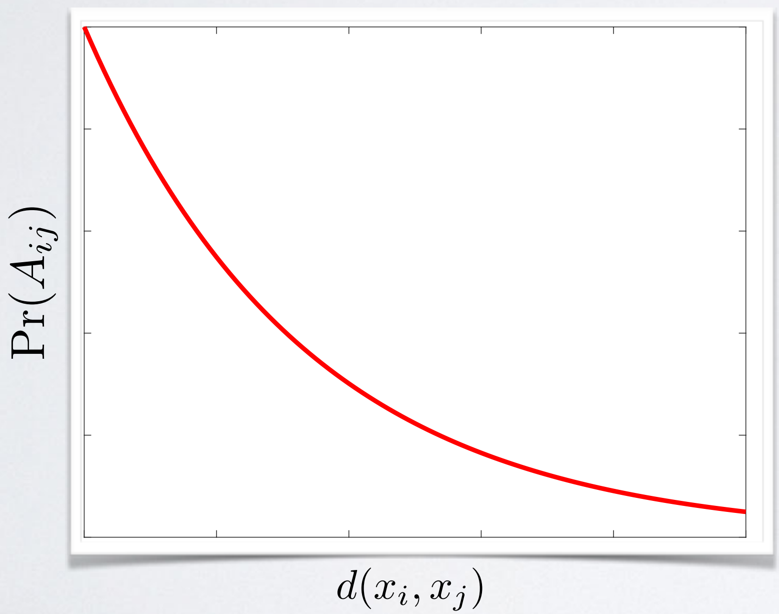
* first formalized in Holland et al., Social Networks 5, 109–137 (1983)

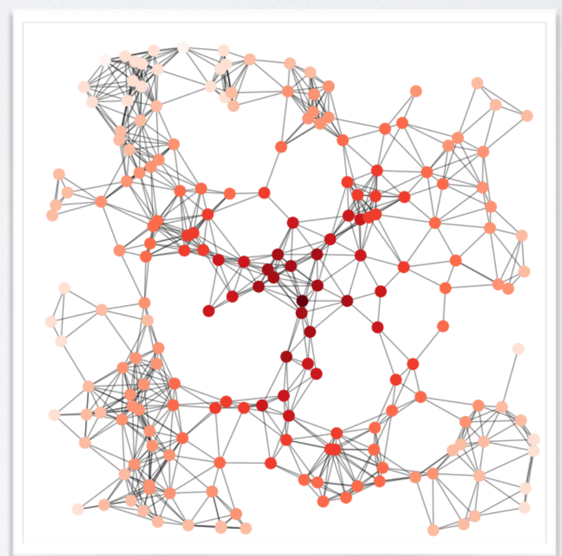
generative models for complex networks



3. Latent space models (random geometric graphs)

- nodes have position in latent space $x_i \in \mathbb{S}$
- $\Pr(A_{ij} \mid d(x_i, x_j))$ depends on distance $d(x_i, x_j)$ of the pair i, j





* formalized in Hoff et al., J. Amer. Stat. Assoc. 97, 1090–1098 (2002)

what patterns should we expect?

feature	real networks
degree distribution	
clustering coefficient	
diameter	
large-scale structure	

what patterns should we expect?

feature	real networks
degree distribution	heavy tailed
clustering coefficient	social: higher non-social: lower
diameter	small, like $O(\ln n)$
large-scale structure	communities, dense core, hierarchies, etc.

Erdos-Renyi random graphs

denoted $G(n, p)$ mean degree

where edges are iid $\Pr(A_{ij}) = p = \dfrac{c}{n-1}$

comments:

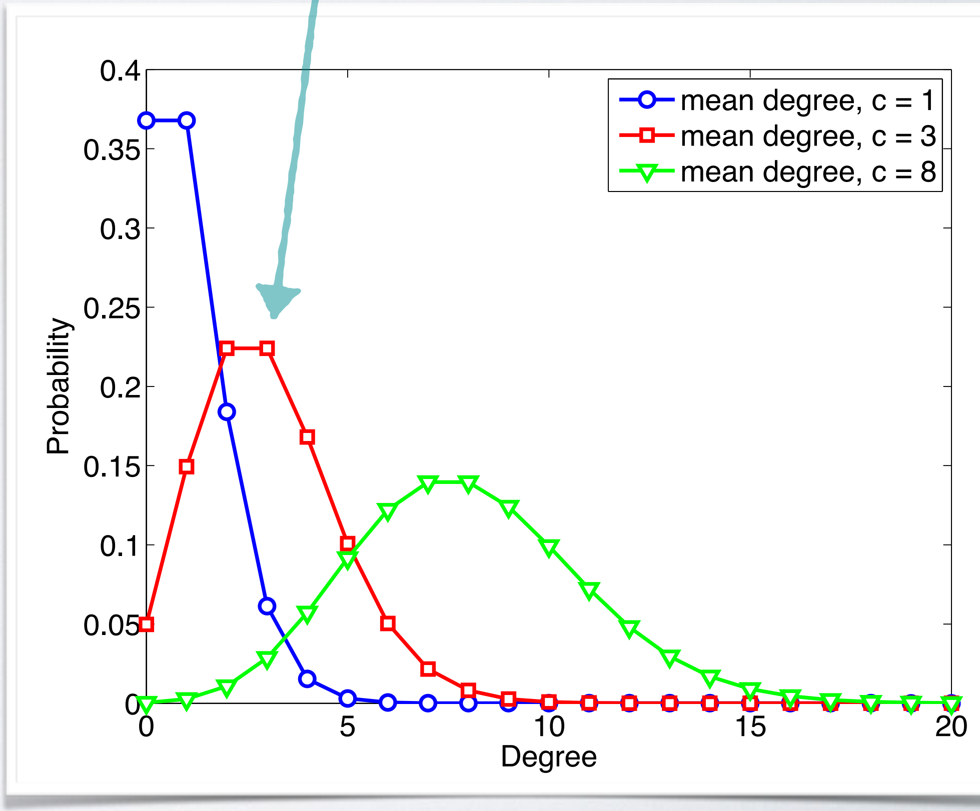
- highly *unrealistic* model (all edges iid)
- but, useful for building intuition & doing math
- the most well-studied random graph model
- warm up for more realistic models

* Erdös & Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5** ,17-61 (1960)

degree distribution

mean degree: $\langle k \rangle = c = (n-1)p$

degree distribution: $\Pr(k) = \mathrm{e}^{-c}\dfrac{c^k}{k!}$ } Poisson distribution



degree distribution

mean degree: $\langle k \rangle = c = (n-1)p$

degree distribution: $\Pr(k) = \mathrm{e}^{-c} \dfrac{c^k}{k!}$ } Poisson distribution

clustering coefficient: $$C = \frac{3 \times \#\text{triangles}}{\#\text{connected triples}}$$

$$= \frac{\binom{n}{3} p^3}{\binom{n}{3} p^2} = p = \frac{c}{n-1} = \underbrace{O(n^{-1})}_{\substack{\text{asymptotically,} \\ \text{zero clustering}}}$$

degree distribution

mean degree: $\langle k \rangle = c = (n-1)p$

degree distribution: $\left.\Pr(k) = \mathrm{e}^{-c}\dfrac{c^k}{k!}\right\}$ Poisson distribution

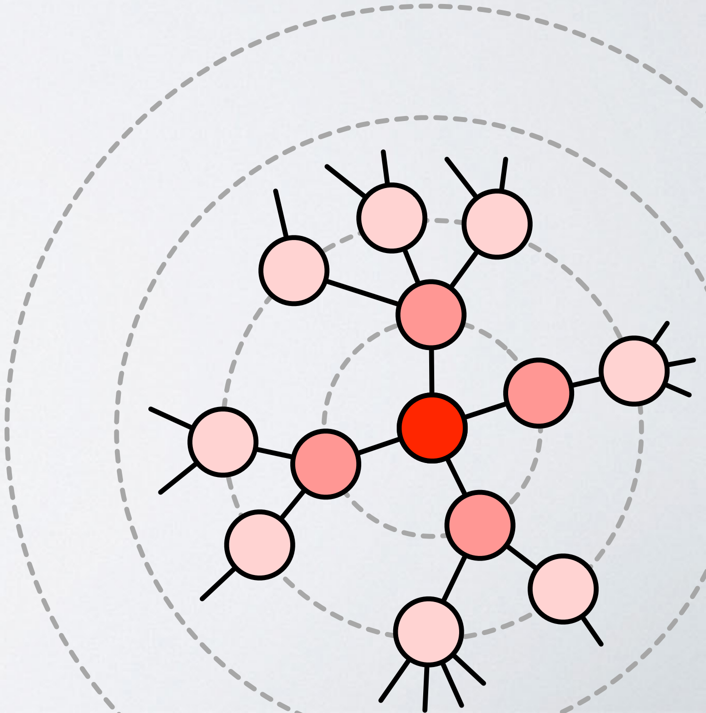
clustering coefficient: $\left.C = O(n^{-1})\right\}$ asymptotically, zero

diameter: $G(n,p)$ is locally tree-like

mean number of vertices within s steps is c^s

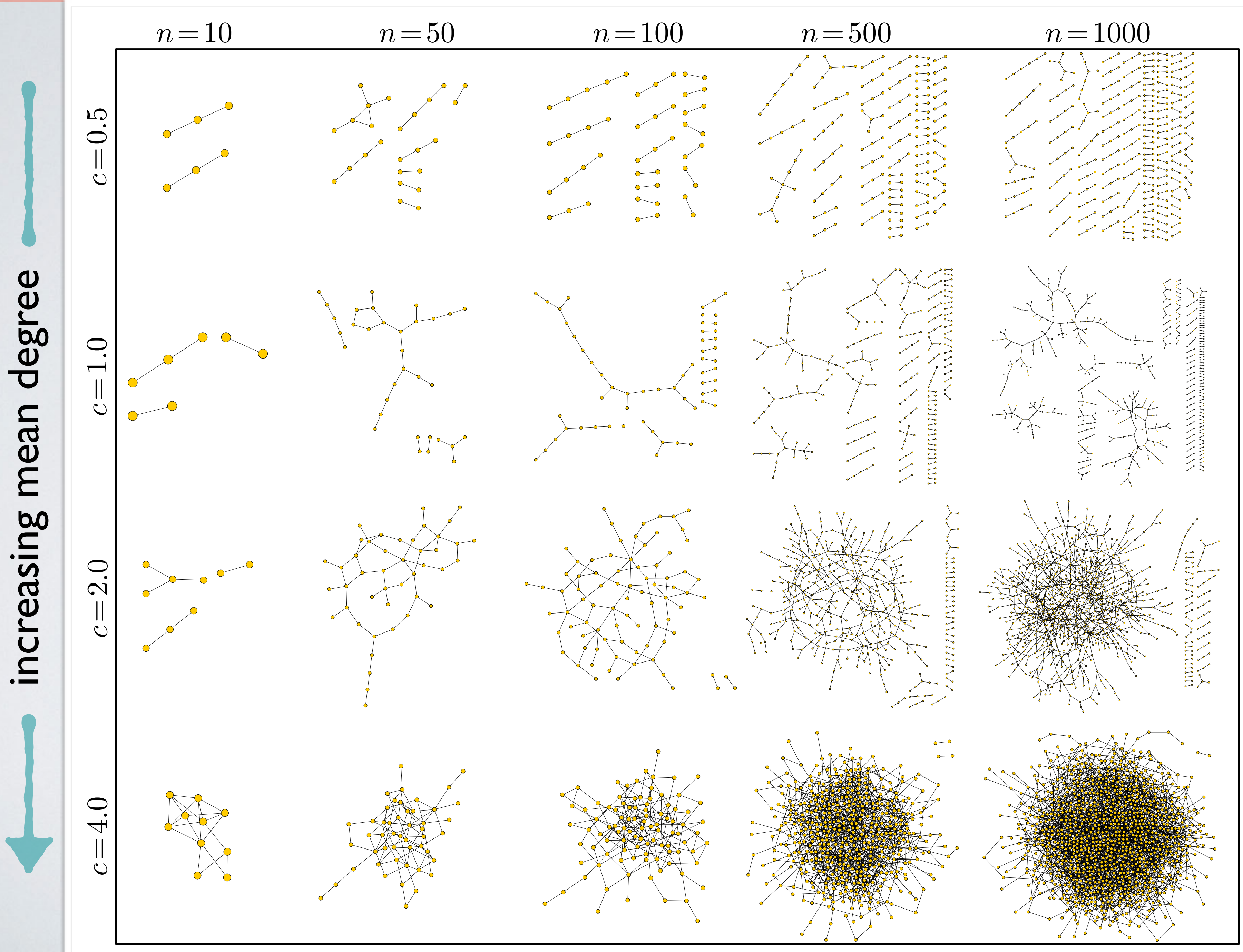
all n vertices within ℓ steps

thus, diameter is $\underbrace{\ell = O(\ln n)}_{\text{a "small" world}}$



* this argument can be made more formal, but yields the same asymptotic result

examples of ER random graphs



how are we doing?

feature	$G(n,p)$	real networks
degree distribution	Poisson	heavy tailed
clustering coefficient	$O(n^{-1})$	social: higher non-social: lower
diameter	$O(\ln n)$	small
large-scale structure	none	communities, dense core, hierarchies, etc.

degree-based random graphs

configuration model : a random graph conditioned on having the *specified* degree sequence $\{k_1, k_2, \ldots, k_n\}$

$$\Pr(i \to j) = \frac{k_i k_j}{2m}$$

* Fosdick et al. SIAM Review 60, 315-355 (2018)
* Chung & Lu, Ann. Comb. 6, 125-145 (2002) specifies a model that produces a *simple graph* with a given degree sequence in *expectation*

degree-based random graphs

configuration model : a random graph conditioned on having the *specified* degree sequence $\{k_1, k_2, \ldots, k_n\}$

$$\Pr(i \to j) = \frac{k_i k_j}{2m}$$

double-edge swap algorithm:

start with a graph G

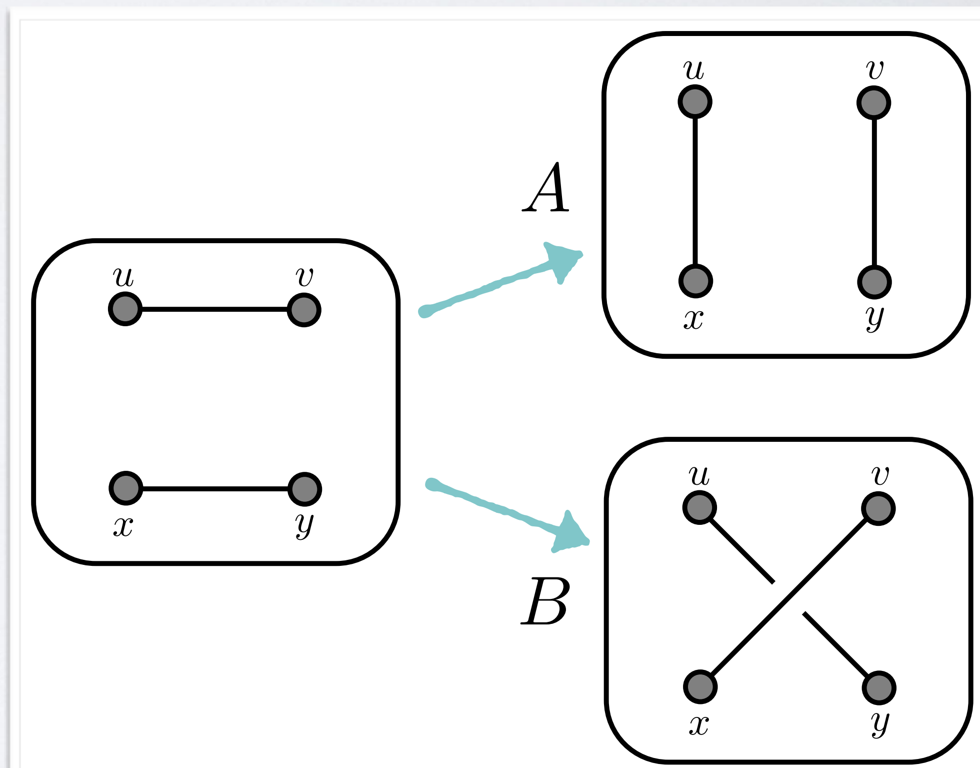
choose $\{(u,v),(x,y)\}$ uniformly

rearrange to A or B

repeat forever

degree preserving*

record a G every now and then

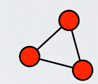
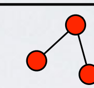


