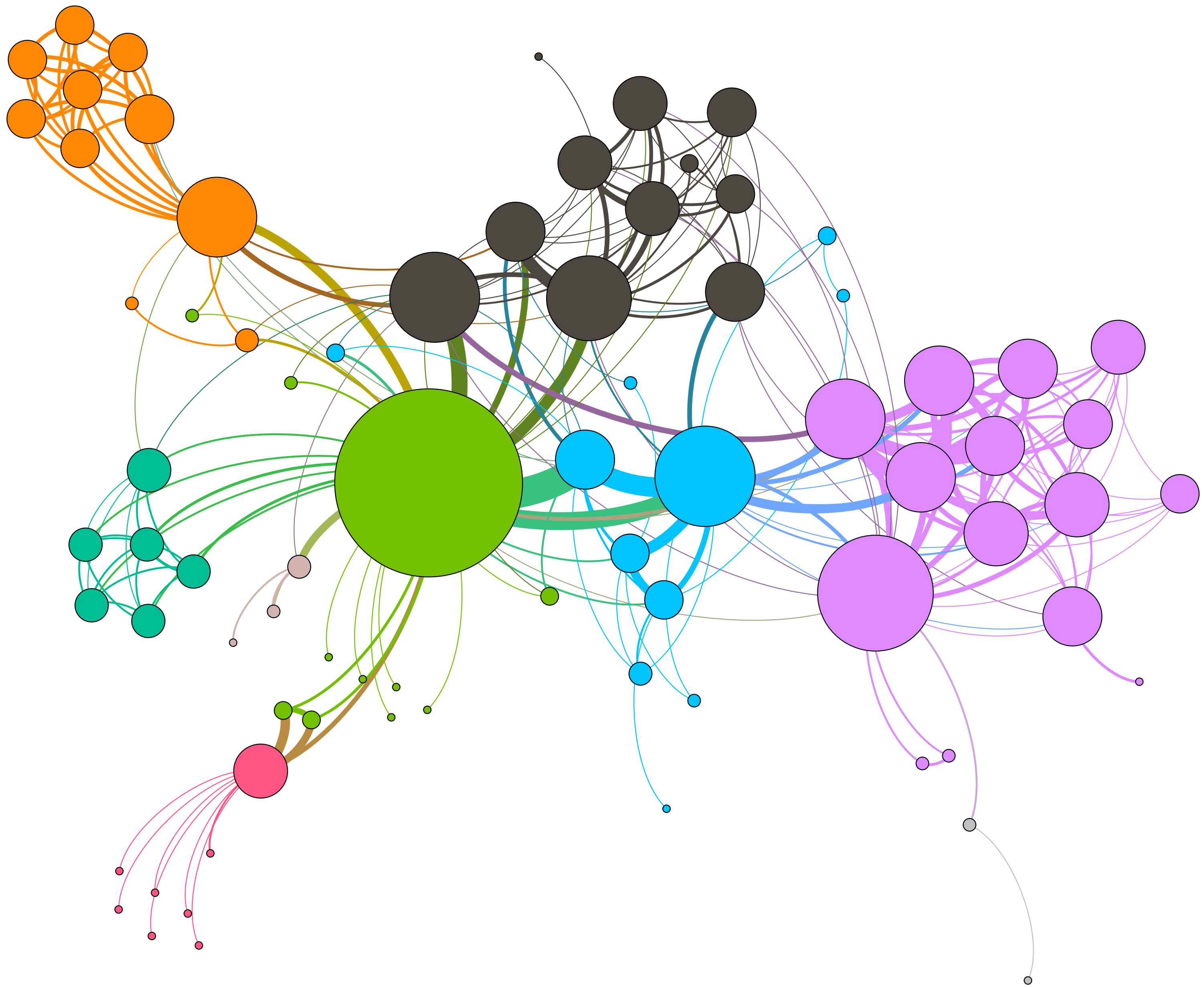
