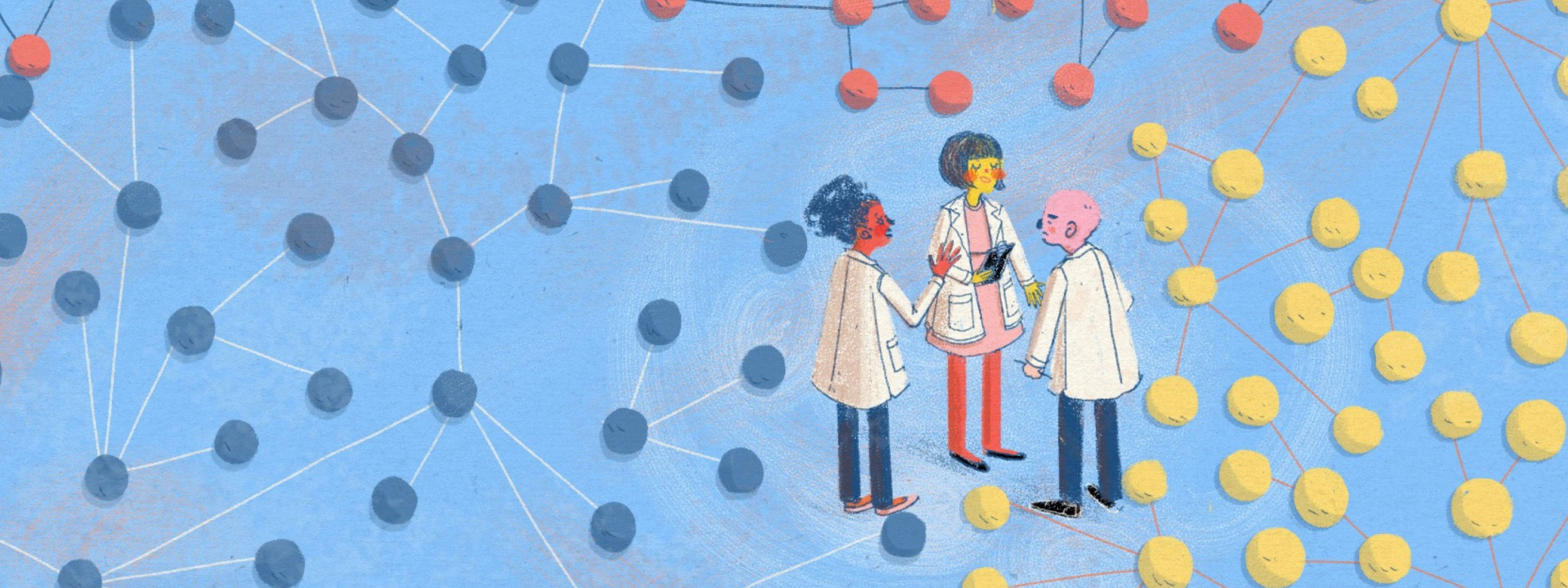


Scale-free networks are rare

joint work with Dr. Anna D. Broido

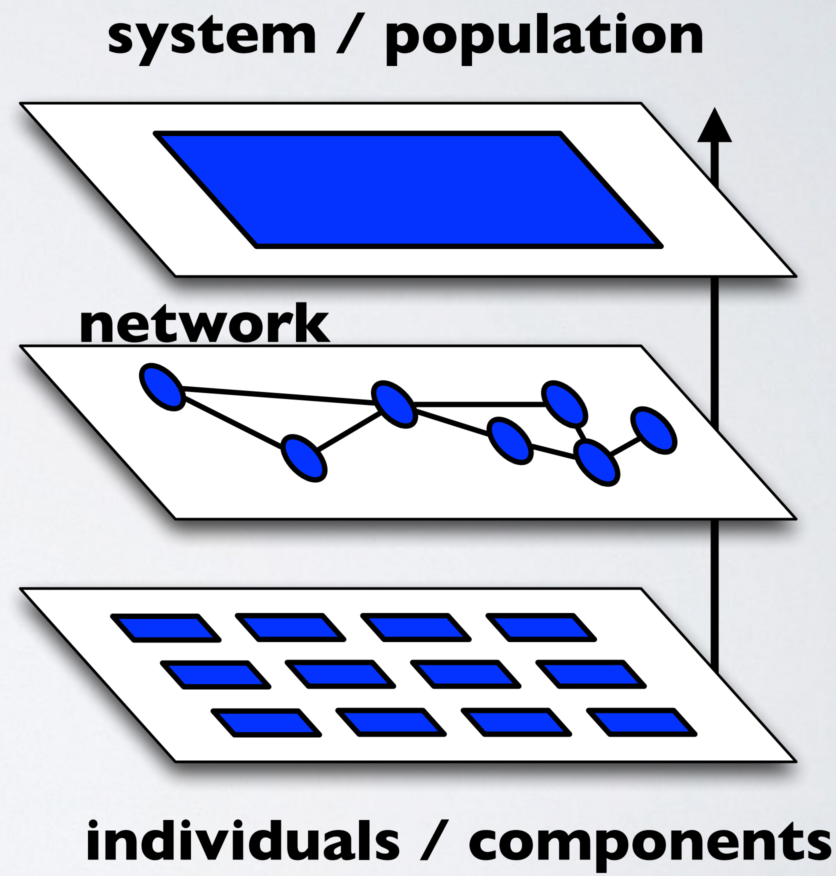
Aaron Clauset
@aaronclauset
Computer Science Dept. & BioFrontiers Institute
University of Colorado, Boulder
External Faculty, Santa Fe Institute

image by Meredith Miotke for Quanta Magazine

the *shape* of a network

network structure :

- shapes system dynamics & drives function
- offers clues about underlying mechanisms
- informs predictive model building



the scale-free hypothesis

ON A CLASS OF SKEW DISTRIBUTION FUNCTIONS

By HERBERT A. SIMON†

Carnegie Institute of Technology

1955

Networks of Scientific Papers

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

Derek J. de Solla Price

1965

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

SCIENCE VOL 286 15 OCTOBER 1999

1999

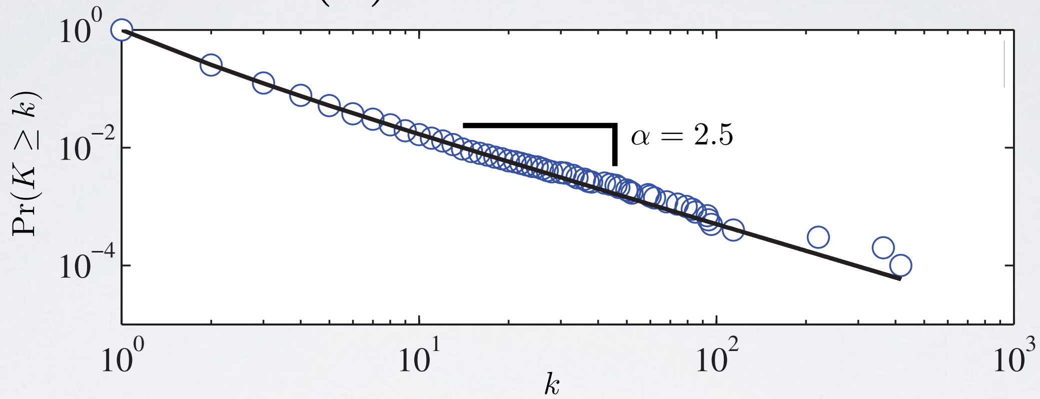
1955	Simon	cumulative advantage = scale-free distributions
1965	Price	cumulative advantage in citation networks
1999	Barabási & Albert	generalized to all networks

Simon, *Biometrika* **42**, 425-440 (1955)
Price, *Science* **149**, 510-515 (1965)
Barabási, Albert, *Science* **286**, 509-512 (1999)

the scale-free hypothesis

scale-free networks have a power-law degree distribution (in upper tail)