* to sample only *simple* graphs, forbid swaps that would create multi edges or self-loops

* multigraph with given degree sequence can be constructed more quickly by constructing a random matching on edge stubs

degree-based random graphs

configuration model : a random graph conditioned on having the *specified* degree sequence $\{k_1, k_2, \ldots, k_n\}$

clustering coefficient: $C = \dfrac{3 \times \#\text{triangles}}{\#\text{connected triples}}$  

$$= \frac{1}{n} \frac{\left[\langle k^2 \rangle - \langle k \rangle\right]^2}{\langle k \rangle^3} = \underbrace{O(n^{-1})}_{\substack{\text{asymptotically,} \\ \text{zero clustering}}}$$

degree-based random graphs

configuration model : a random graph conditioned on having the *specified* degree sequence $\{k_1, k_2, \ldots, k_n\}$

clustering coefficient: $C = O(n^{-1})$ } asymptotically, zero

diameter: also locally tree-like (if variance of degrees is finite)

following similar argument as ER graphs $\longrightarrow$ $\underbrace{\ell = O(\ln n)}_{\text{a "small" world}}$

degree-based random graphs

the standard **null model** for empirical patterns

defines a probability distribution $\Pr(G \mid \vec{k})$

e.g. from an empirical $G_\circ$
or a preferred $\Pr(k)$

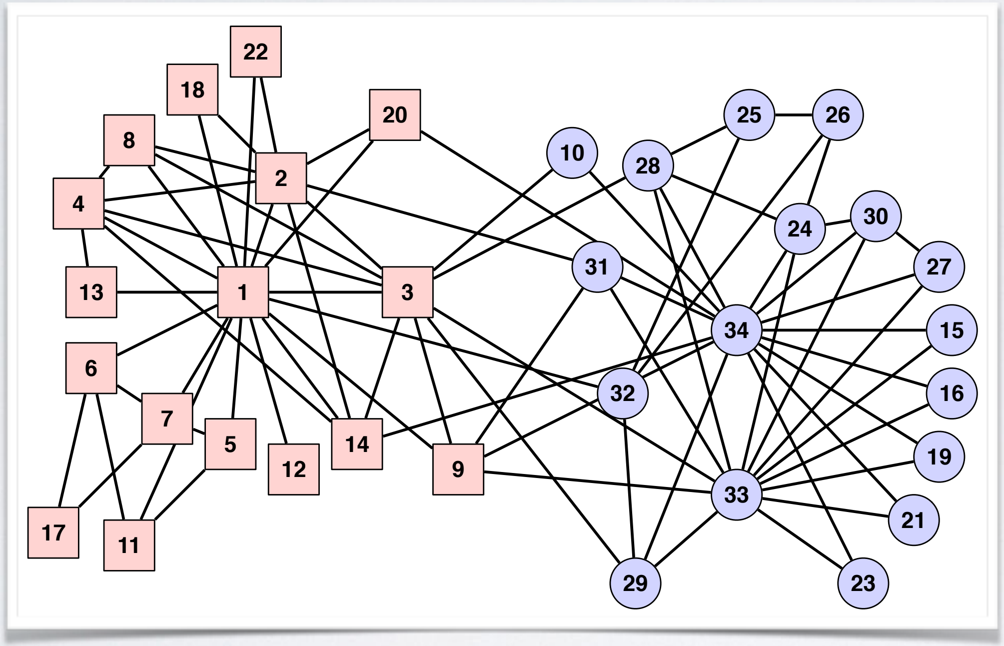
if $f(G_\circ)$ is "typical" within $\Pr(f(G) \mid \vec{k})$

then we say that $\vec{k}$ "explains" $f(G_\circ)$

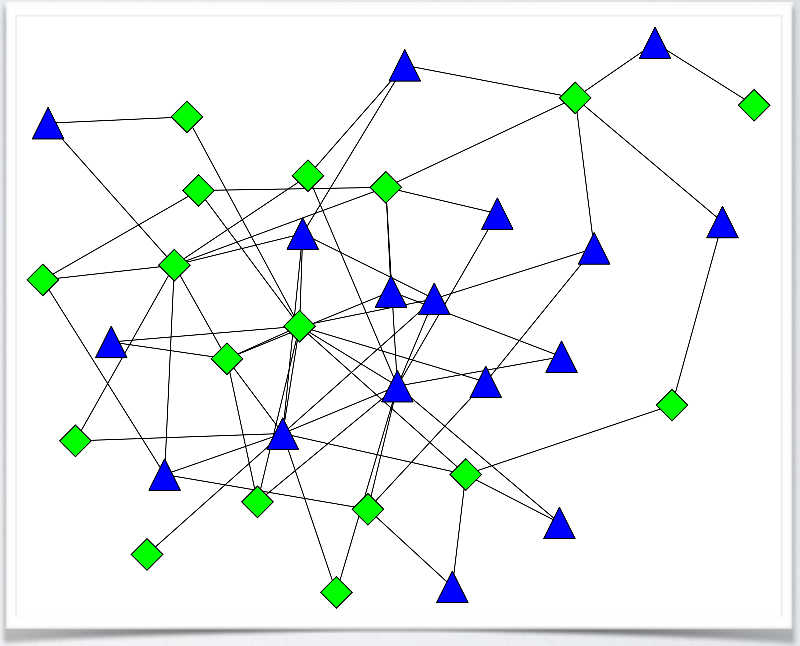
* when a $\Pr(k)$ is drawn from a power-law distribution, we call these "power-law random graphs", which are a popular model for mathematical calculations

degree-based random graphs

the standard **null model** for empirical patterns



Zachary's Karate Club

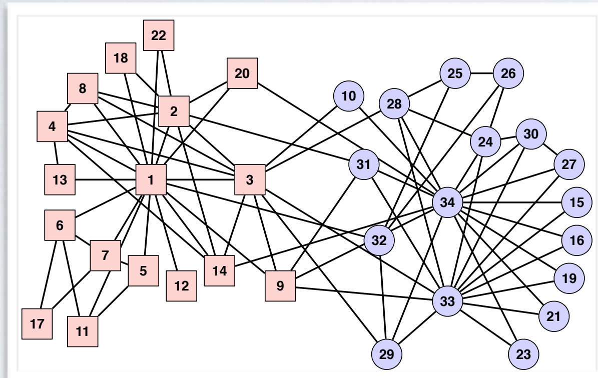


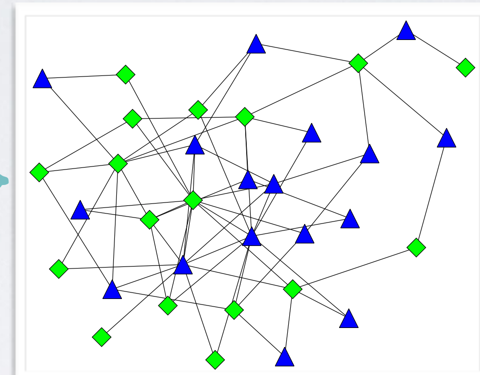
random graph (config. model)

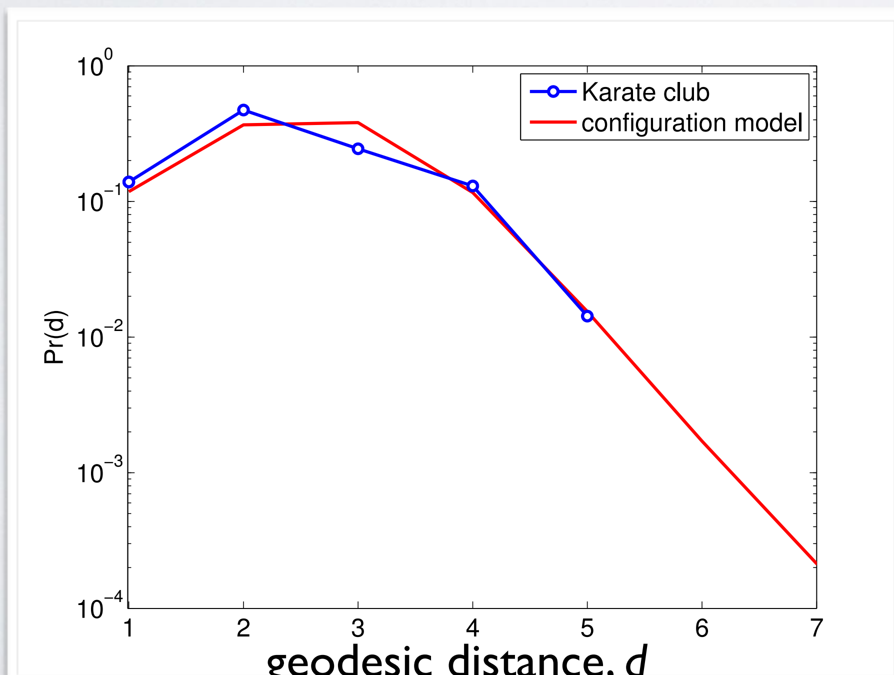
* a very famous network, with its own Prize and its own tumblr

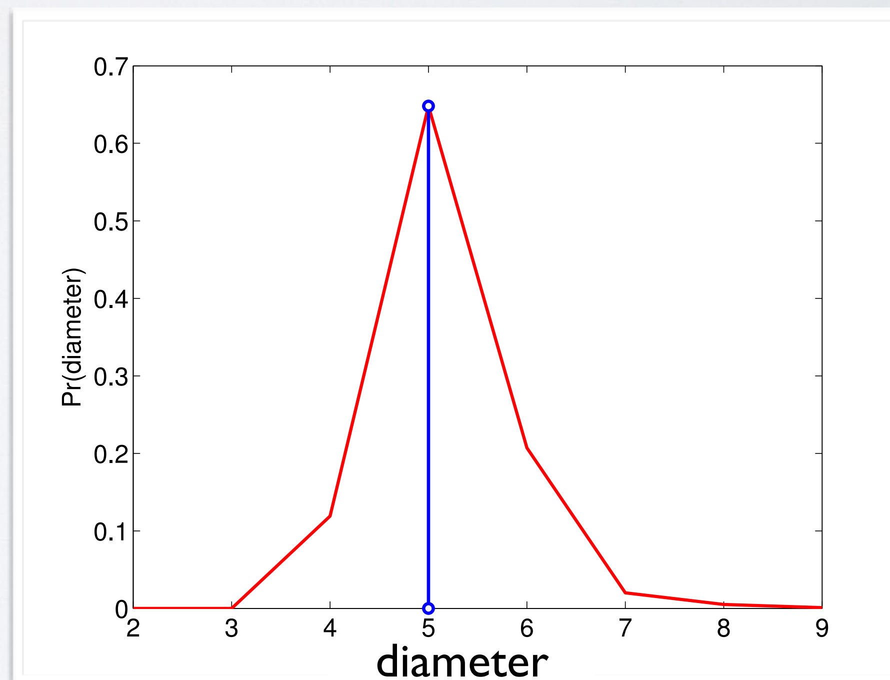
degree-based random graphs

the standard **null model** for empirical patterns





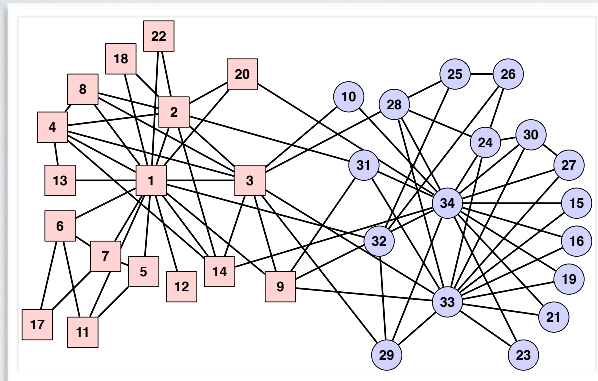


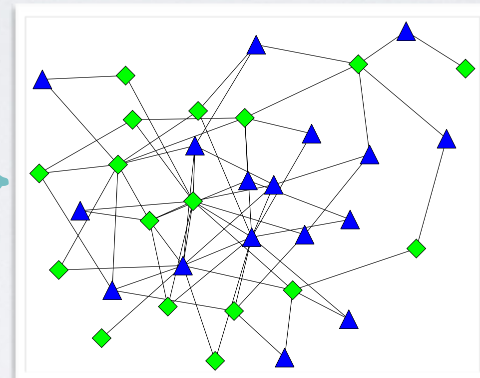


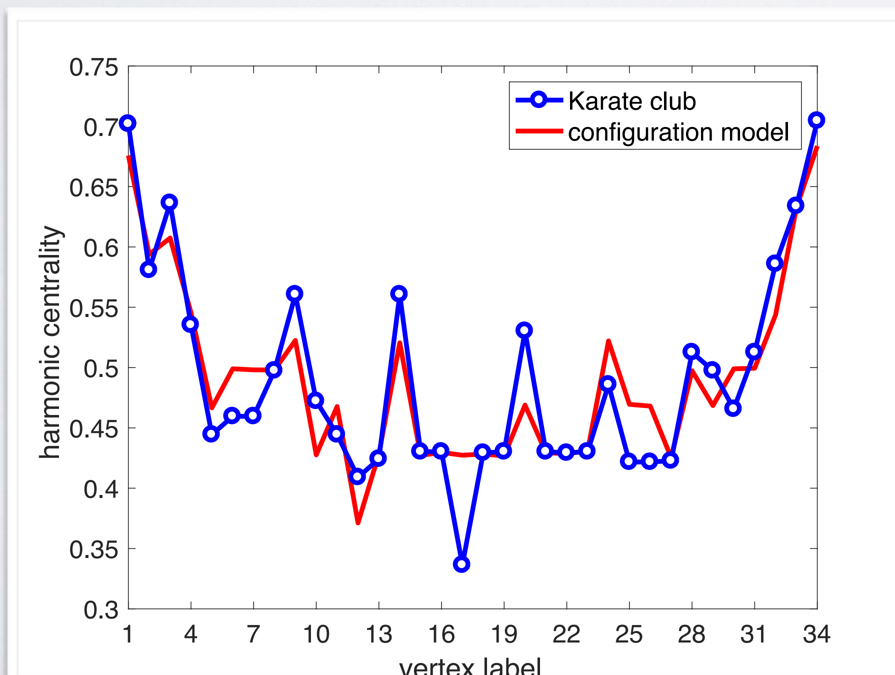
* null distribution from 100 configuration models. what the configuration model gets *wrong* is the community structure. most everything else is well-explained by the degree structure alone