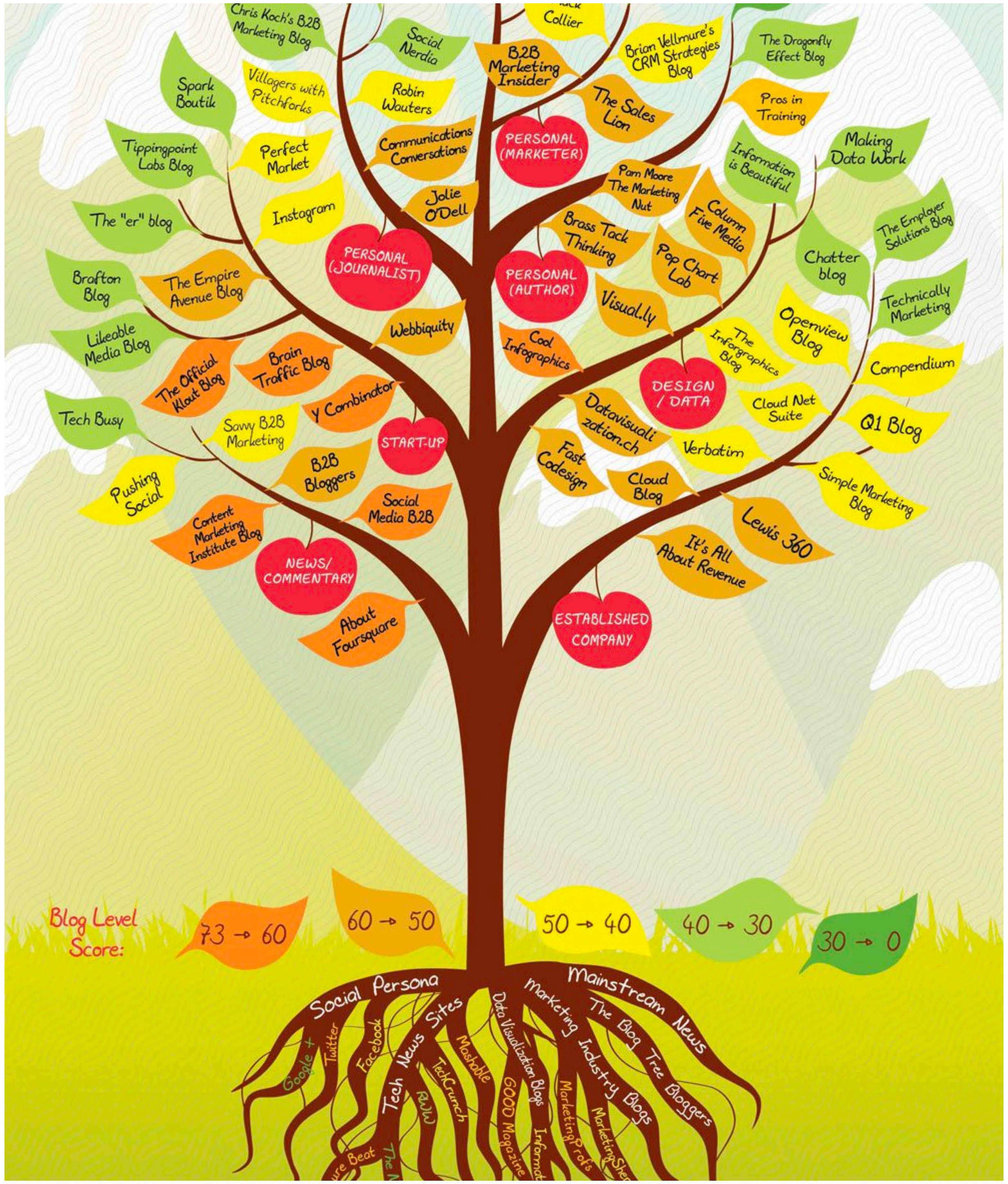