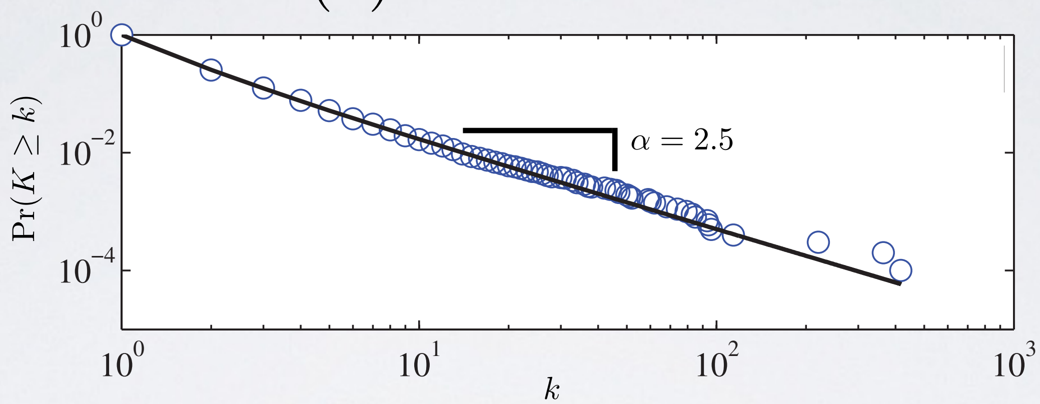
for example: $$\Pr(k) \propto k^{-\alpha} \qquad \alpha > 1$$



the scale-free hypothesis

scale-free networks have a power-law degree distribution (in upper tail)

for example: $$\Pr(k) \propto k^{-\alpha} \qquad \alpha > 1$$



scale-free networks represent a foundational idea in network science

- widely believed that scale-free networks are empirically universal
  (social, biological, informational, technological, economic, etc.)
- standard model of heterogeneous network structure (in all the textbooks)
- basis for many theoretical results (epidemics, influence, vulnerability, controllability, etc.)
- now, an enormous literature : Cited by 38886

the scale-free hypothesis

also empirically and theoretically controversial

the scale-free hypothesis

also empirically and theoretically controversial

Epidemic Spreading in Scale-Free Networks

Romualdo Pastor-Satorras[1] and Alessandro Vespignani[2]
[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain*
[2]*The Abdus Salam International Centre for Theoretical Physics (ICTP), P.O. Box 586, 34100 Trieste, Italy*
(Received 20 October 2000)

Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications

Lun Li, David Alderson, John C. Doyle, and Walter Willinger

Internet

Diameter of the World-Wide Web

Réka Albert, Hawoong Jeong, Albert-László Barabási
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

Classification of scale-free networks

Kwang-Il Goh*, Eulsik Oh*, Hawoong Jeong†, Byungnam Kahng*‡, and Doochul Kim*

SCALE-FREE NETWORKS IN BIOLOGY

Eivind Almaas, Alexei Vázquez and Albert-László Barabási
Center for Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556,USA

Scale-Rich Metabolic Networks

Reiko Tanaka*
Bio-Mimetic Control Research Center, RIKEN, Moriyama-ku, Nagoya 463-0003, Japan
(Received 1 October 2004; published 25 April 2005)

Revisiting "scale-free" networks

Evelyn Fox Keller*

PERSPECTIVE

Scale-Free Networks: A Decade and Beyond

Albert-László Barabási

Subnets of scale-free networks are not scale-free: Sampling properties of networks

Michael P. H. Stumpf†‡, Carsten Wiuf§, and Robert M. May¶

the scale-free hypothesis

why is their empirical status difficult to resolve?

- *comparing theory and data is complicated** 🤔

* in the sense of Mayo, there are many *auxiliary* hypotheses that surround the scale-free hypothesis

the scale-free hypothesis

why is their empirical status difficult to resolve?

- *comparing theory and data is complicated** 🤔
- statistically ***unsound methods*** for assessing $\Pr(k) \propto k^{-\alpha}$

* in the sense of Mayo, there are many *auxiliary* hypotheses that surround the scale-free hypothesis

the scale-free hypothesis

why is their empirical status difficult to resolve?

- *comparing theory and data is complicated* * 🤔
- statistically ***unsound methods*** for assessing $\Pr(k) \propto k^{-\alpha}$
- relatively ***little empirical data*** (esp. in 2000-2010)

* in the sense of Mayo, there are many *auxiliary* hypotheses that surround the scale-free hypothesis

the scale-free hypothesis

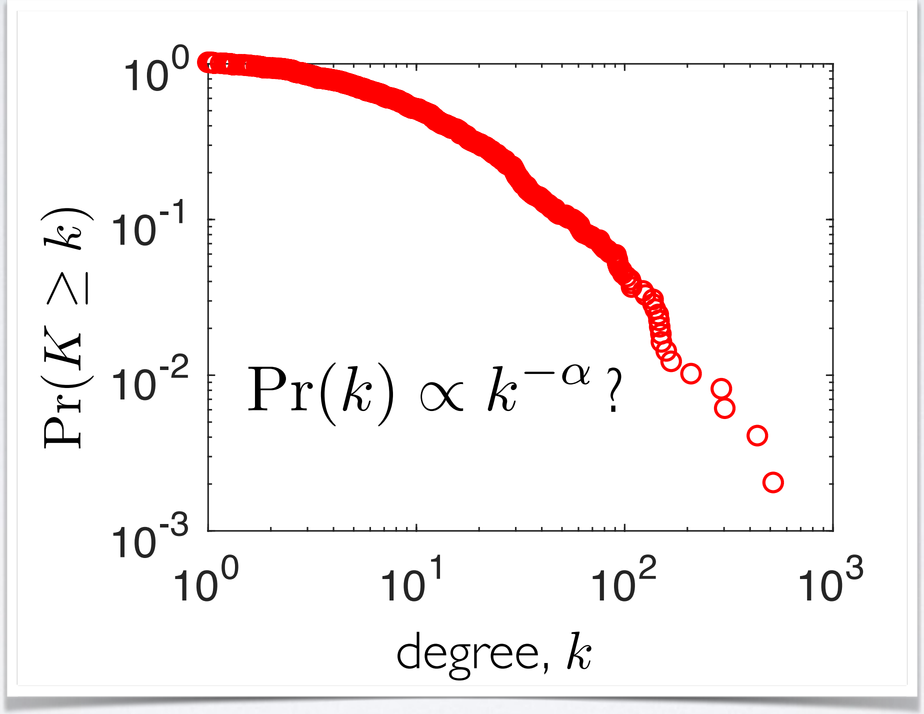
why is their empirical status difficult to resolve?

- *comparing theory and data is complicated*[*] 🤔
- statistically ***unsound methods*** for assessing $\Pr(k) \propto k^{-\alpha}$
- relatively ***little empirical data*** (esp. in 2000-2010)
- ***people mean different things*** when they say "scale free" :
  - just more heavy-tailed than exponential, $\Pr(k) \propto \mathrm{e}^{-\lambda k}$
  - not $\Pr(k) \propto k^{-\alpha}$ exactly, but more plausible than alternatives
  - plausibly $\Pr(k) \propto k^{-\alpha}$ for any α
  - plausibly $\Pr(k) \propto k^{-\alpha}$ but for $2 < \alpha < 3$ only
  - plausibly $\Pr(k) \propto k^{-\alpha}$ but only in upper tail
  - plausibly $\Pr(k) \propto k^{-\alpha}$ except for exponential cutoff
  - scale free is well defined mainly for simple graphs

* in the sense of Mayo, there are many *auxiliary* hypotheses that surround the scale-free hypothesis

the scale-free hypothesis

tl;dr : is this degree distribution "scale free"? 🤷‍♀️



resolving the controversy

we aim to construct a *severe test** of the scale-free hypothesis:

- nearly 1000 empirical network data sets
- sound statistical methods
- organize & assess multiple definitions of "scale free"
- evaluate robustness of results