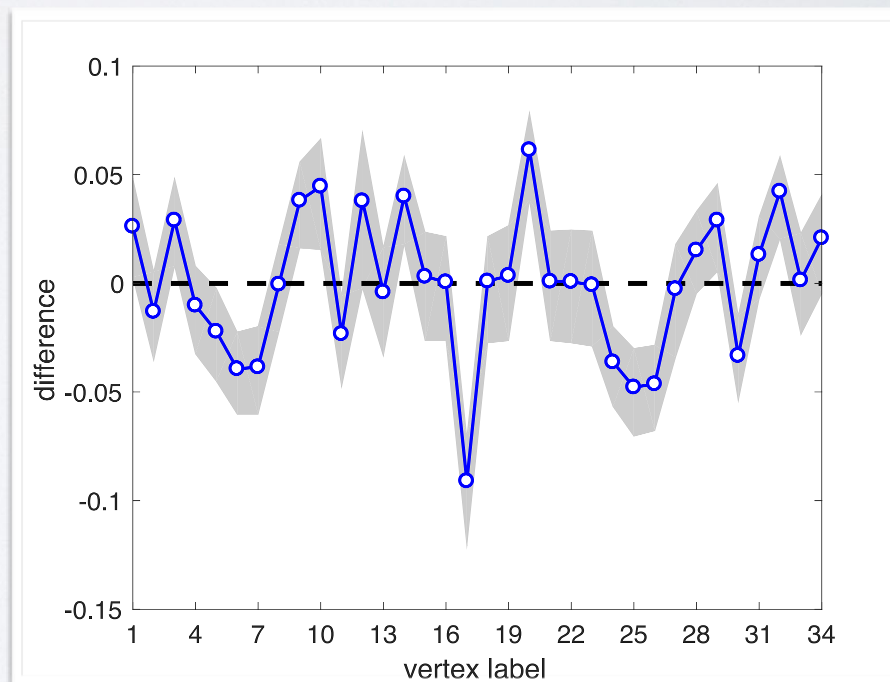
degree-based random graphs

the standard **null model** for empirical patterns









* null distribution from 100 configuration models. what the configuration model gets *wrong* is the community structure. most everything else is well-explained by the degree structure alone

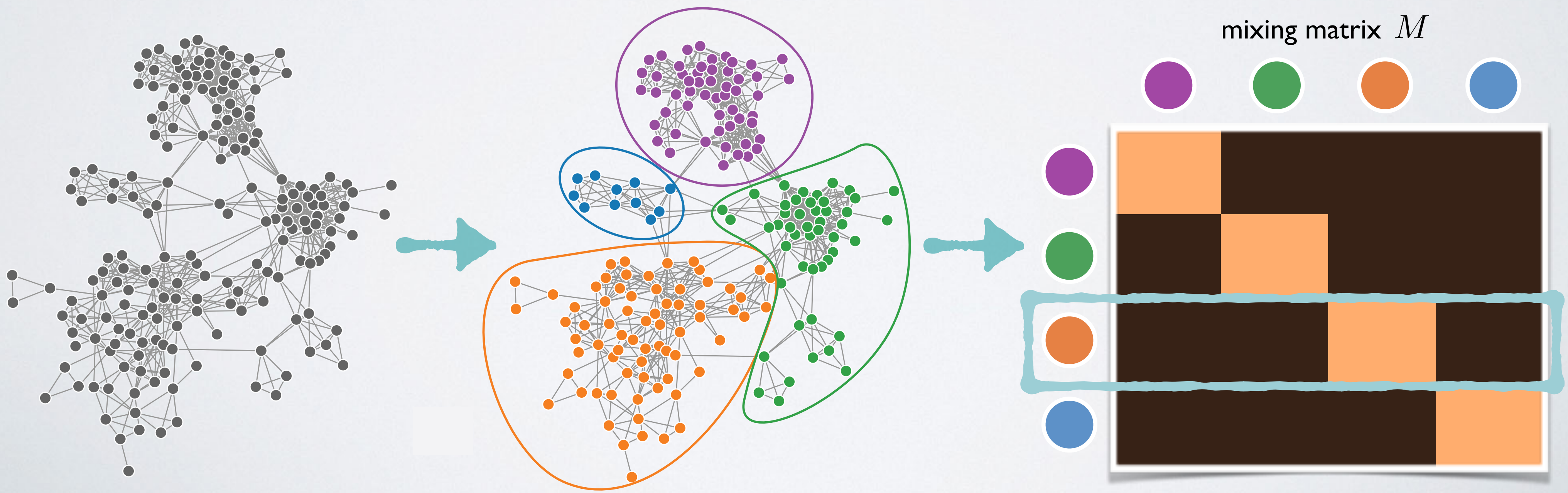
how are we doing?

feature	😐 $G(n,p)$	🙂 configuration	real networks
degree distribution	Poisson	specified	heavy tailed
clustering coefficient	$O(n^{-1})$	$O(n^{-1})$	social: higher non-social: lower
diameter	$O(\ln n)$	$O(\ln n)$	small
large-scale structure	none	none	communities, dense core, hierarchies, etc.

stochastic block models

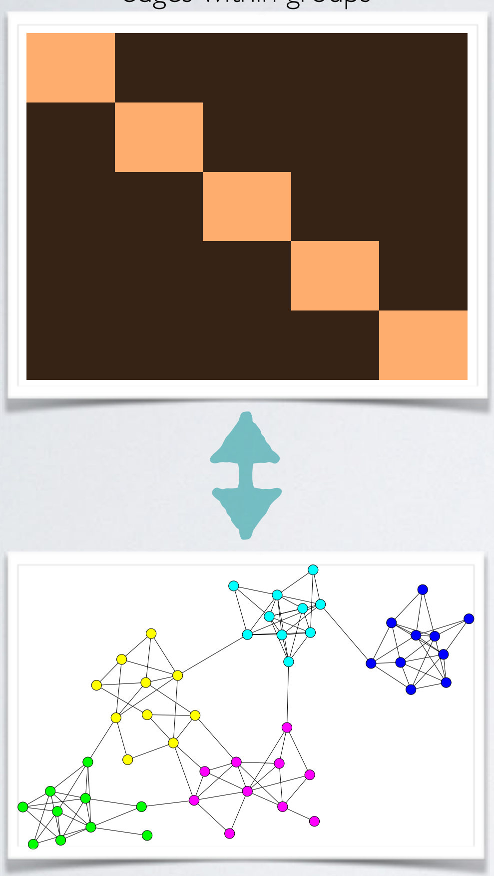
- each vertex i has type $z_i \in \{1, \ldots, k\}$ (k vertex types or groups)
- stochastic block matrix M of group-level connection probabilities
- probability that i, j are connected = M_{z_i, z_j}

community = vertices with same pattern of inter-community connections

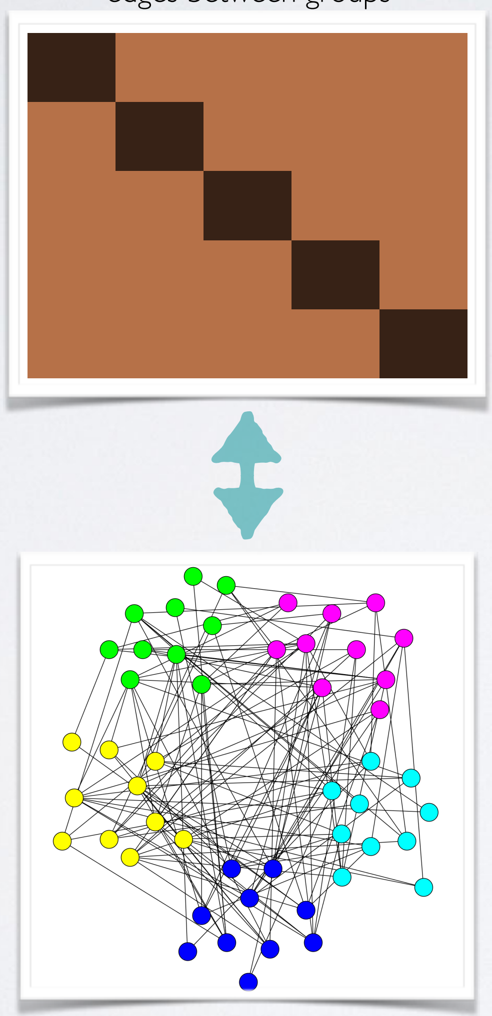


stochastic block models

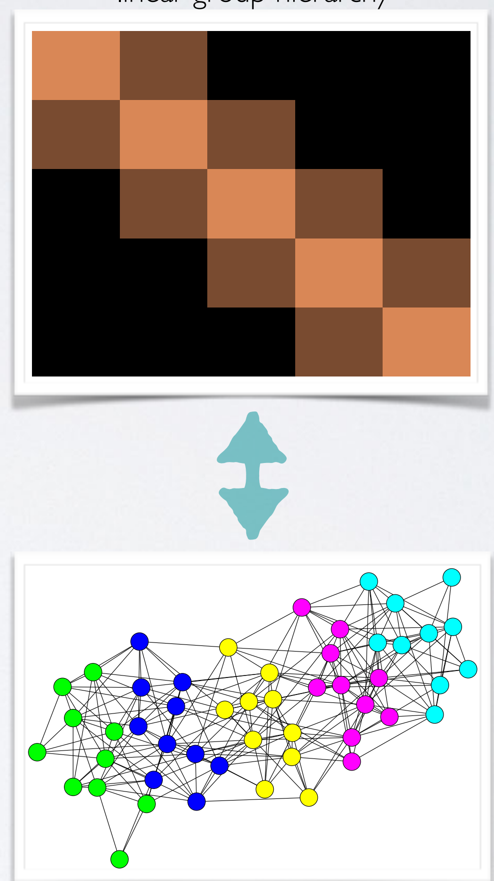
assortative
edges within groups



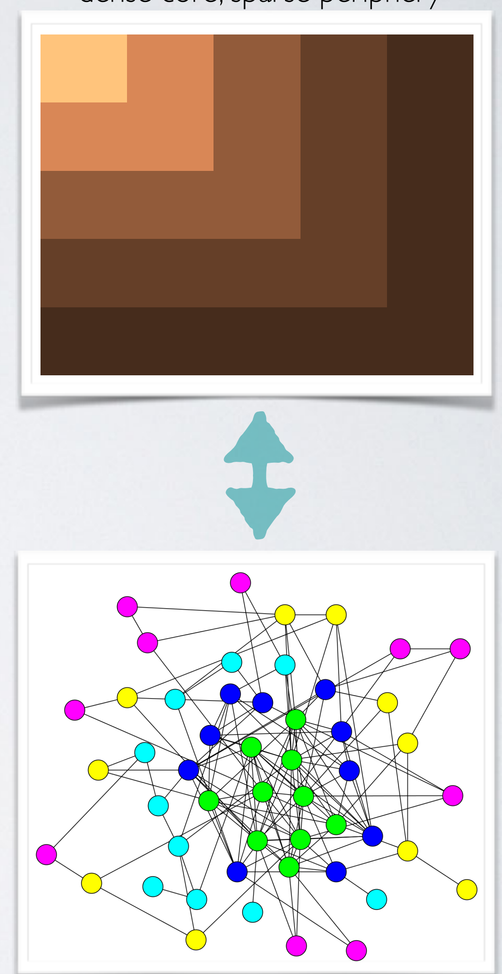
disassortative
edges between groups



ordered
linear group hierarchy



core-periphery
dense core, sparse periphery



stochastic block models

likelihood function

the probability of G given labeling z and block matrix M

$$\Pr(G \mid z, M) = \underbrace{\prod_{(i,j) \in E} M_{z_i, z_j}}_{\text{edge}} \; \underbrace{\prod_{(i,j) \notin E} (1 - M_{z_i, z_j})}_{\text{non-edge probability}}$$

edge / non-edge probability

stochastic block models

likelihood function

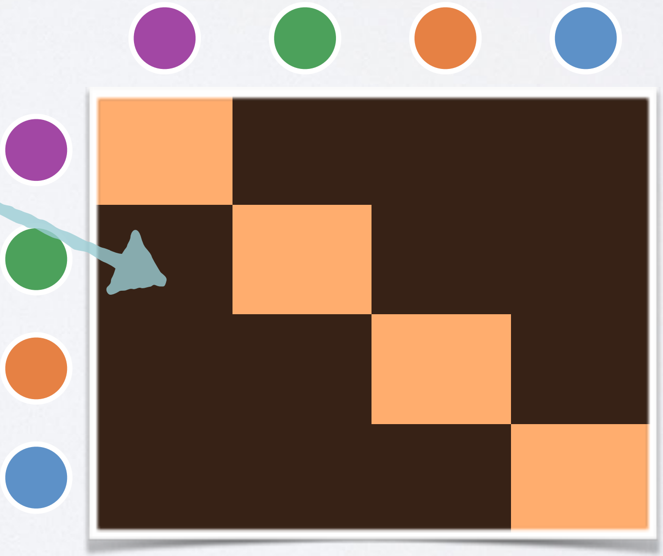
the probability of G given labeling z and block matrix M

$$\Pr(G \mid z, M) = \prod_{(i,j) \in E} M_{z_i, z_j} \prod_{(i,j) \notin E} (1 - M_{z_i, z_j})$$

$$= \prod_{rs} M_{r,s}^{e_{r,s}} (1 - M_{r,s})^{n_s n_r - e_{r,s}}$$

(Bernoulli edges)

Bernoulli random graph
with parameter $M_{r,s}$



stochastic block models

the most general SBM

$$\Pr(A \mid z, \theta) = \prod_{i,j} f\left(A_{ij} \mid \theta_{\mathcal{R}(z_i, z_j)}\right)$$

A_{ij} : value of adjacency

$\mathcal{R}$: partition of adjacencies

f : probability function

$\theta_{a,*}$: pattern for a-type adjacencies

Binomial = simple graphs
Poisson = multi-graphs
Normal = weighted graphs
etc.