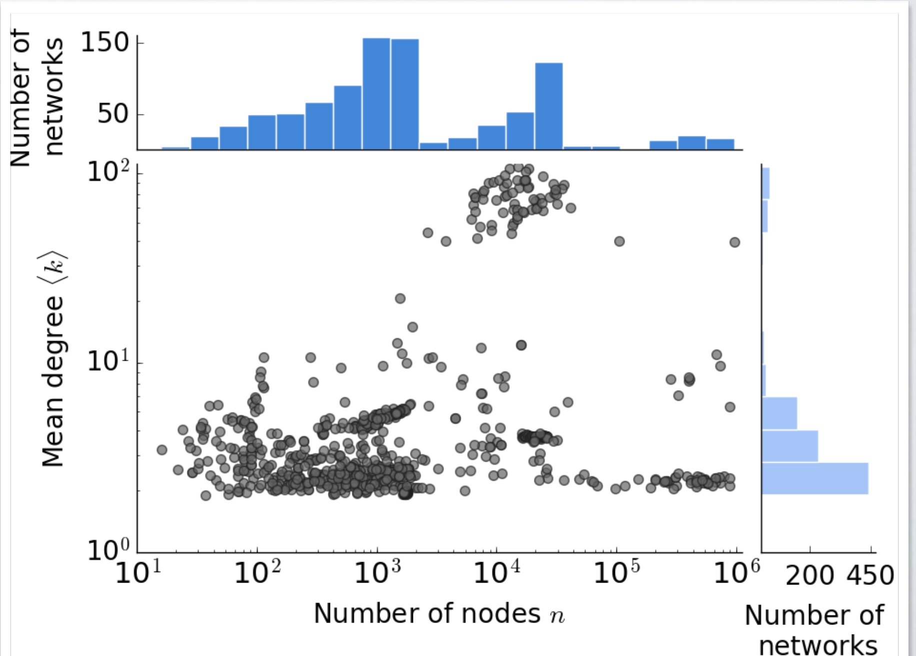
* in the sense of Mayo, *Error and the Growth of Experimental Knowledge* (1996)

empirical data & issues

a novel corpus of 928 network data sets

- drawn from Index of Complex Networks (all domains of science)

Domain	Num(Prop)	Multip.	Bip.	Multig.	Weigh.	Dir.	Simp.
Bio.	495 (0.53)	273	41	378	29	37	39
Info.	16 (0.02)	0	0	4	0	5	7
Social	147 (0.16)	7	0	6	8	0	129
Tech.	203 (0.22)	122	0	3	1	195	5
Trans.	67 (0.07)	48	0	65	3	2	0
Total	928 (1.00)	450	41	456	41	239	180



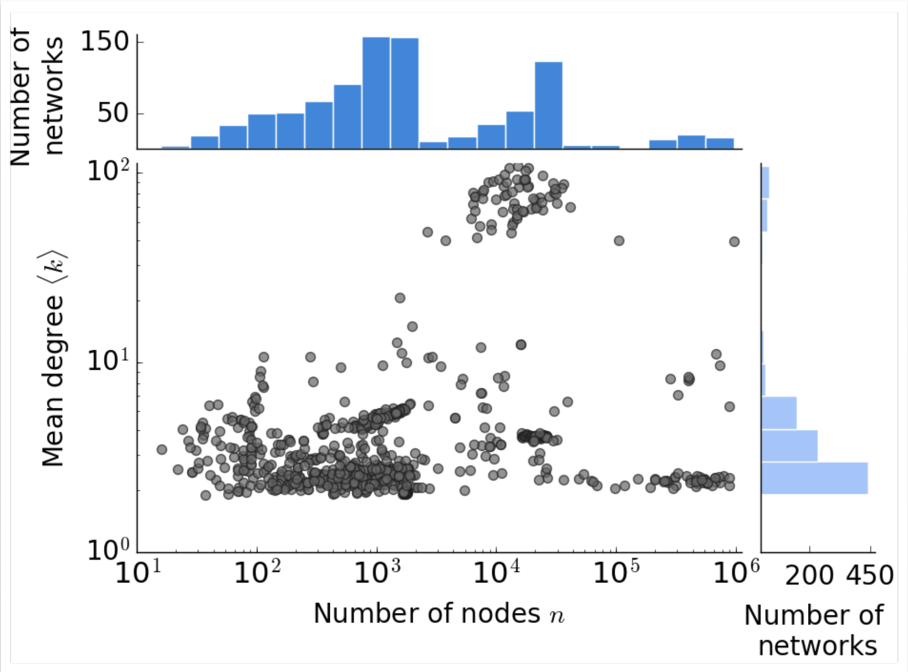
Index of Complex Networks : icon.colorado.edu

empirical data & issues

a novel corpus of 928 network data sets

- drawn from Index of Complex Networks (all domains of science)
- includes multiplex, bipartite, multi-edge, weighted, directed graphs
- **problem**: *scale-free network* not well-defined for non-simple graphs

Domain	Num(Prop)	Multip.	Bip.	Multig.	Weigh.	Dir.	Simp.
Bio.	495 (0.53)	273	41	378	29	37	39
Info.	16 (0.02)	0	0	4	0	5	7
Social	147 (0.16)	7	0	6	8	0	129
Tech.	203 (0.22)	122	0	3	1	195	5
Trans.	67 (0.07)	48	0	65	3	2	0
Total	928 (1.00)	450	41	456	41	239	180



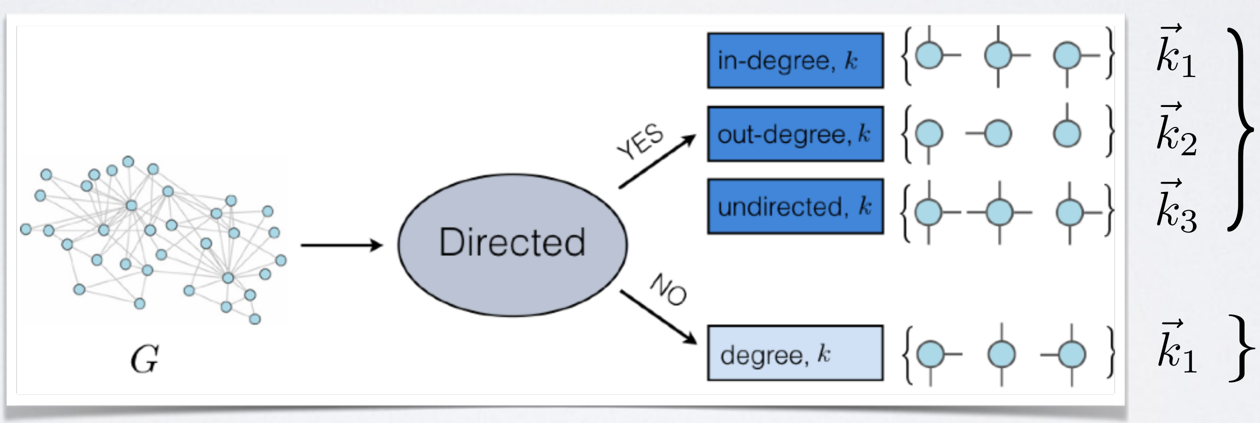
Index of Complex Networks : icon.colorado.edu

empirical data & issues

solution: convert† each network $G \to$ set of degree sequences $\{\vec{k}\}$

- exclude any $\vec{k}$ that is too sparse or too dense*
- assess each $\vec{k}$ for scale-free pattern $\Pr(k) \propto k^{-\alpha}$
- consensus among $\{\vec{k}\}$ determines conclusion for G

for example:



- 928 data sets $\to$ 3662 degree sequences

† see slide 48
* we include only simple graphs with $2 < \langle k \rangle < \sqrt{n}$ which removes 14,786 of 18,448 graphs (80.4%) for being too sparse or too dense to be plausibly scale free. More than 90% of these excluded graphs were generated by only 3 input graphs. Fully 94% of the input graphs yielded no exclusions.

resolving the controversy

we aim to construct a *severe test** of the scale-free hypothesis:

✅ nearly 1000 empirical network data sets

- sound statistical methods
- organize & assess multiple definitions of "scale free"
- evaluate robustness of results

* in the sense of Mayo, Error and the Growth of Experimental Knowledge (1996)

statistical methods

for each $\vec{k}$:

- estimate power-law model M_{PL} with $\hat{\theta} = (\hat{k}_{\mathrm{min}}, \hat{\alpha})$
- compute n_{tail} : #nodes in scale-free region
- compute p-value for $\Pr(k \mid \hat{\theta})$ via Monte Carlo

- store $(\hat{k}_{\mathrm{min}}, \hat{\alpha}, n_{\mathrm{tail}}, p$

standard methods for testing for power-law patterns

SIAM REVIEW
Vol. 51, No. 4, pp. 661–703

Power-Law Distributions in Empirical Data*

Clauset, Shalizi, Newman, *SIAM Review* **51**, 661 (2009)

statistical methods

for each $\vec{k}$:

- estimate power-law model $M_{\rm PL}$ with $\hat{\theta} = (\hat{k}_{\rm min}, \hat{\alpha})$
- compute $n_{\rm tail}$: #nodes in scale-free region
- compute p-value for $\Pr(k \,|\, \hat{\theta})$ via Monte Carlo
- estimate 4 alternative models $M_{\rm Alt}{}^{*}$
- Vuong likelihood ratio test $\mathcal{V}$ on 4 alternatives
  (exponential, log-normal, power law with exp. cutoff, stretched exponential)
- store $(\hat{k}_{\rm min}, \hat{\alpha}, n_{\rm tail}, p, \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4)$

standard methods for testing for power-law patterns

SIAM Review
Vol. 51, No. 4, pp. 661–703

Power-Law Distributions in Empirical Data*

* on degrees $k \geq \hat{k}_{\rm min}$, for technical reasons

Vuong, *Econometrica* **57**, 307 (1989)

Clauset, Shalizi, Newman, *SIAM Review* **51**, 661 (2009)

resolving the controversy

we aim to construct a *severe test** of the scale-free hypothesis:

✅ nearly 1000 empirical network data sets

✅ sound statistical methods

- organize & assess multiple definitions of "scale free"
- evaluate robustness of results

* in the sense of Mayo, Error and the Growth of Experimental Knowledge (1996)

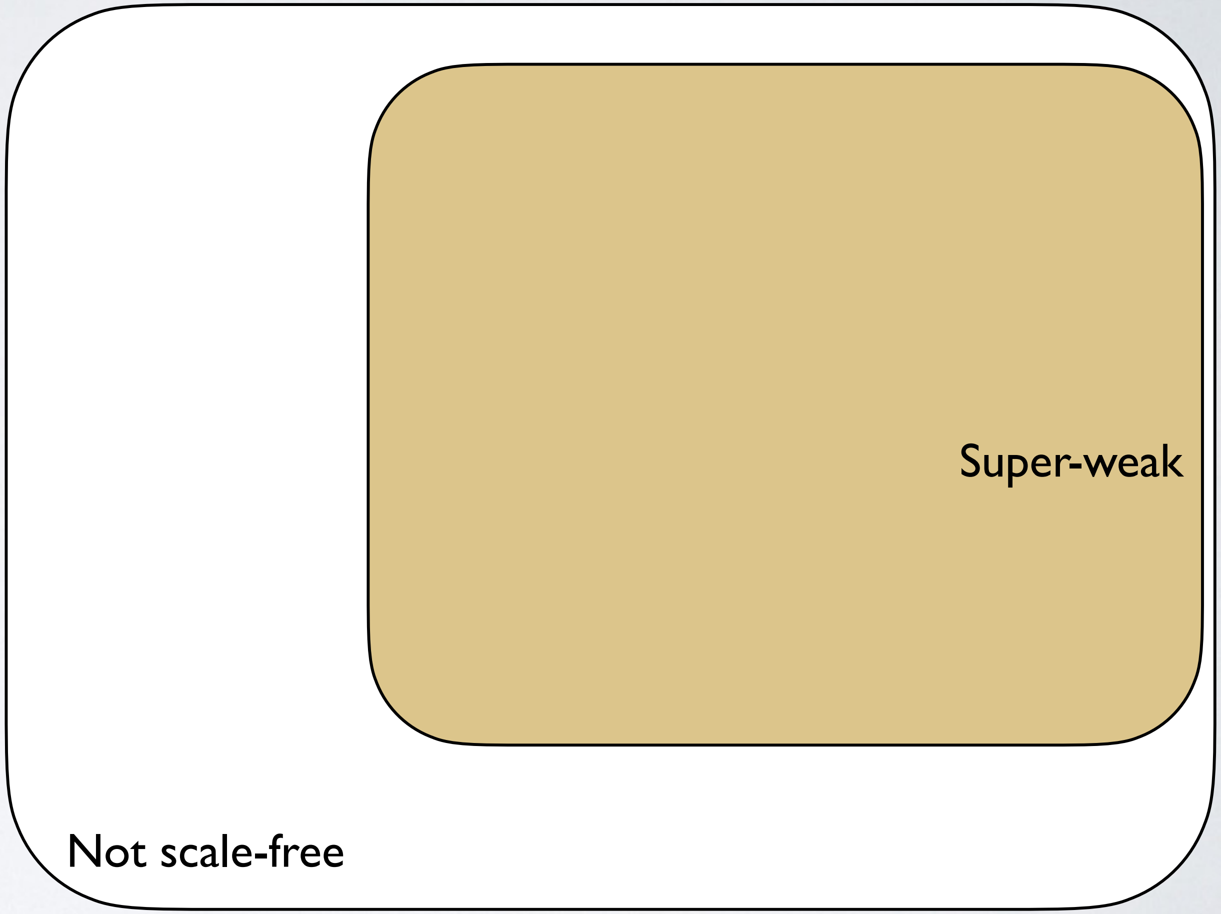
definitions of scale free

two main flavors:

- scale free (power law) is **better model** of data than alternatives
- scale free (power law) is **a good model** of the data

definitions of scale free

Super-weak : "closer to scale free"
no M_{Alt} favored over M_{PL}
for $\geq 50\%$ of $\{\vec{k}\}$