θ_{11}	θ_{12}	θ_{13}	θ_{14}
θ_{21}	θ_{22}	θ_{23}	θ_{24}
θ_{31}	θ_{32}	θ_{33}	θ_{34}
θ_{41}	θ_{42}	θ_{43}	θ_{44}

many stochastic block models

stochastic block models

k types of vertices, $\Pr(A_{ij} \mid M, z)$ depends only on node types z_i, z_j
originally invented by sociologists [Holland, Laskey, Leinhardt 1983]

many, many flavors, including

- binomial SBM [Holland et al. 1983, Wang & Wong 1987]
- simple assortative SBM [Hofman & Wiggins 2008]
- mixed-membership SBM [Airoldi et al. 2008]
- hierarchical SBM [Clauset et al. 2006,2008, Peixoto 2014]
- fractal SBM [Leskovec et al. 2005]
- infinite relational model [Kemp et al. 2006]
- degree-corrected SBM [Karrer & Newman 2011]
- SBM + topic models [Ball et al. 2011]
- SBM + vertex covariates [Mariadassou et al. 2010, Newman & Clauset 2016]
- SBM + edge weights [Aicher et al. 2013,2014, Peixoto 2015]
- bipartite SBM [Larremore et al. 2014]
- multilayer SBM [Peixoto 2015, Valles-Catata et al. 2016]
- and many others

one important stochastic block model

degree-corrected SBM (f = Poisson)

* Karrer & Newman, Phys. Rev. E 83, 016107 (2011)

one important stochastic block model

degree-corrected SBM (f = Poisson)

key assumption $\Pr(i \to j) = \theta_i \theta_j \omega_{z_i,z_j}$

stochastic block matrix $\omega_{r,s}$

(degree) propensity of node θ_i

likelihood:

$$\Pr(A \mid z, \theta, \omega) = \prod_{i<j} \frac{\left(\theta_i \theta_j \omega_{z_r,z_j}\right)^{A_{ij}}}{A_{ij}!} \exp\left(-\theta_i \theta_j \omega_{z_r,z_j}\right)$$

where $\underbrace{\hat{\theta}_i = \frac{k_i}{\sum_j k_j \delta_{z_i,z_j}}}_{\text{fraction of } i\text{'s group's stubs on } i}$ $\underbrace{\hat{\omega}_{rs} = m_{rs}}_{\text{total number of edges between } r \text{ and } s} = \sum_{ij} A_{ij} \delta_{z_i,r} \delta_{z_j,s}$

* Karrer & Newman, Phys. Rev. E 83, 016107 (2011)

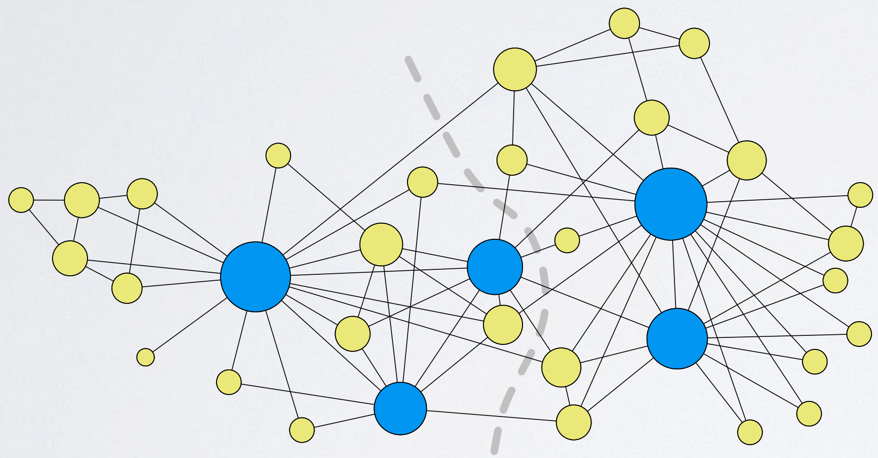
one important stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club

* Karrer & Newman, Phys. Rev. E 83, 016107 (2011)

one important stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club



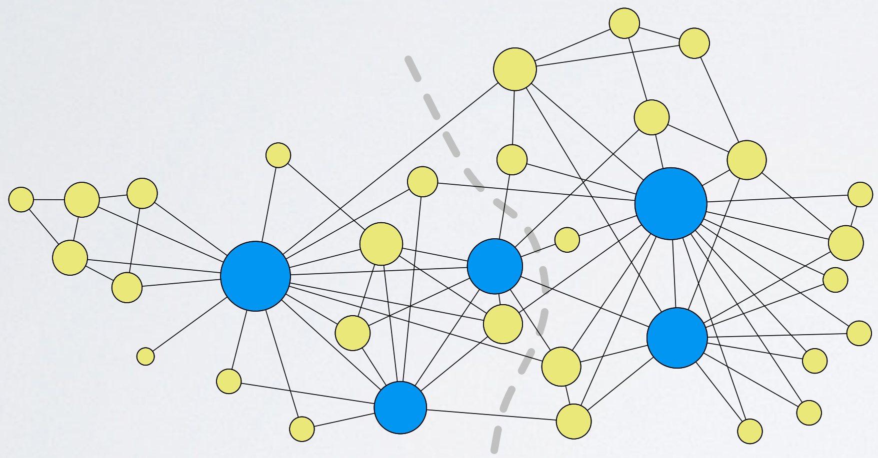
SBM
leader/follower division

DC-SBM
assortative group division

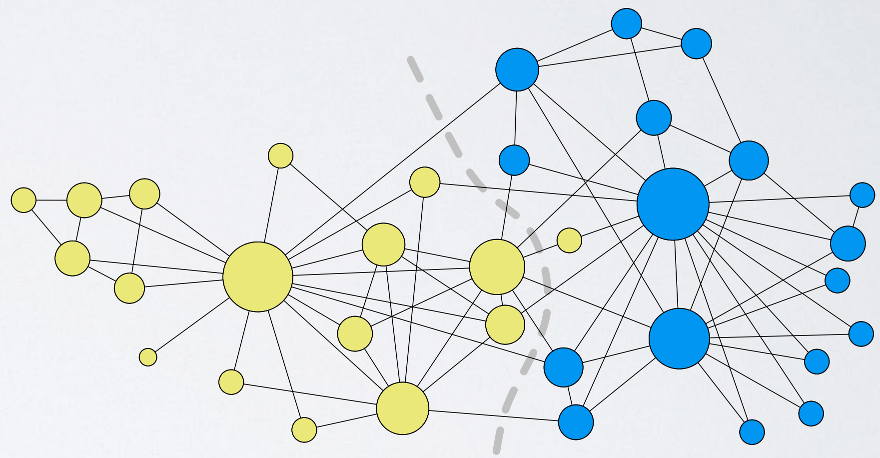
* Karrer & Newman, Phys. Rev. E 83, 016107 (2011)

one important stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club



SBM
leader/follower division

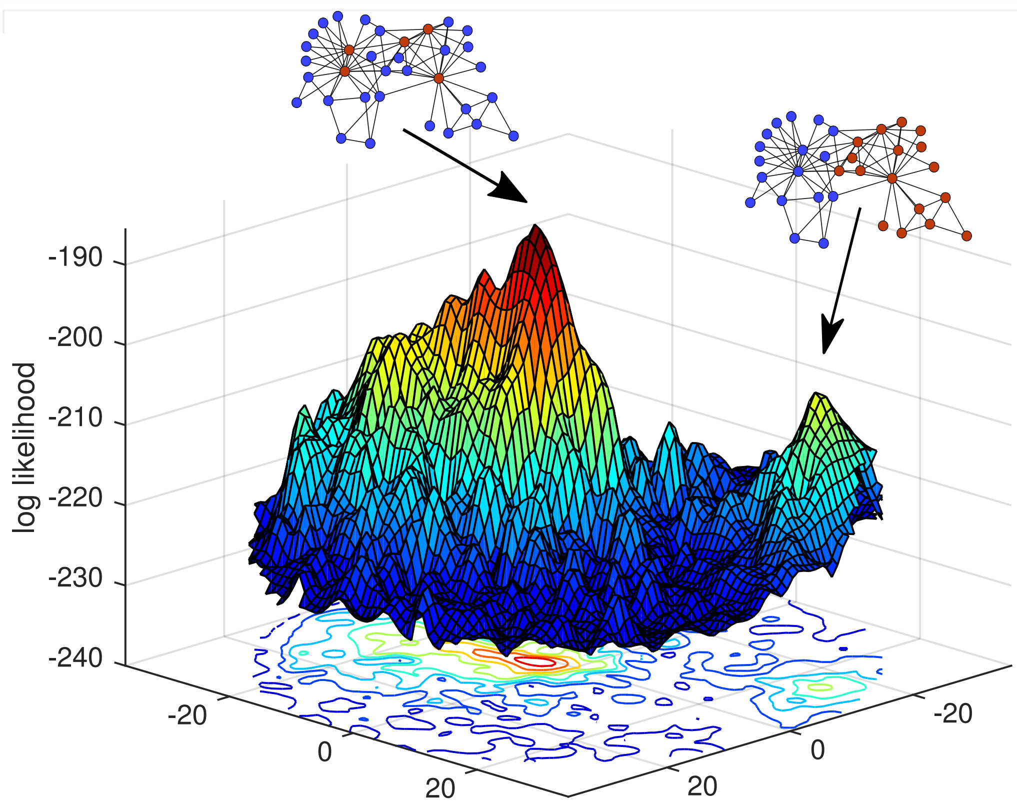


DC-SBM
assortative group division

* Karrer & Newman, Phys. Rev. E 83, 016107 (2011)

different models, different insights

comparing SBM vs. DC-SBM : Zachary karate club



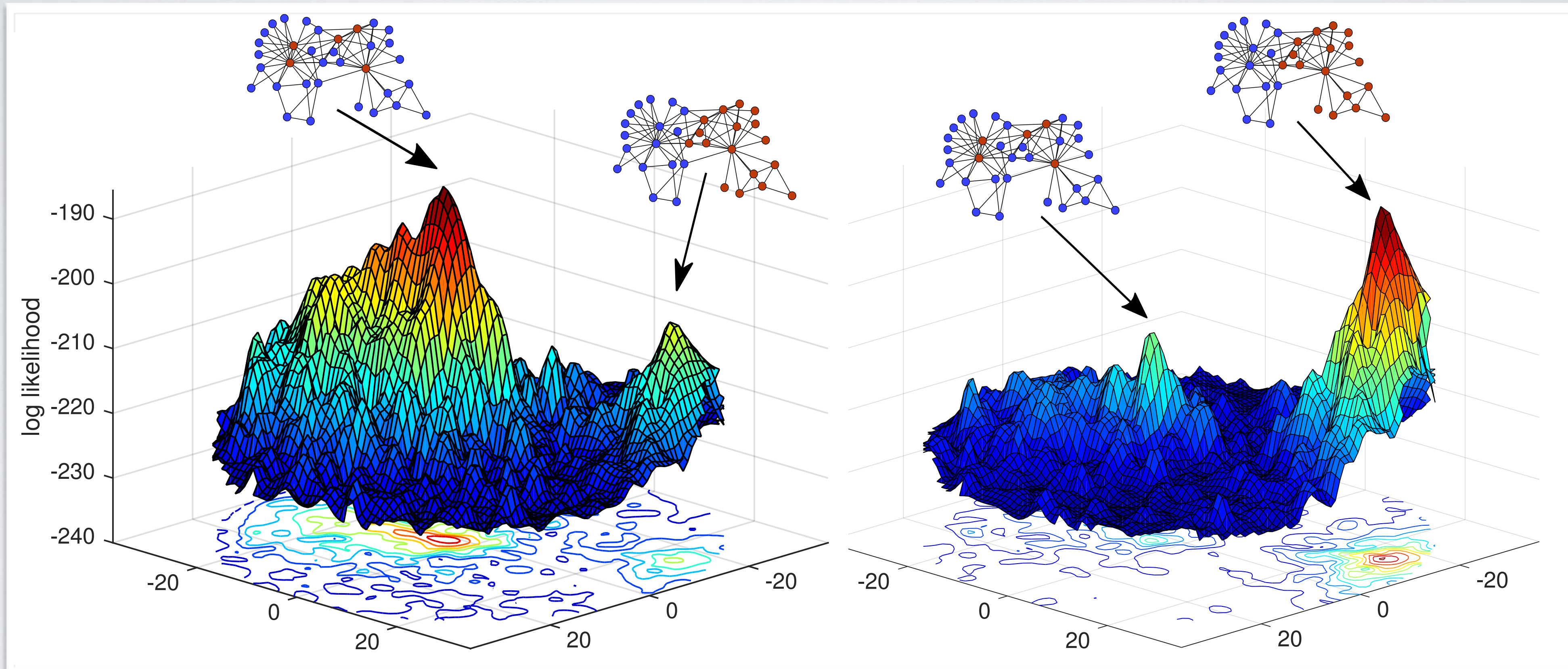
SBM
leader/follower division

DC-SBM
assortative group division

* Peel et al., Science Advances 3, e1602548 (2017)

different models, different insights

comparing SBM vs. DC-SBM : Zachary karate club



SBM
leader/follower division

DC-SBM
assortative group division

* Peel et al., Science Advances 3, e1602548 (2017)

stochastic block models

SBM properties

k Erdos-Renyi random graphs

each with size n_r and internal density $M_{r,r}$

joined pairwise as random bipartite graph with density $M_{r,s}$

degree distribution: mixture of Poissons

diameter: $O(\ln n)$ or $O(\ln(kn))$

triangle density: low, except when $M_{r,s} \gg 0$

local structure: like a random graph

large-scale: mixtures of assortative & disassortative structure

stochastic block models

DC-SBM properties

- k 'configuration model' random multi-graphs
- each with size n_r, internal density $M_{r,r}$ and propensities $\{\theta_i\}_r$
- joined pairwise as random bipartite graph with parameters $M_{r,s}$ and $\{\theta_i\}_{r,s}$
  - degree distribution: arbitrary ($\{\theta_i\}$)
  - diameter: $O(\ln n)$ or $O(\ln(kn))$
  - triangle density: low, except when $M_{r,s} \gg 0$
  - local structure: like a random multi-graph
  - large-scale: mixtures of assortative & disassortative structure

how are we doing?

feature	😐 $G(n,p)$	🙂 configuration	😄 DC SBM	real networks
degree distribution	Poisson	specified	specified	heavy tailed
clustering coefficient	$O(n^{-1})$	$O(n^{-1})$	$O(n^{-1})$	social: high non-social: low
diameter	$O(\ln n)$	$O(\ln n)$	$O(\ln n)$	small
large-scale structure	none	none	specified: communities, hierarchies, etc.	communities, dense core, hierarchies, etc.

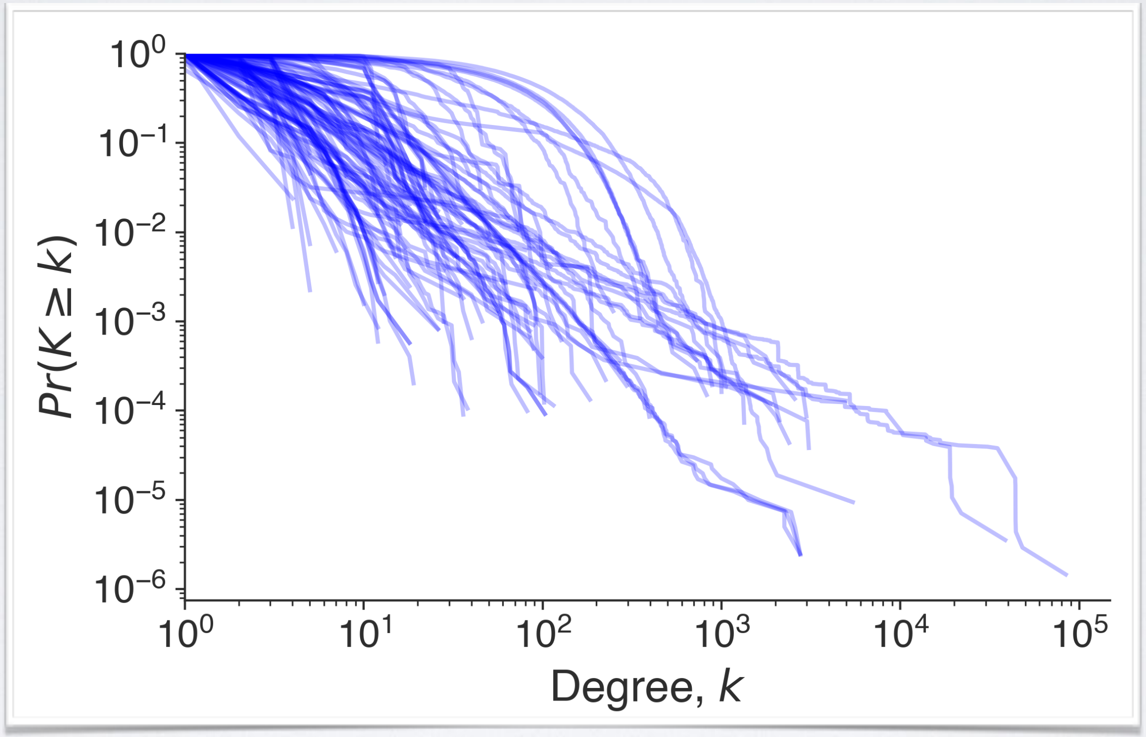
how are we doing?