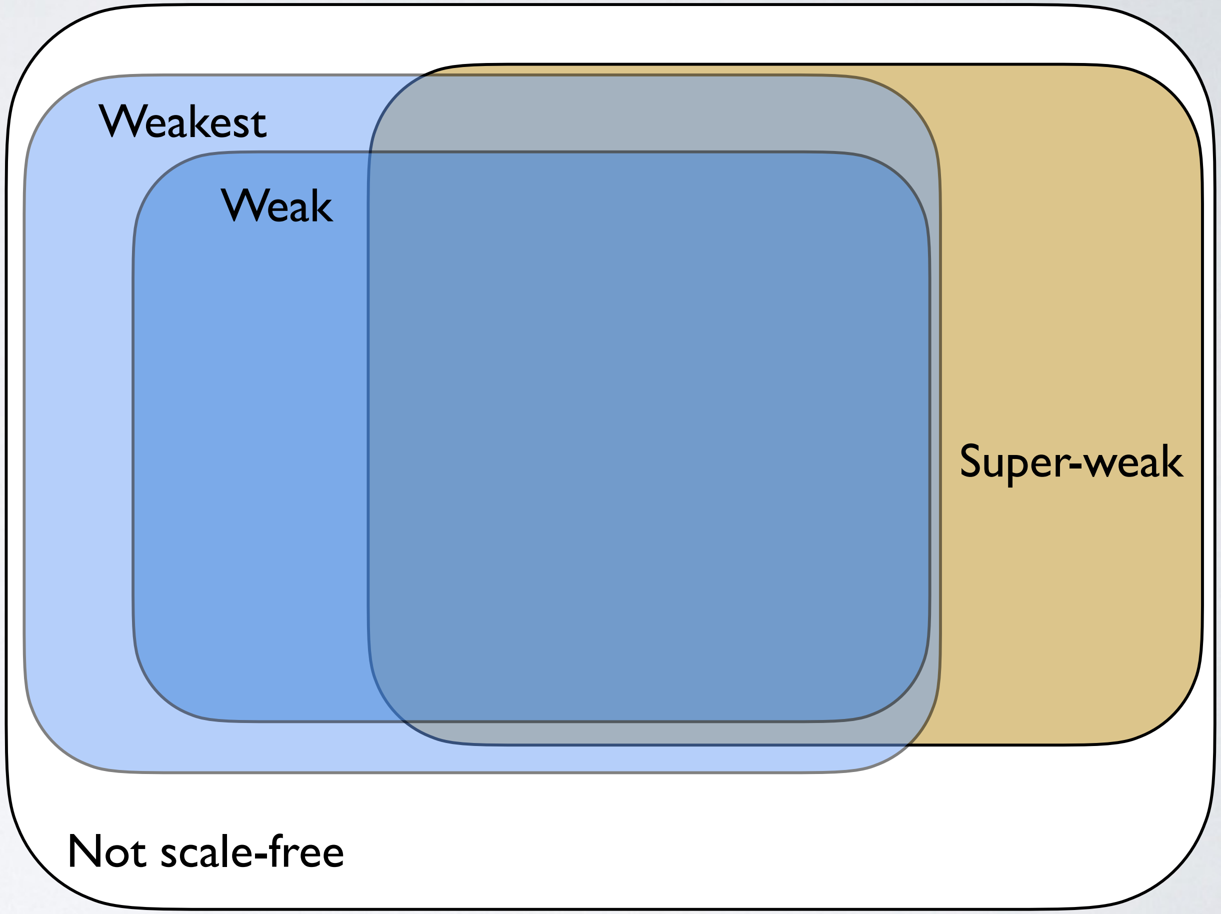


definitions of scale free

Super-weak : "closer to scale free"
no M_{Alt} favored over M_{PL}
for $\geq 50\%$ of $\{\vec{k}\}$

Weakest : "maybe scale free"
M_{PL} gives $p > 0.1$ for $\geq 50\%$ of $\{\vec{k}\}$

Weak : "maybe scale free, with power"
Weakest + $n_{\text{tail}} \geq 50$



definitions of scale free

Super-weak : "closer to scale free"
no $M_{\rm Alt}$ favored over $M_{\rm PL}$
for $\geq 50\%$ of $\{\vec{k}\}$

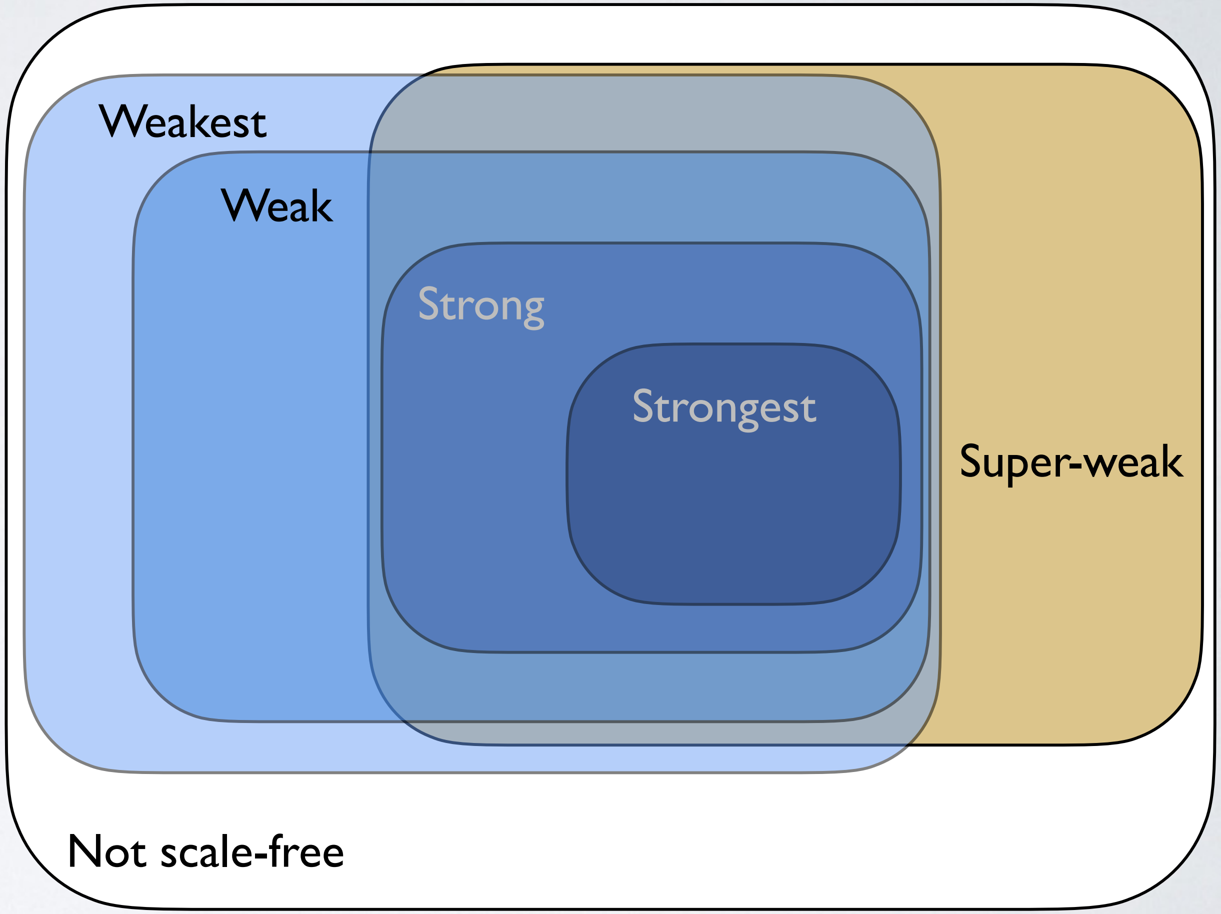
Weakest : "maybe scale free"
$M_{\rm PL}$ gives $p > 0.1$ for $\geq 50\%$ of $\{\vec{k}\}$

Weak : "maybe scale free, with power"
Weakest + $n_{\rm tail} \geq 50$

Strong : "pretty scale free"
Weak + *Super-weak* + $2 < \hat{\alpha} < 3$

Strongest : "really scale free"
Strong for $\geq 90\%$ of $\{\vec{k}\}$ and no $M_{\rm Alt}$
favored over $M_{\rm PL}$ for $\geq 95\%$ of $\{\vec{k}\}$

Not scale-free :
neither *Super-weak* nor *Weakest*



resolving the controversy

we aim to construct a *severe test** of the scale-free hypothesis:

- ✅ nearly 1000 empirical network data sets
- ✅ sound statistical methods
- ✅ organize & assess multiple definitions of "scale free"
- evaluate robustness of results

* in the sense of Mayo, Error and the Growth of Experimental Knowledge (1996)

resolving the controversy

we aim to construct a *severe test** of the scale-free hypothesis:

- ✅ nearly 1000 empirical network data sets
- ✅ sound statistical methods
- ✅ organize & assess multiple definitions of "scale free"
- ✅ evaluate robustness of results: we consider
  - ⭐ consistency across different domains
  - ⭐ lower threshold : **any** $\vec{k}$ is plausibly scale free (maximally permissive)
  - simple graphs only
  - moment ratio divergence $\langle k^2 \rangle / \langle k \rangle^2$ vs. n (model free)
  - accuracy on 4 types of synthetic networks with known structure

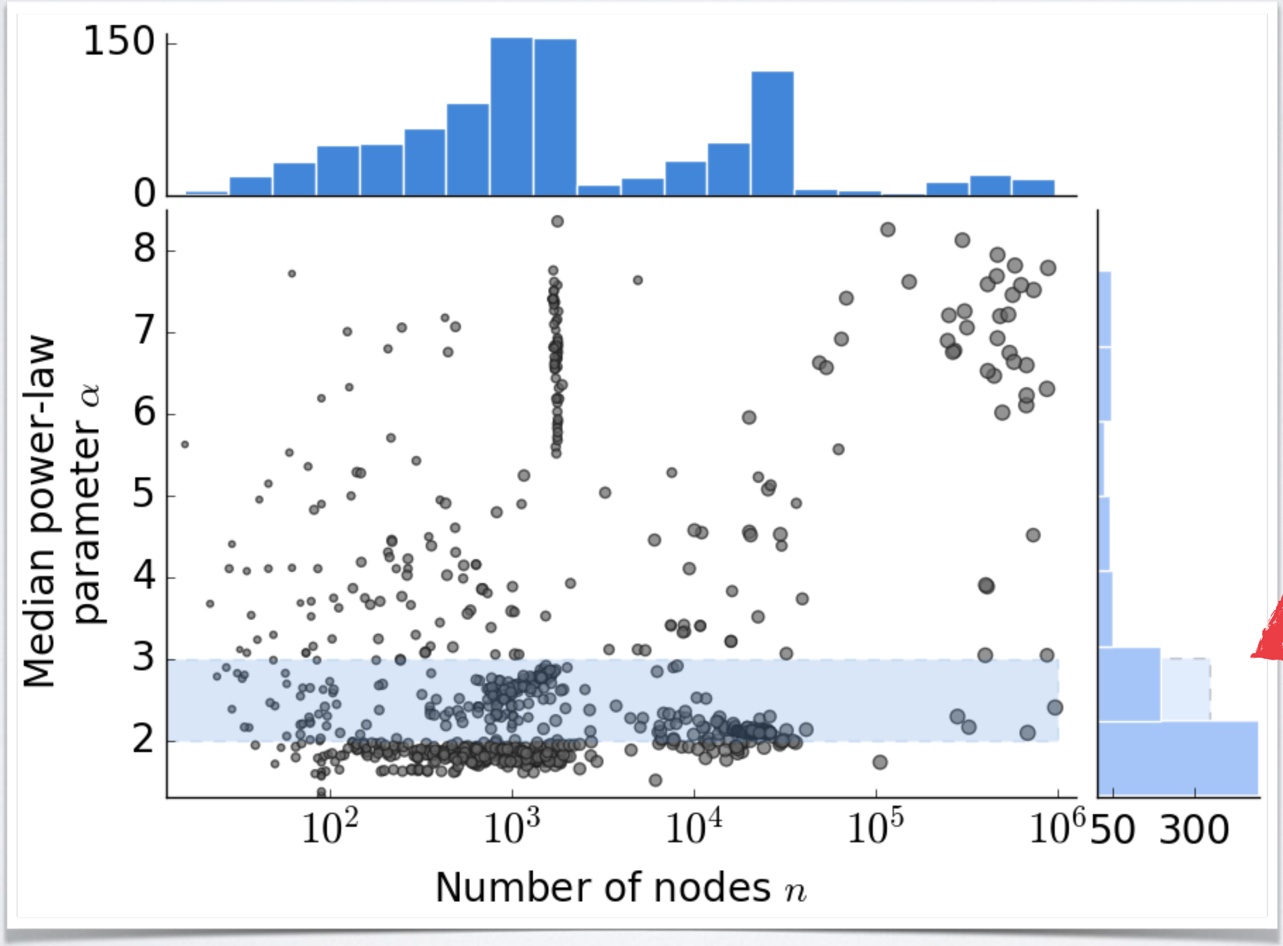
* in the sense of Mayo, Error and the Growth of Experimental Knowledge (1996)

results

- model parameters

results

- model parameters



Out of 928 data sets

results

- model parameters
- comparison to alternatives

results

- model parameters
- comparison to alternatives

Alternative	$p(x) \propto f(x)$	M_{PL}	Inconclusive	M_{Alt}
			Test Outcome	
Exponential	$e^{-\lambda x}$	**33%**	26%	41%
Log-Normal	$\frac{1}{x} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$	**12%**	40%	48%
Weibull	$e^{-\left(\frac{x}{b}\right)^a}$	**33%**	20%	47%
Power law with Cutoff	$x^{-\alpha} e^{-\lambda x}$	—	44%	56%

Proportion of 928 data sets

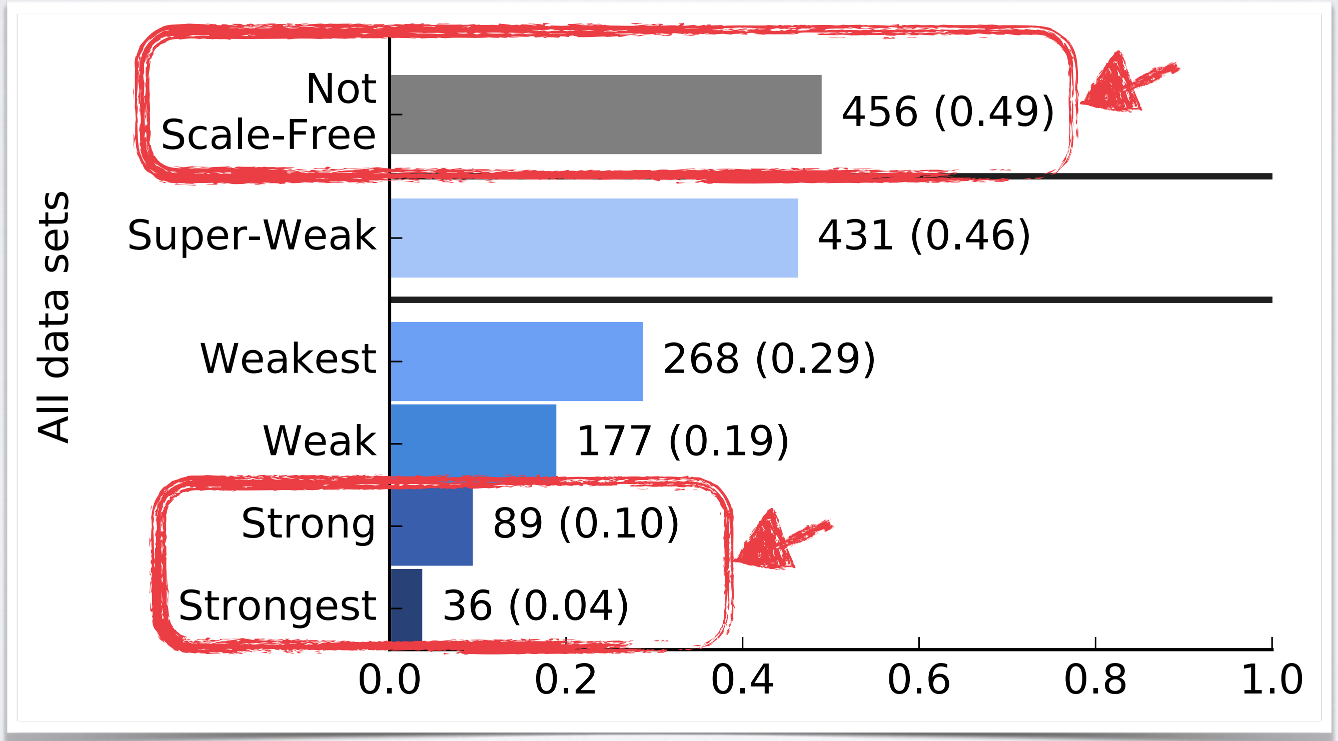
each row gives the expected fractions of outcomes for a randomly chosen data set

results

- model parameters
- comparison to alternatives
- distribution over levels of evidence

results

- model parameters
- comparison to alternatives
- distribution over levels of evidence