	😐	🙂	😄	🤔
feature	$G(n,p)$	configuration	DC SBM	real networks
degree distribution	Poisson	specified	specified	heavy tailed
clustering coefficient	$O(n^{-1})$	$O(n^{-1})$	$O(n^{-1})$	social: high non-social: low
diameter	$O(\ln n)$	$O(\ln n)$	$O(\ln n)$	small
large-scale structure	none	none	specified: communities, hierarchies, etc.	communities, dense core, hierarchies, etc.

what patterns *do* real networks exhibit?

degree distributions:

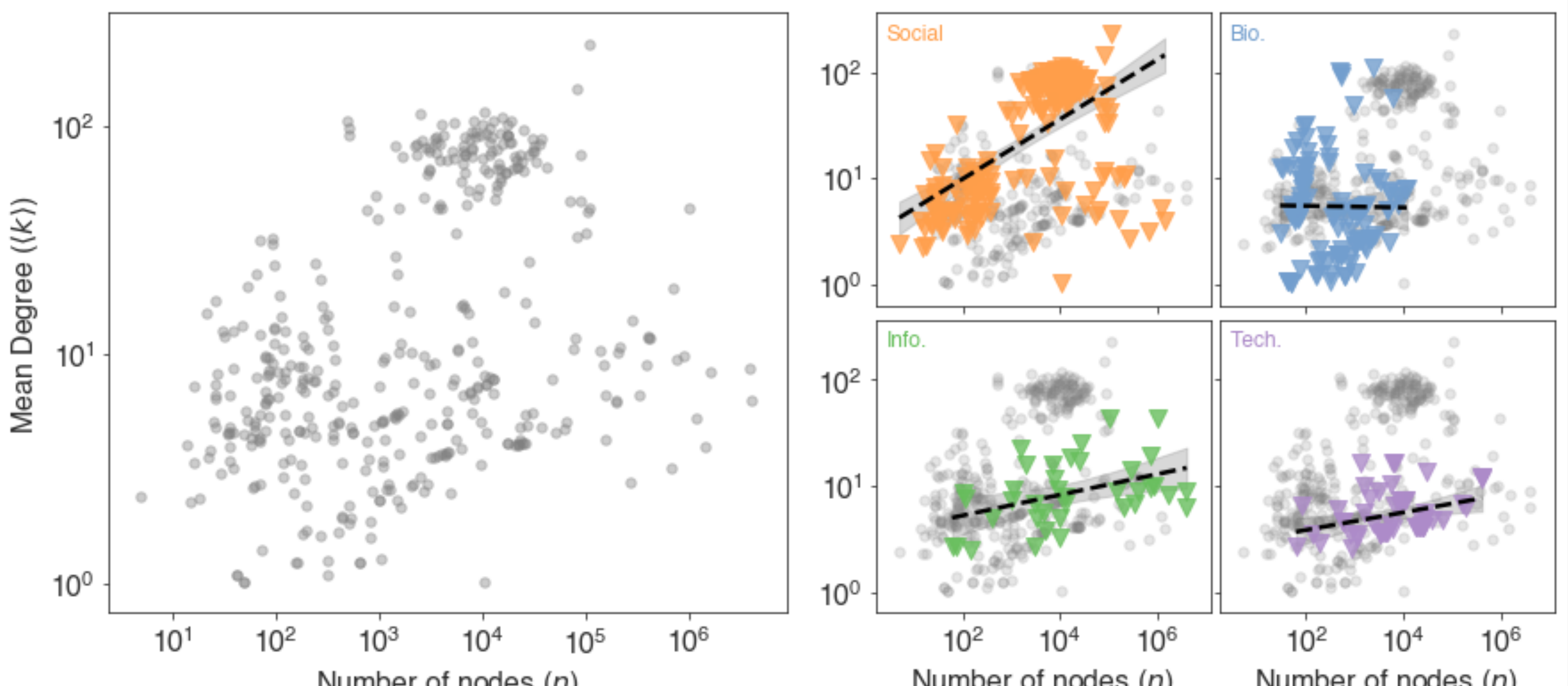
✅ **heavy-tailed**, with enormous diversity across networks and domains



what patterns *do* real networks exhibit?

mean degree (are networks sparse?):

🤔 $O(n^{\alpha})$, social networks generally far more dense than other types

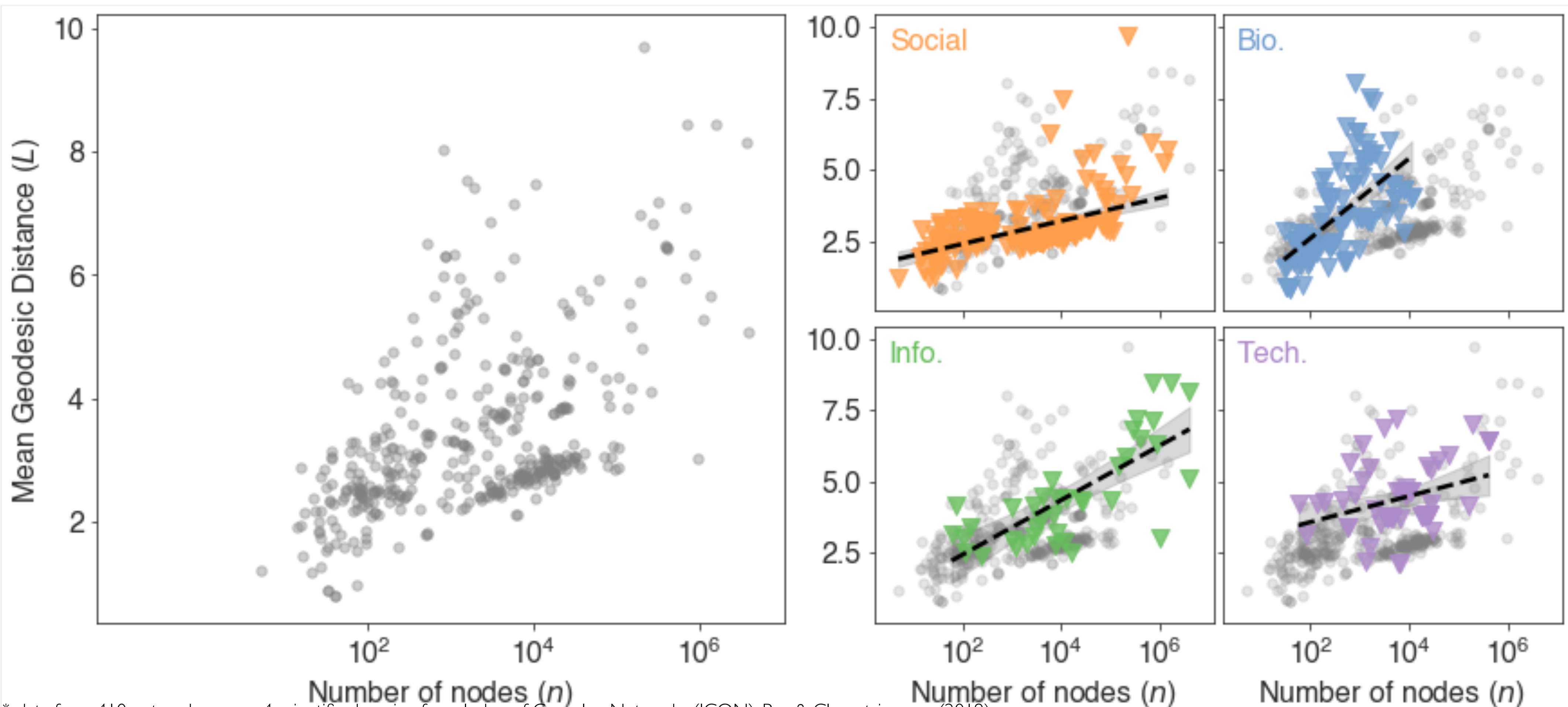


* data from 410 networks across 4 scientific domains, from Index of Complex Networks (ICON); Ray & Clauset, in prep (2019)

what patterns *do* real networks exhibit?

mean geodesic distance (also, diameter):

🥳 $O(\ln n)$, but with ***different coefficients*** for different domains

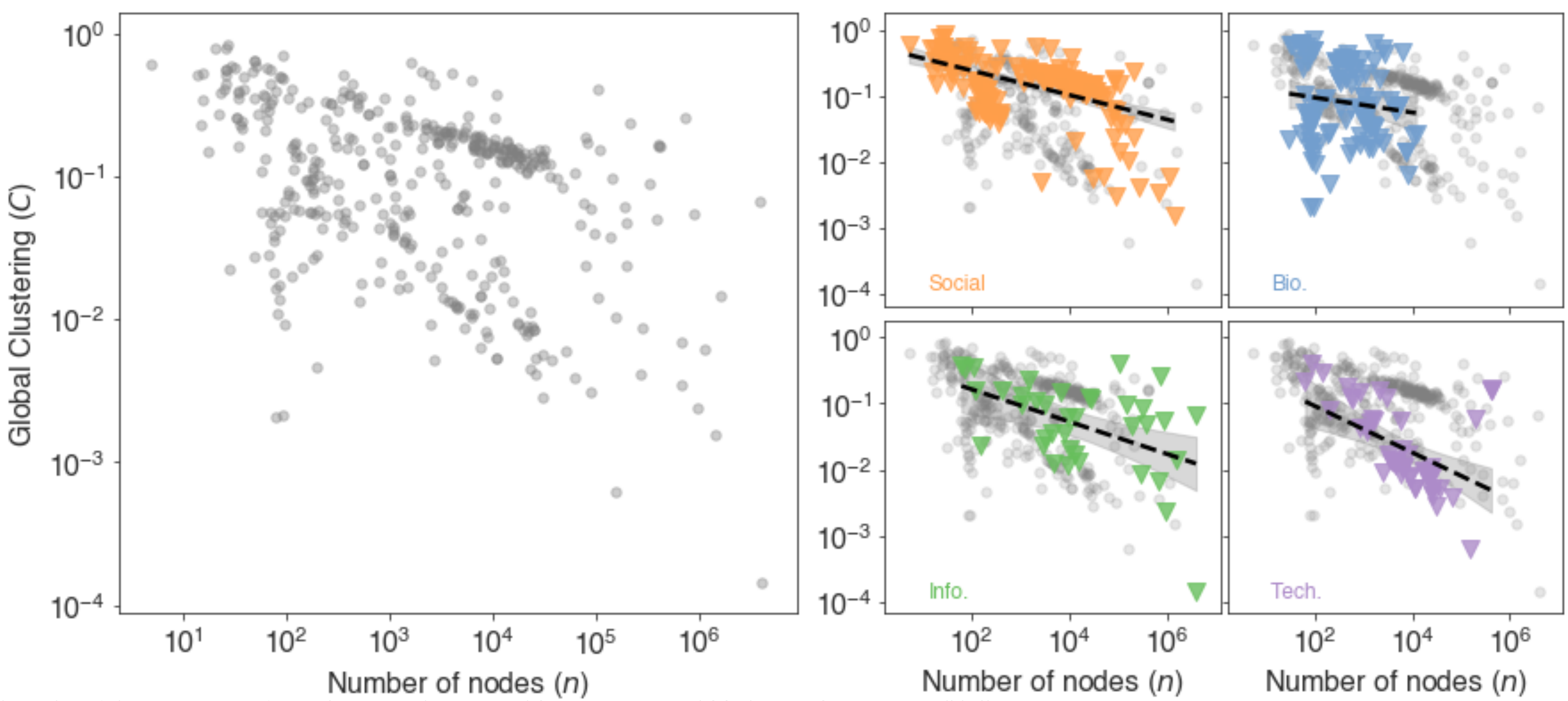


* data from 410 networks across 4 scientific domains, from Index of Complex Networks (ICON); Ray & Clauset, in prep (2019)

what patterns *do* real networks exhibit?

clustering coefficient:

😮 $O(n^{-1})$, social networks have $5-10\times$ more triangles at a given scale n, but all networks scale down

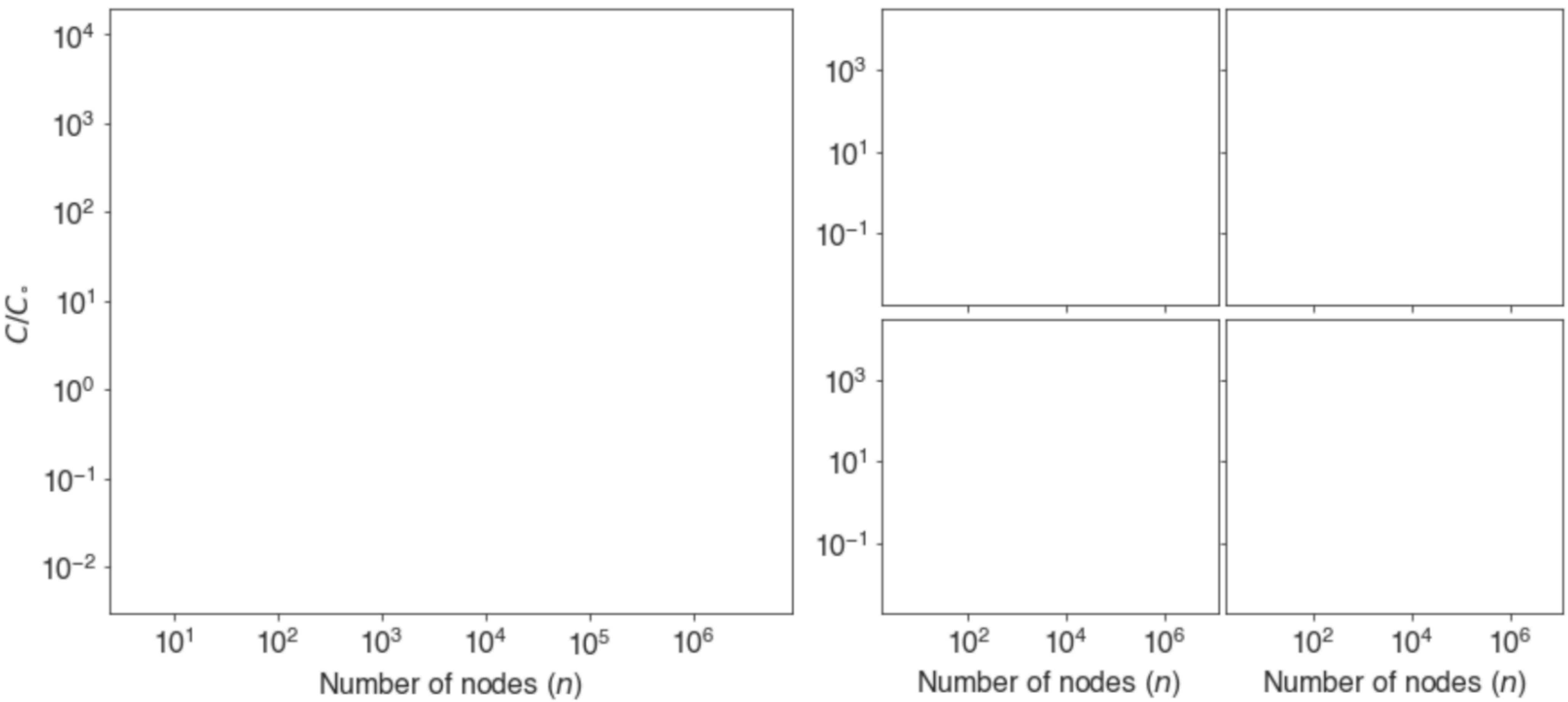


* data from 410 networks across 4 scientific domains, from Index of Complex Networks (ICON); Ray & Clauset, in prep (2019)

what patterns *do* real networks exhibit?

how much of clustering coefficient is due to degree structure?

null models! compare empirical vs. configuration model: C/C_o

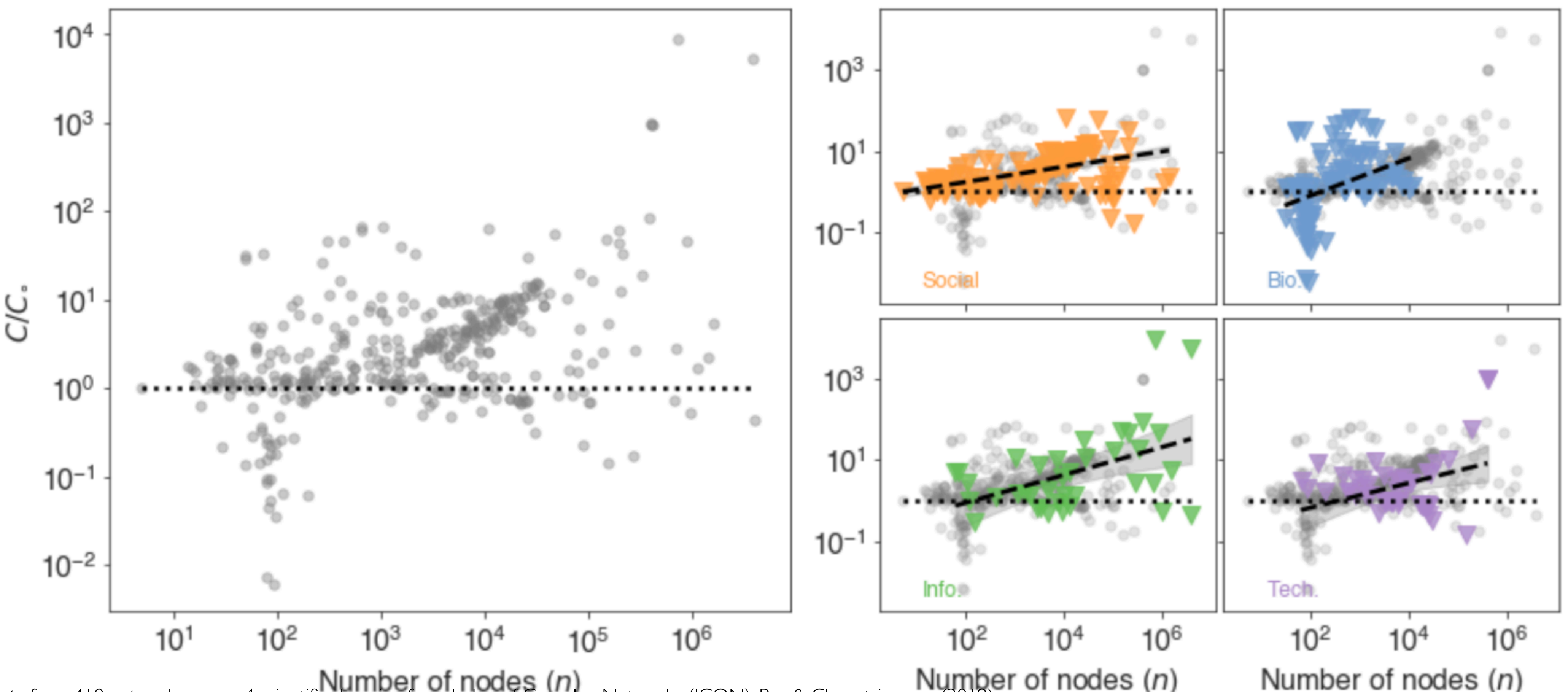


what patterns *do* real networks exhibit?

how much of clustering coefficient is due to degree structure?

😮 social networks' higher C is ***explained*** by their degree distributions

all domains exhibit similar triangle-enrichment across scales (a bit more for bio)



* data from 410 networks across 4 scientific domains, from Index of Complex Networks (ICON); Ray & Clauset, in prep (2019)

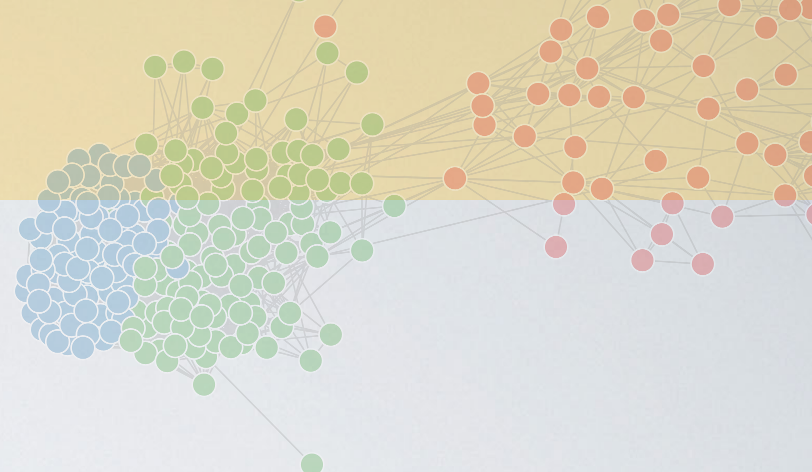
how are we doing?

	😐	🙂	😄	🥳
feature	$G(n, p)$	configuration	DC SBM	real networks
degree distribution	Poisson	specified	specified	heavy tailed
clustering coefficient	$O(n^{-1})$	$O(n^{-1})$	$O(n^{-1})$	$O(n^{-1})$
diameter	$O(\ln n)$	$O(\ln n)$	$O(\ln n)$	$O(\ln n)$
large-scale structure	none	none	specified: communities, hierarchies, etc.	communities, dense core, hierarchies, etc.

parting thoughts on networks

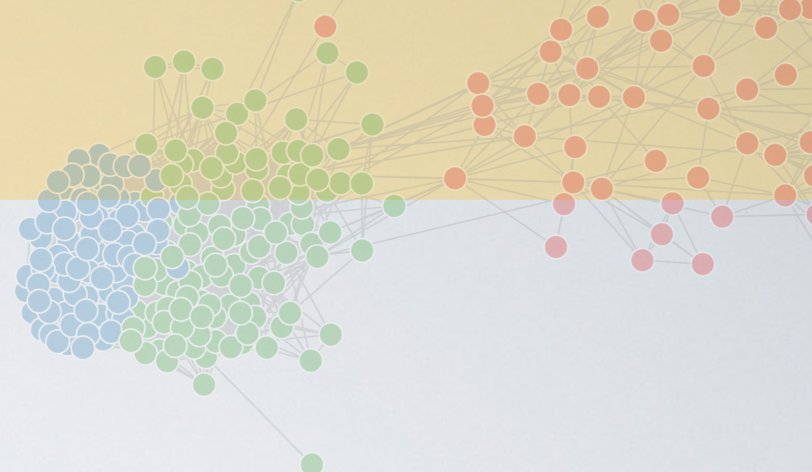
- networks are cool!

parting thoughts on networks



- **networks are cool!**
  but also complicated objects = enormous structural diversity
  many ways to describe a network's structure
- **null models & statistical inference**
  among most powerful tools for describing network structure
  highly flexible, scalable, useful
  auxiliary data (weights, attributes, time)
  applications abound [new ideas often come from these]
- **structure + dynamics = function**
  how does structure constrain dynamics, robustness, etc.
  to what degree does structure = function?

analyzing networks



6 major approaches

1. **exploratory data analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)
2. **simple regressions:** convert network structure into node-level features, and do traditional explanatory modeling
3. **null models:** use some kind of random graph to identify non-random patterns as deviations from the null
4. **mechanisms / simulations:** explain structural or dynamical patterns as caused by specific process
5. **predictive models:** fit parametric model of network structure & use it to predict missing or future data (edges, labels, etc.)
6. **network experiments:** manipulate structure and measure node-level or graph-level behavior as function of changes

end of lecture 3

end of lecture 3

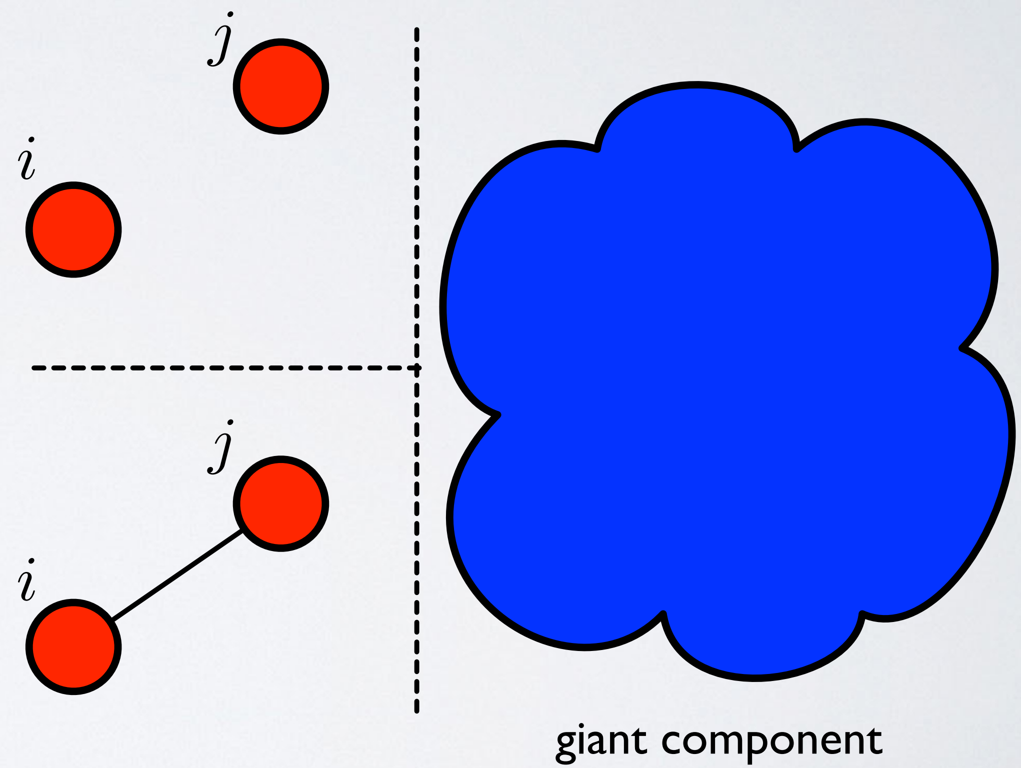
giant component (Erdos-Renyi graph)

let u be fraction of vertices *not* in **giant component**

for i not to be in the giant component, then for every j

1. i is not connected to j, or

2. i connects to j, and j is not part of the giant component



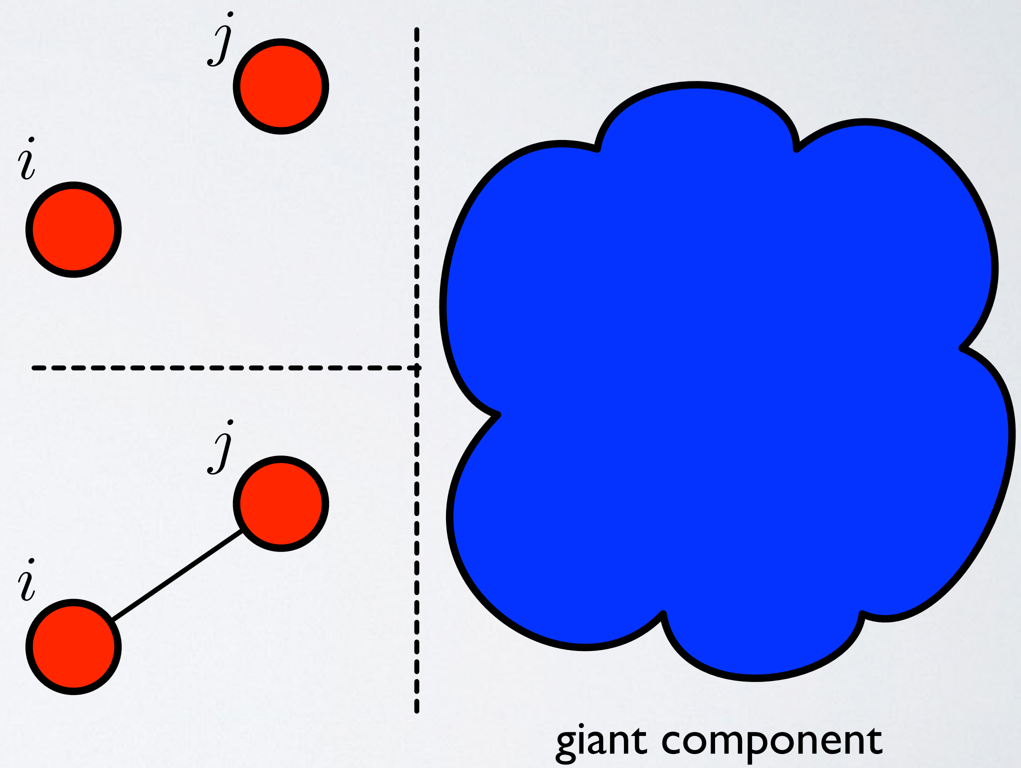
giant component (Erdos-Renyi graph)

let u be fraction of vertices *not* in **giant component**

for i not to be in the giant component, then for every j

1. with probability $1-p$

2. with probability pu



giant component (Erdos-Renyi graph)

total probability that i **not** in giant component via any of the $n-1$ choices of j:

$$u = (1 - p + pu)^{n-1} = \left[1 - \frac{c}{n-1}(1-u)\right]^{n-1}$$

remember: $p = c/(n-1)$

giant component (Erdos-Renyi graph)

total probability that i **not** in giant component via any of the $n-1$ choices of j:

$$u = (1-p+pu)^{n-1} = \left[1 - \frac{c}{n-1}(1-u)\right]^{n-1}$$

taking logs of both sides, and approximating:

$$\begin{aligned}\ln u &= (n-1)\ln\left[1 - \frac{c}{n-1}(1-u)\right] \\ &\approx -(n-1)\frac{c}{n-1}(1-u) \\ &= -c(1-u)\end{aligned}$$

remember: $\ln(1+x) \approx x$ if $x \ll 1$

giant component (Erdos-Renyi graph)

total probability that i **not** in giant component via any of the $n-1$ choices of j:

$$\boxed{u = \mathrm{e}^{-c(1-u)}}$$

and the fraction of vertices **in** the giant component is

$$S = 1 - u$$

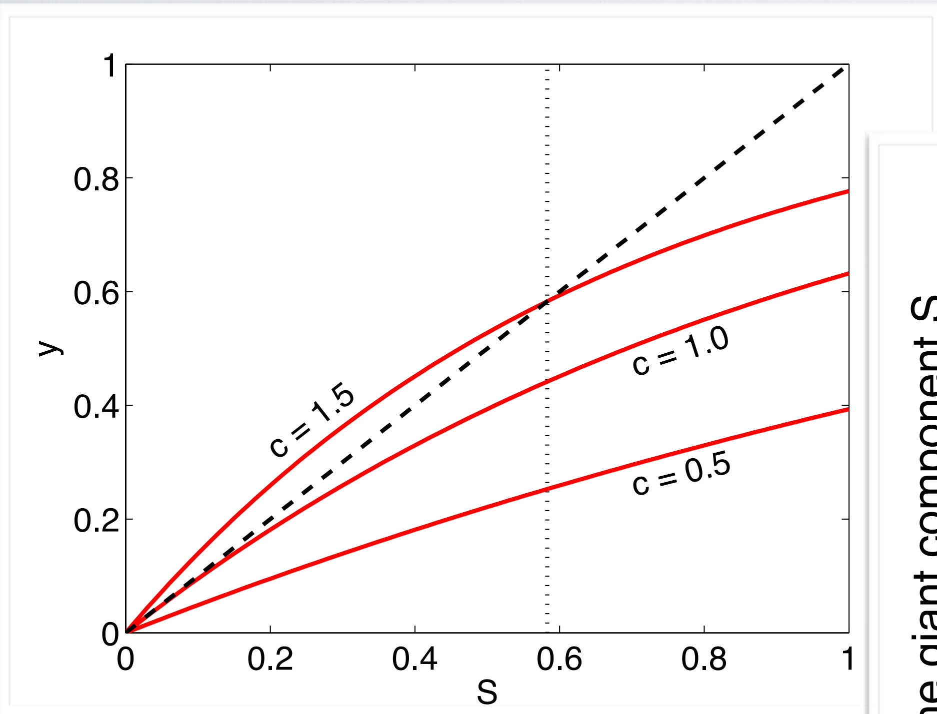
eliminating u for S yields the transcendental equation