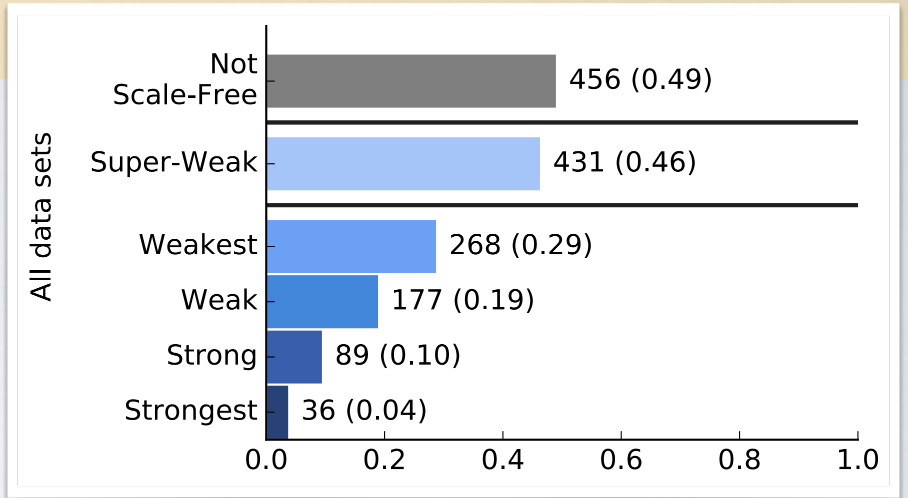


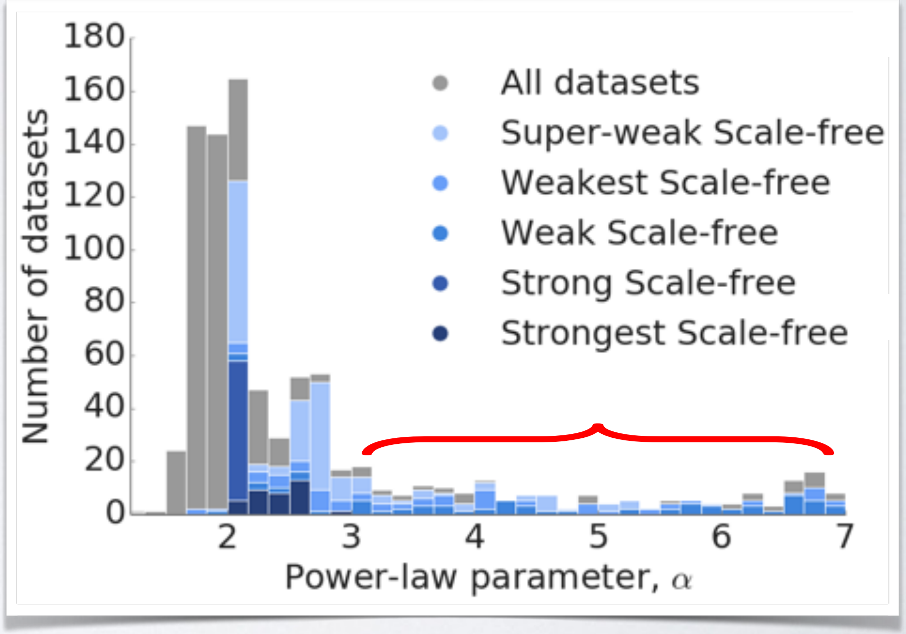
Proportion of 928 data sets

results : discussion

genuinely scale-free networks seem *rare*

- only 4% meet **Strongest** criteria
- only half meet **Super-Weak** criteria
- broad distribution of α
- *log-normal* as good/better for 88%
- even *exponential* better for 41%
- scale-free networks not universal





but, are these results robust? let's check.

robustness of results

- consistency across different domains

Biological

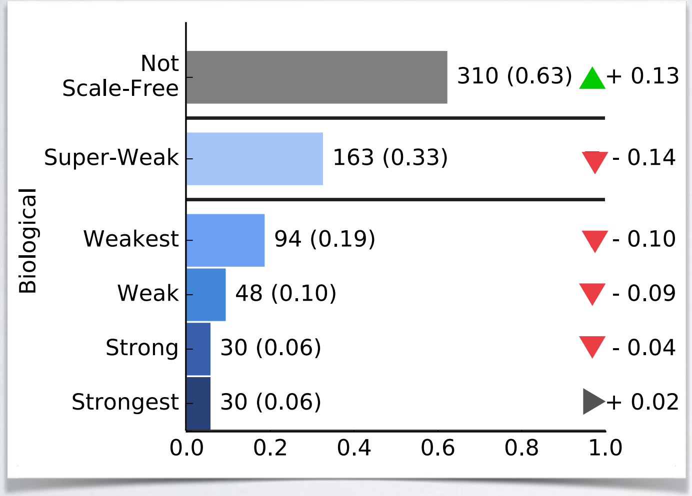
Social

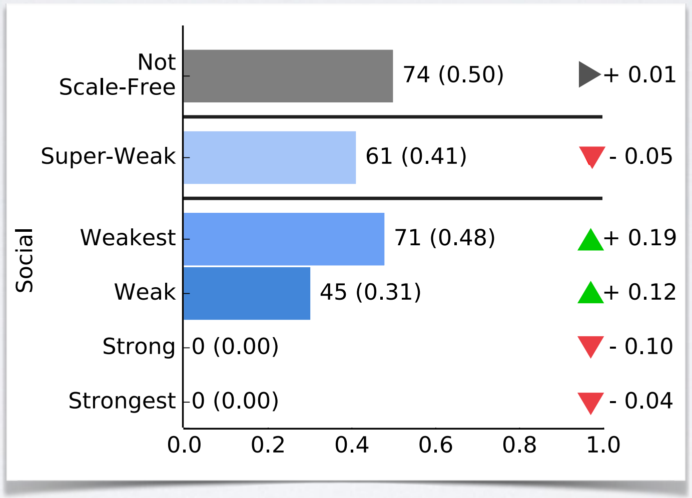
Technological

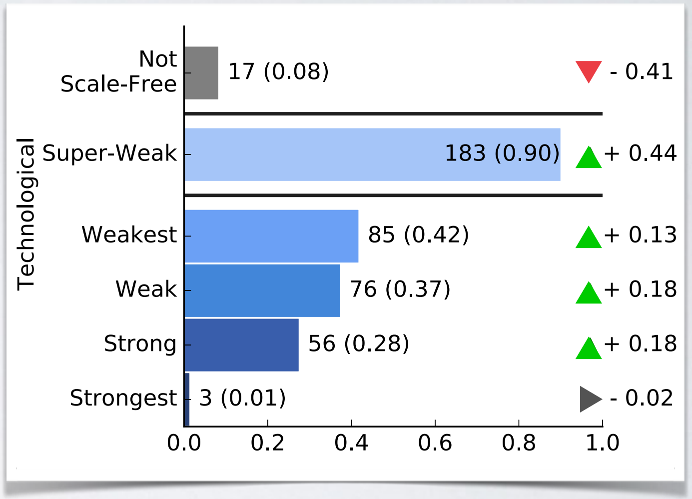
Biological (metabolic, protein, food web, connectomes, etc.)   Social (online, offline, etc.)   Technological (circuits, water pipes, auton. systems, p2p, etc.)

robustness of results

- consistency across different domains







Biological Social Technological

- rarity of scale-free networks is **consistent across domains**
- *biological networks* : less evidence overall
- *social networks* : mainly weak evidence
- *tech. networks* : marginally more evidence

Biological (metabolic, protein, food web, connectomes, etc.) Social (online, offline, etc.) Technological (circuits, water pipes, auton. systems, p2p, etc.)