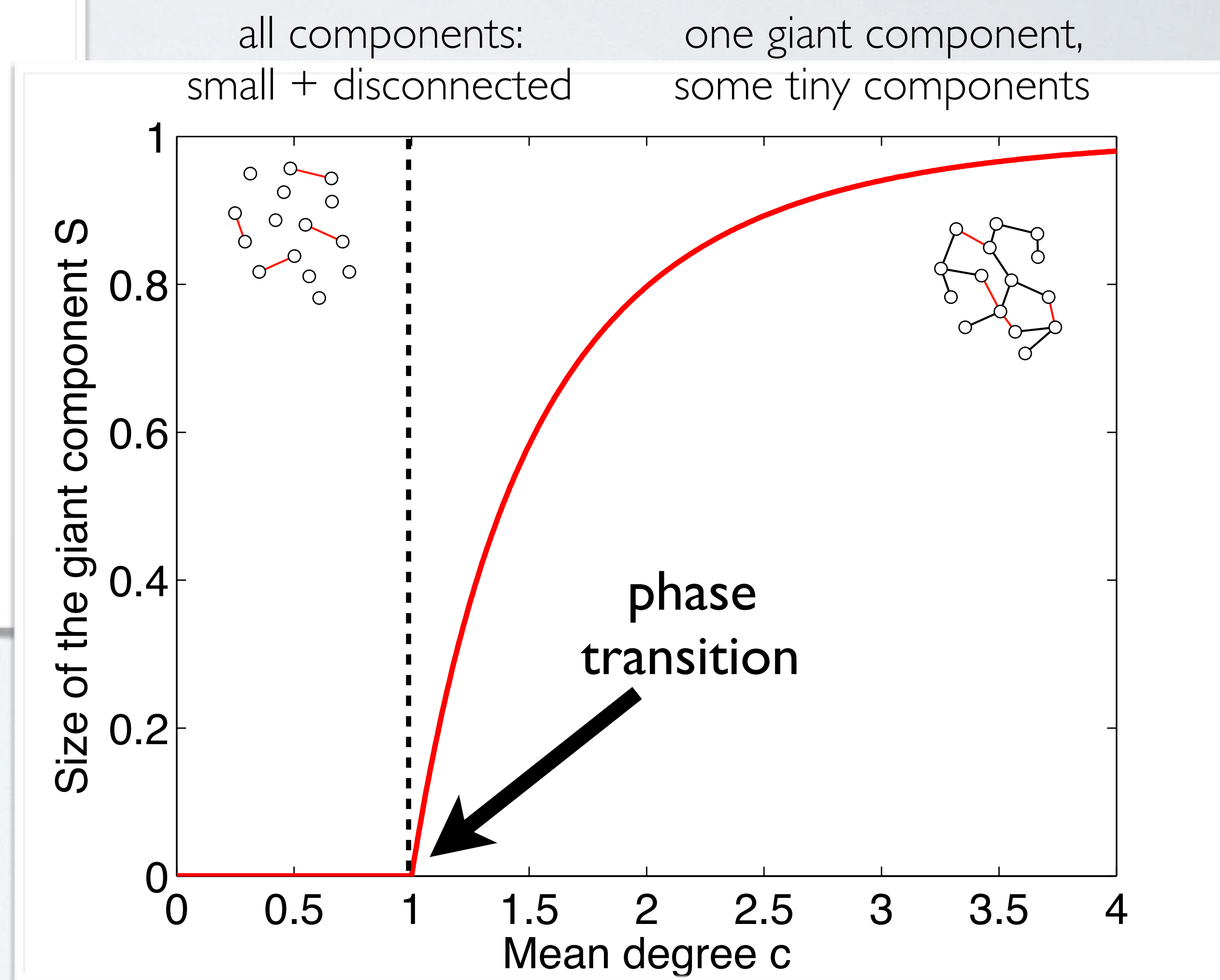
$$S = 1 - \mathrm{e}^{-c\,S}$$

[first given by Erdos and Renyi in 1959]

giant component

size of the giant component: $S = 1 - \mathrm{e}^{-c\,S}$





citation networks

example of a **dynamic, growing network** model

example of a network **mechanistic** model

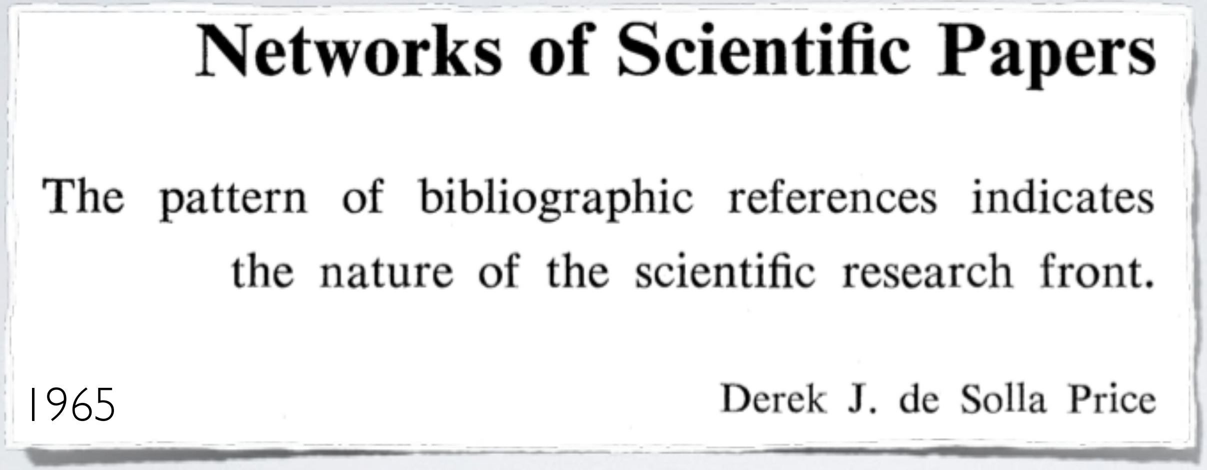
ample data

pleasing narcissistic qualities

long history of study

generally well understood

citation networks


Networks of Scientific Papers

The pattern of bibliographic references indicates the nature of the scientific research front.

1965

Derek J. de Solla Price

Price's model:

- papers are published continually [growing network]
- each paper has bibliography of length c [mean out degree]
- new papers cite previously published only [directed acyclic graph]
- attachment mechanism:

so long as $\Pr(c)$ has finite variance, the model is well defined

citation networks

Networks of Scientific Papers

The pattern of bibliographic references indicates the nature of the scientific research front.

1965 Derek J. de Solla Price

Price's model:

- papers are published continually [growing network]
- each paper has bibliography of length c [mean out degree]
- new papers cite previously published only [directed acyclic graph]
- attachment mechanism:

$$p(j \text{ cites some paper } i) \propto k_i + a$$

preferential attachment

uniform attachment

preferential attachment = cumulative advantage = the Matthew effect

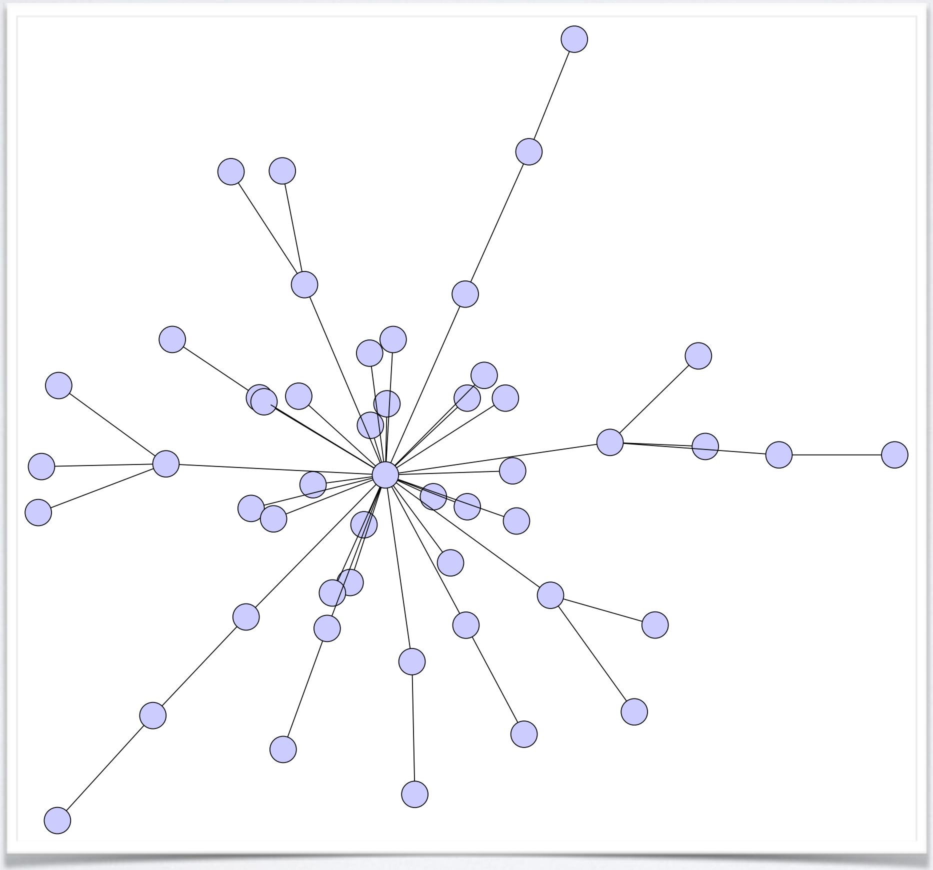
so long as $\Pr(c)$ has finite variance, the model is well defined

preferential attachment networks

$$n = 50$$

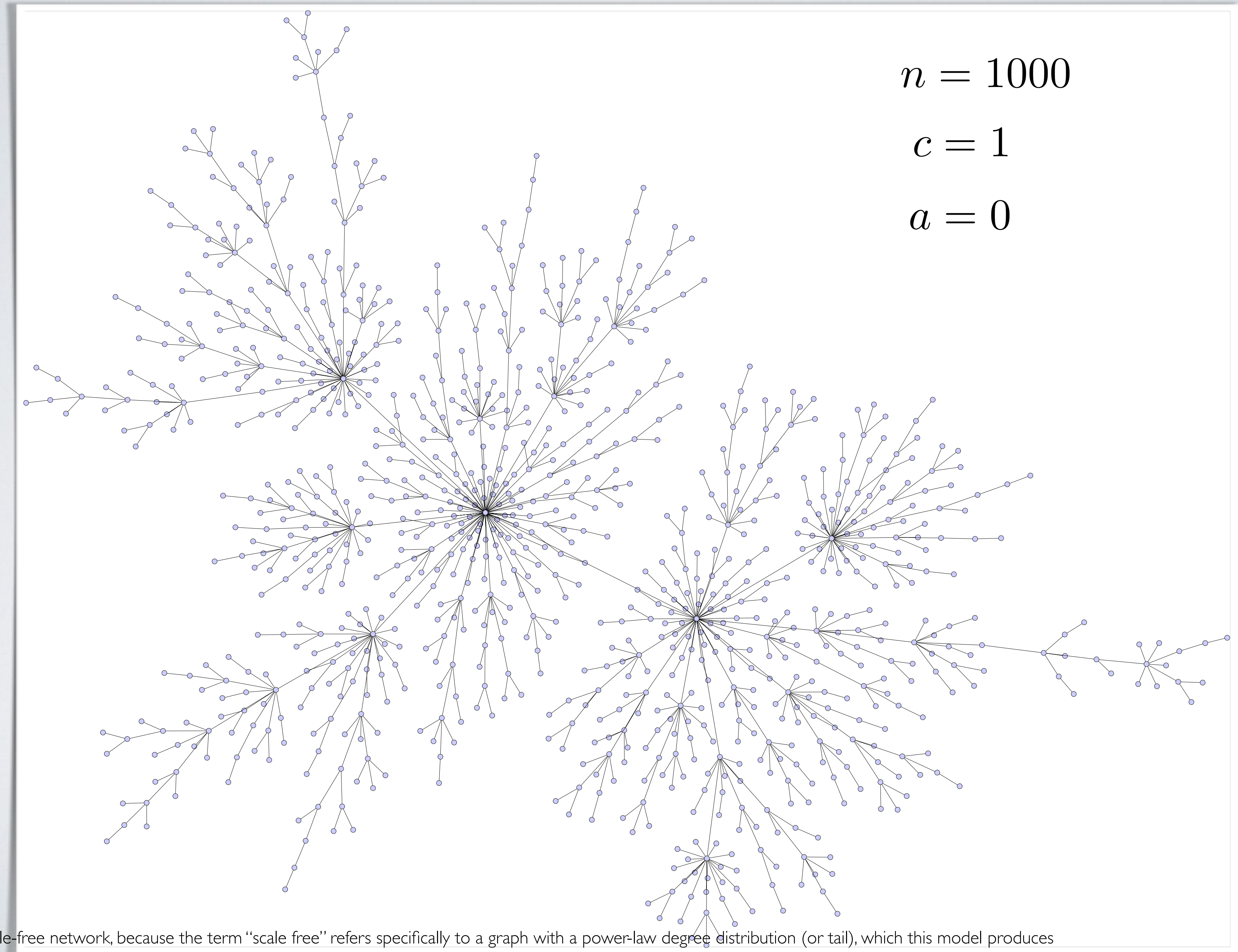
$$c = 1$$

$$a = 0$$



* this is a scale-free network, because the term "scale free" refers specifically to a graph with a power-law degree distribution (or tail), which this model produces

preferential attachment networks



* this is a scale-free network, because the term "scale free" refers specifically to a graph with a power-law degree distribution (or tail), which this model produces

degree distribution

exactly solvable in the limit [originally by Simon 1955]

$$p_k = \frac{\mathrm{B}(k+a,\alpha)}{\mathrm{B}(a,\alpha-1)} \qquad\qquad \alpha = 2 + a/c$$

see Section 14.1 in *Networks: An Introduction*

degree distribution

exactly solvable in the limit [originally by Simon 1955]

$$p_k = \frac{\mathrm{B}(k+a,\alpha)}{\mathrm{B}(a,\alpha-1)} \qquad \alpha = 2 + a/c$$

recall that

$$\mathrm{B}(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$$

$\mathrm{B}(a,b) \sim a^{-b}$ (in the tail)

degree distribution

exactly solvable in the limit [originally by Simon 1955]

$$p_k = \frac{\mathrm{B}(k+a,\alpha)}{\mathrm{B}(a,\alpha-1)} \qquad \alpha = 2 + a/c$$

recall that

$$\mathrm{B}(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$$

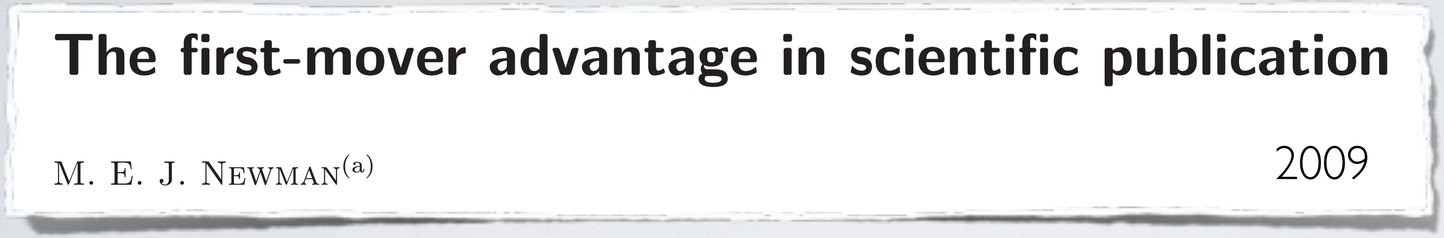
$$\mathrm{B}(a,b) \sim a^{-b}$$ (in the tail)

thus, distribution of citations

$$p_k \approx (k+a)^{-\alpha}$$

see Section 14.1 in *Networks: An Introduction*

degree distribution

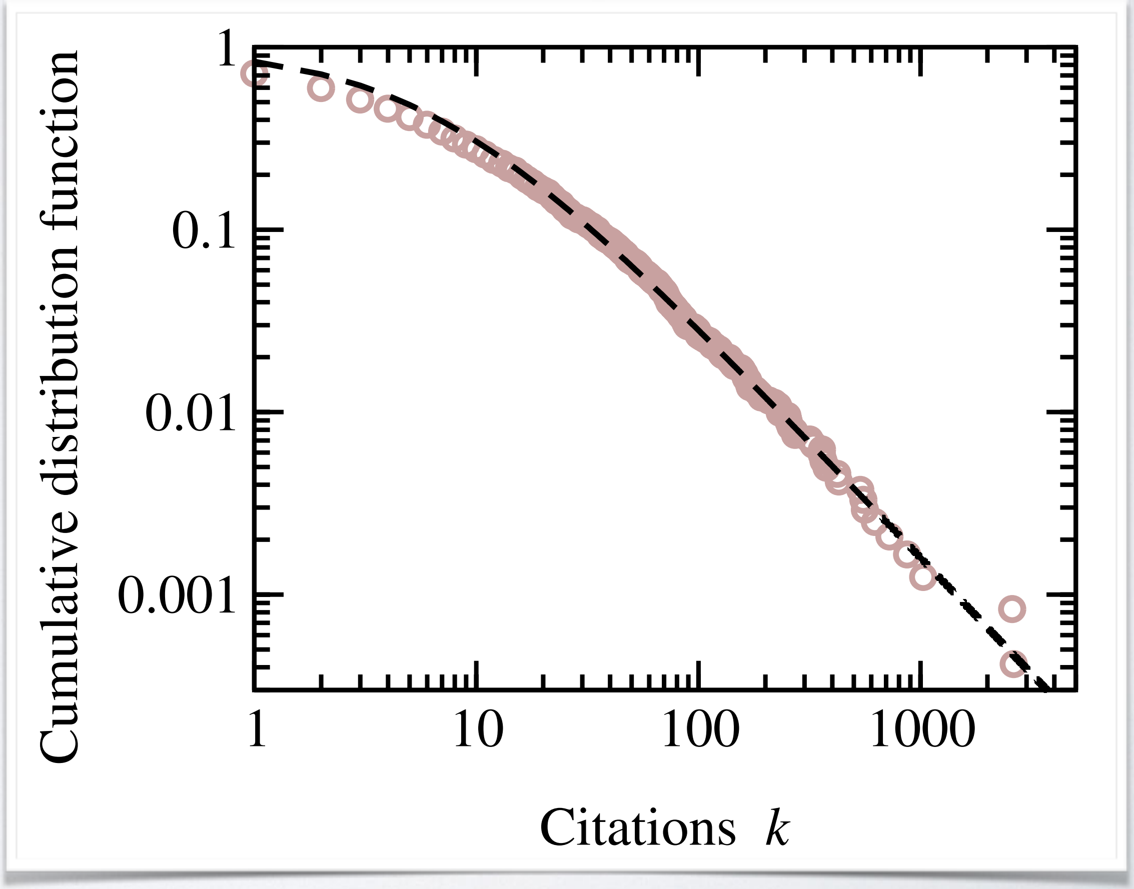


$$p_k \approx (k+a)^{-\alpha}$$

- 2407 network science papers
- from 1998-2008
- fitted parameters

$\alpha = 2.28$

$a = 6.38$



the first-mover effect

The first-mover advantage in scientific publication

M. E. J. NEWMAN[a] 2009

- let t_i denote *time* that paper i was published
- new papers only cite older papers
- thus, first-mover effect: $k_i \propto 1/t_i$

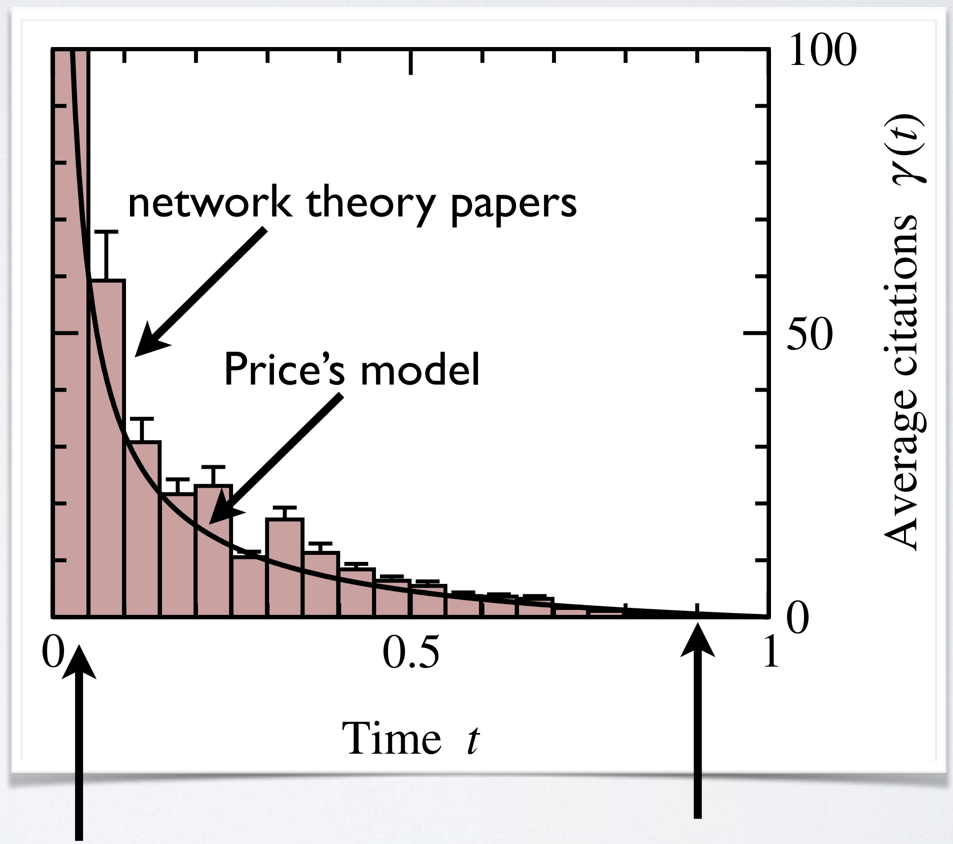
- Price's model fully specified by α and a
- idea:

1. estimate them from total citation distribution
2. derive predictions about citation counts vs. age of paper

the first-mover effect

average citations $\langle k \rangle$ vs. time of publication t

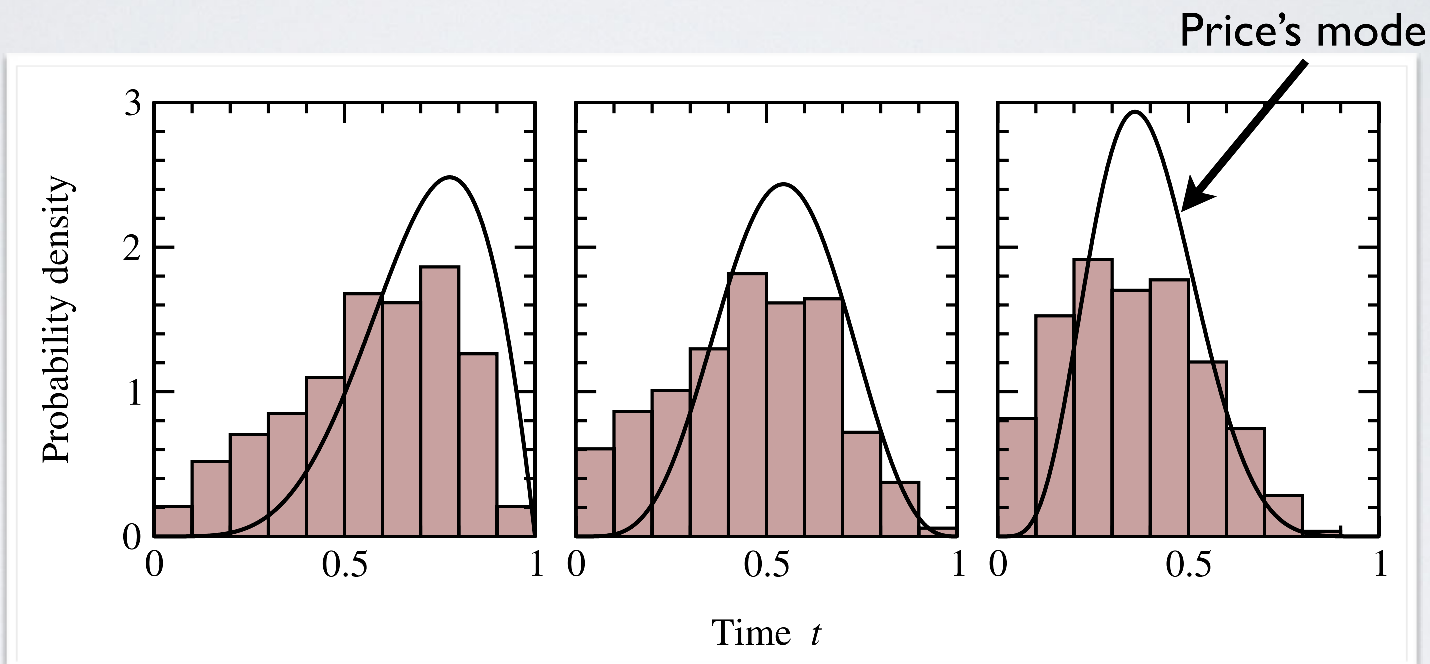
no free parameters



the first-mover effect

given k citations at time $t = 1$, probability of publication time t_i

no free parameters



$k \in [1, 2]$ $k \in [3, 5]$ $k \in [6, 10]$

checking the model

Citation Statistics from 110 Years of *Physical Review* 2004

Sidney Redner

110 years of data (July 1893 - June 2003)

3.1 millions citations

330,000 papers with at least one citation

key question: is attachment function $\propto k_i$?

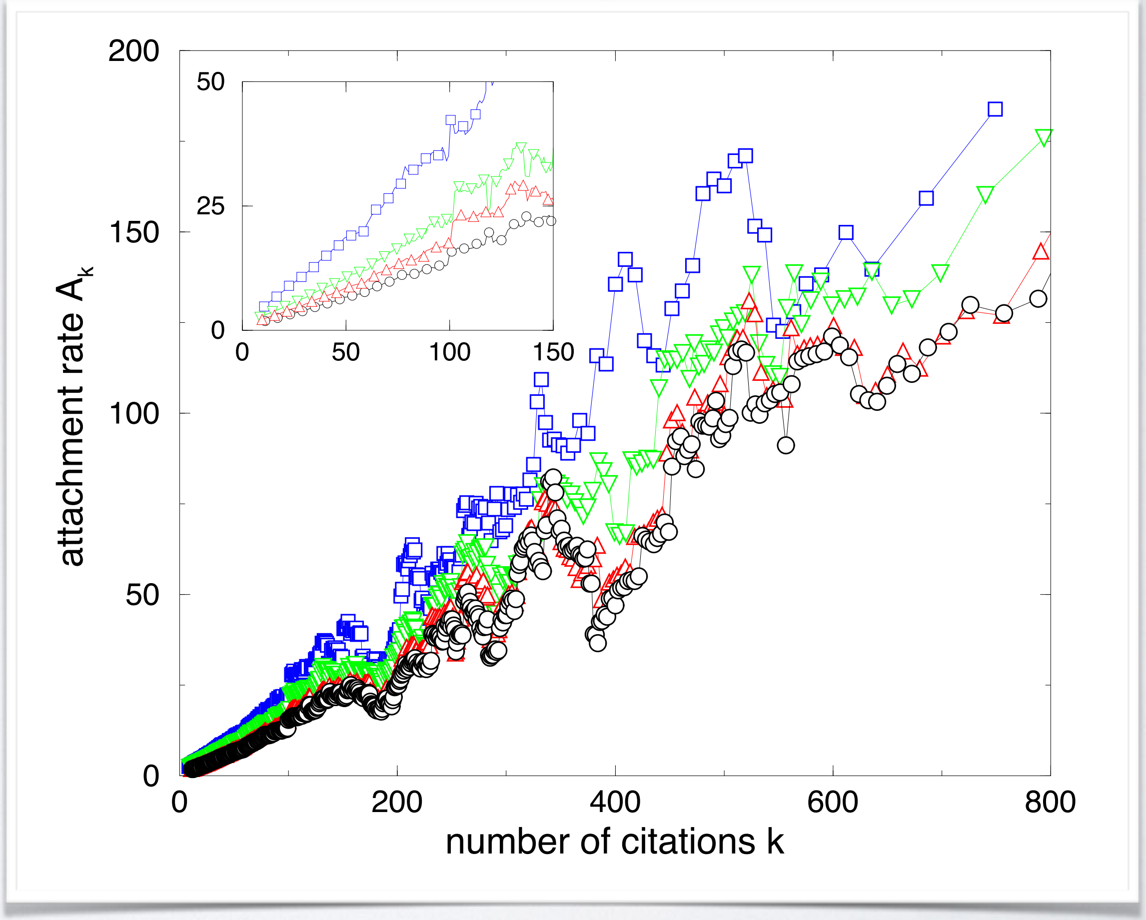
checking the model

key question: is attachment function $\propto k_i$?

pretty much.

caveat:

- ensemble only
  (*not* individual papers)



citation networks

networks of scientific publications

summary of features

- Price's model: *preferential + uniform attachment*
  - excellent model of citation networks
  - also good model of WWW
  - a variation (duplication-mutation) good for gene networks
- not a great model of many other networks
  - especially social and spatial networks
  - ignores constraints (cost of edges)
- many additional mathematical, empirical results
  - see Redner's, Newman's, Fortunato's work