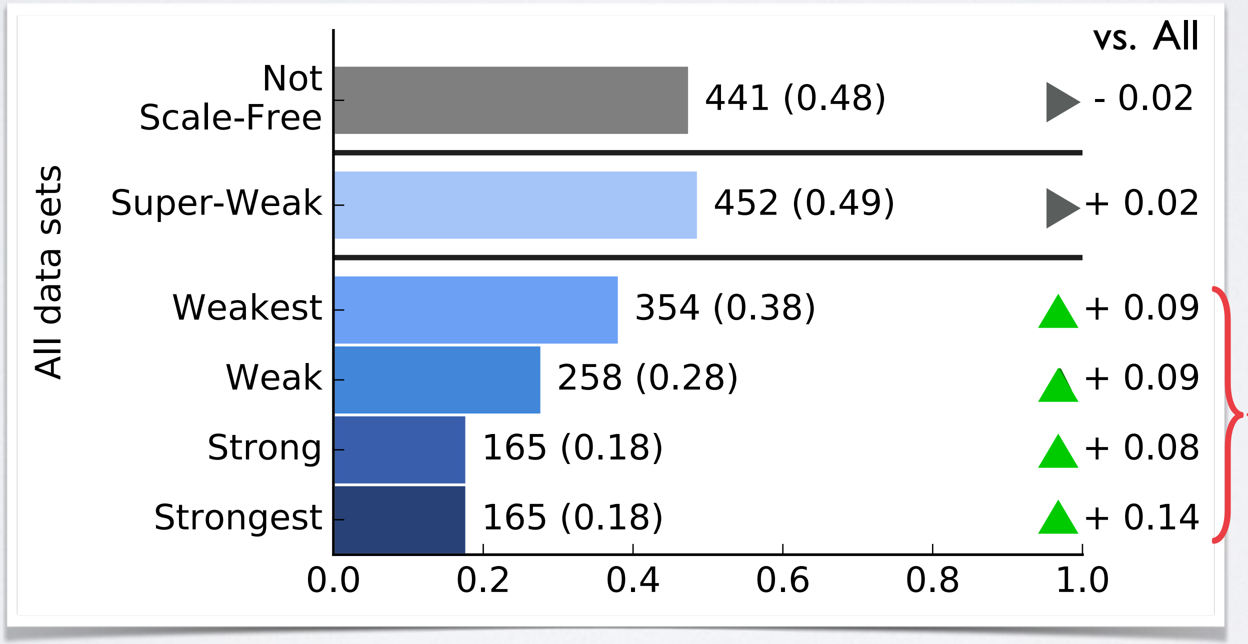
robustness of results

- consistency across different domains
- lower threshold : **any** $\vec{k}$ is plausibly scale free (**maximally permissive**)

* thanks to Albert-László Barabási for this idea

robustness of results

- consistency across different domains
- lower threshold : **any** $\vec{k}$ is plausibly scale free (**maximally permissive**)



all categories go up, *but not by much*

* thanks to Albert-László Barabási for this idea

conclusions

scale-free networks represent a foundational idea in network science

- widely believed that scale-free networks are empirically universal (social, biological, informational, technological, economic, etc.)
- standard model of heterogeneous network structure (in all the textbooks)
- basis for many theoretical results (epidemics, influence, vulnerability, controllability, etc.)
- enormous literature based around it

genuinely scale-free networks are empirically rare ✨

conclusions

scale-free networks represent a foundational idea in network science

- widely believed that scale-free networks are empirically universal (social, biological, informational, technological, economic, etc.)
- standard model of heterogeneous network structure (in all the textbooks)
- basis for many theoretical results (epidemics, influence, vulnerability, controllability, etc.)
- enormous literature based around it

genuinely scale-free networks are empirically rare ✨

- a highly robust result (simple / non-simple + across domains + lowest-possible bar)
- theoretical claims based on scale-free structure likely need revisiting or have limited real-world utility

conclusions & future directions

genuinely scale-free networks are empirically rare ✨

- ✅ evidence varies by domain in interesting ways → clues for future modeling
  (social networks weakly scale free at best | biological / technological networks mainly in areas with known mechanistic models, like protein interactions or auton. systems)
- ✅ log-normal distributions as good/better model → what makes log-normals?
- ✅ mechanisms that *create* scale-free networks (e.g., preferential attachment) may play limited role in real-world complex systems → need new mech.
- ✅ strong theoretical results in epidemiology (e.g., loss of epi. threshold) may not apply as broadly as believed
- ✅ real-world networks appear more structurally diverse than believed
  → what have we missed?
- ✅ comparing theories and data is complicated…

Dr. Anna D Broido
(now at Apple)

nature
COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-08746-5 OPEN

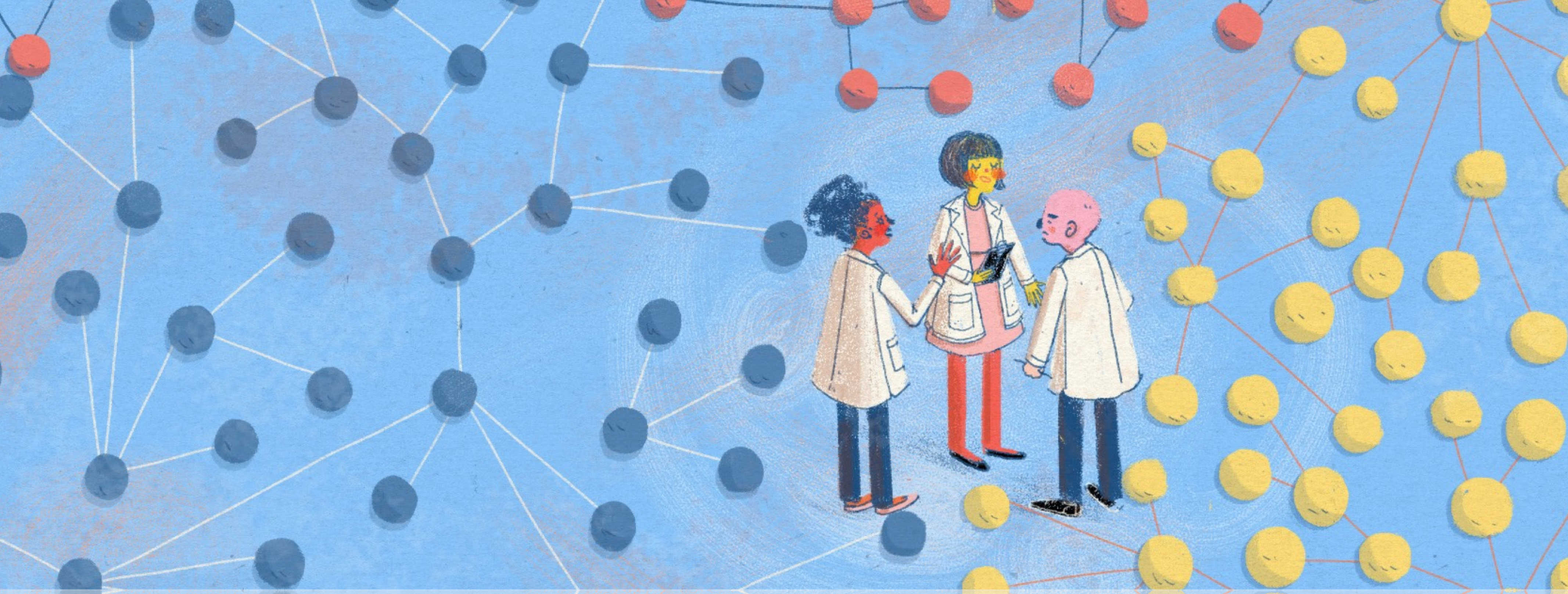
Scale-free networks are rare

Anna D. Broido[1] & Aaron Clauset[2,3,4]

acknowledgements: Eric Kightley, Johan Ugander, Dan Larremore, Duncan Watts, Mark Newman, Cris Moore, Cosma Shalizi, Marc Barthelemy, Petter Holme, Michael Stumpf, Albert-László Barabási, and Alessandro Vespignani

funding and support: NSF & NIH National Institutes of Health

data & code: https://github.com/adbroido/SFAnalysis



fin

image by Meredith Miotke for Quanta Magazine