An Introduction to Biological Networks

Roberto Álvarez
UAQ
roberto.carlos.alvarez.martinez@gmail.com

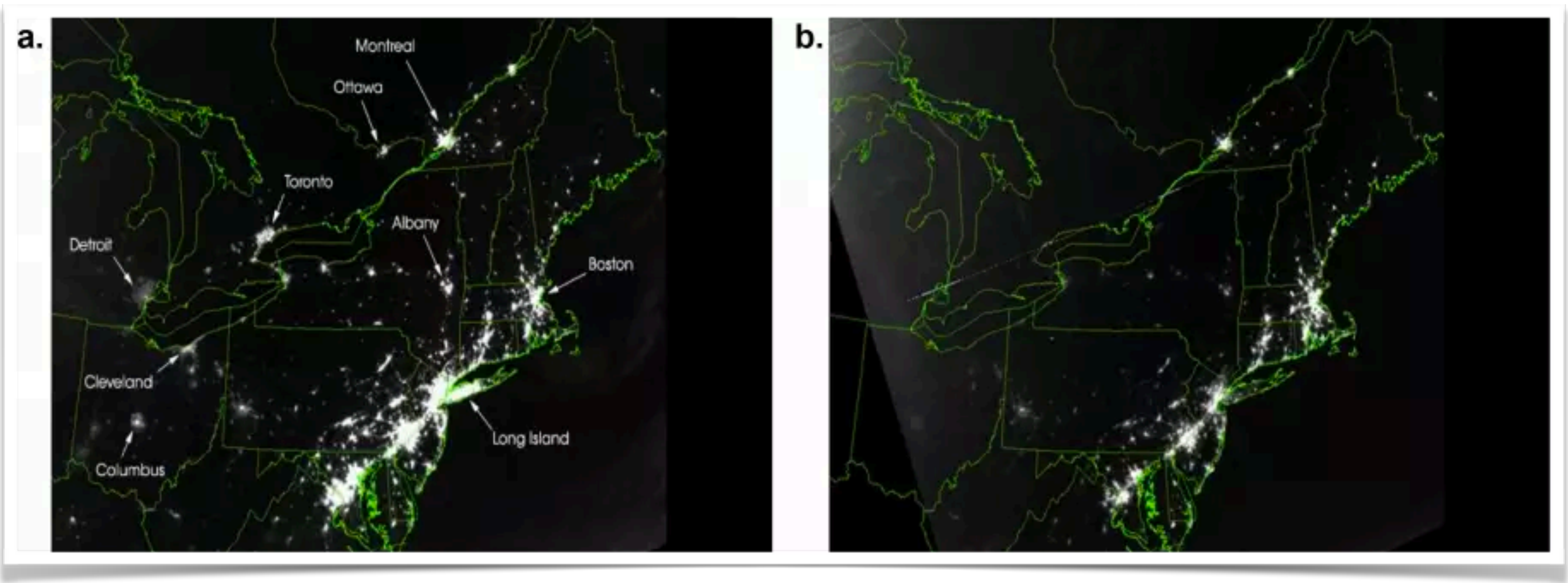
Introduction

- Complexity
- Vulnerability Due to Interconnectivity
- Networks at the Heart of Complex Systems
- Two Forces that Helped Network Science
- The Characteristics of Network Science
- Societal Impact
- Scientific Impact



Why do we need to understand networks?

2003 North American Black Out



- Satellite image on Northeast United States on August 13th, 2003,at 9:29pm (EDT), 20 hours *before* the 2003 blackout.
- The same as above, but 5 hours *after* the blackout.

Complex

[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]

1. composed of many interconnected parts; compound; composite: a complex highway system

2. characterized by a very complicated or involved arrangement of parts, units, etc.: complex machinery

3. so complicated or intricate as to be hard to understand or deal with: a complex problem

Source: *Dictionary.com*

“I think the next century will be the century of complexity”

– Stephen Hawking,2000

Complexity

- Complexity is a scientific theory that propose that some systems behave in such way that it is not understandable by standard analysis.

- Emergent phenomena are best (only?) understood through the (nonlinear) interactions of (many) internal components of system.

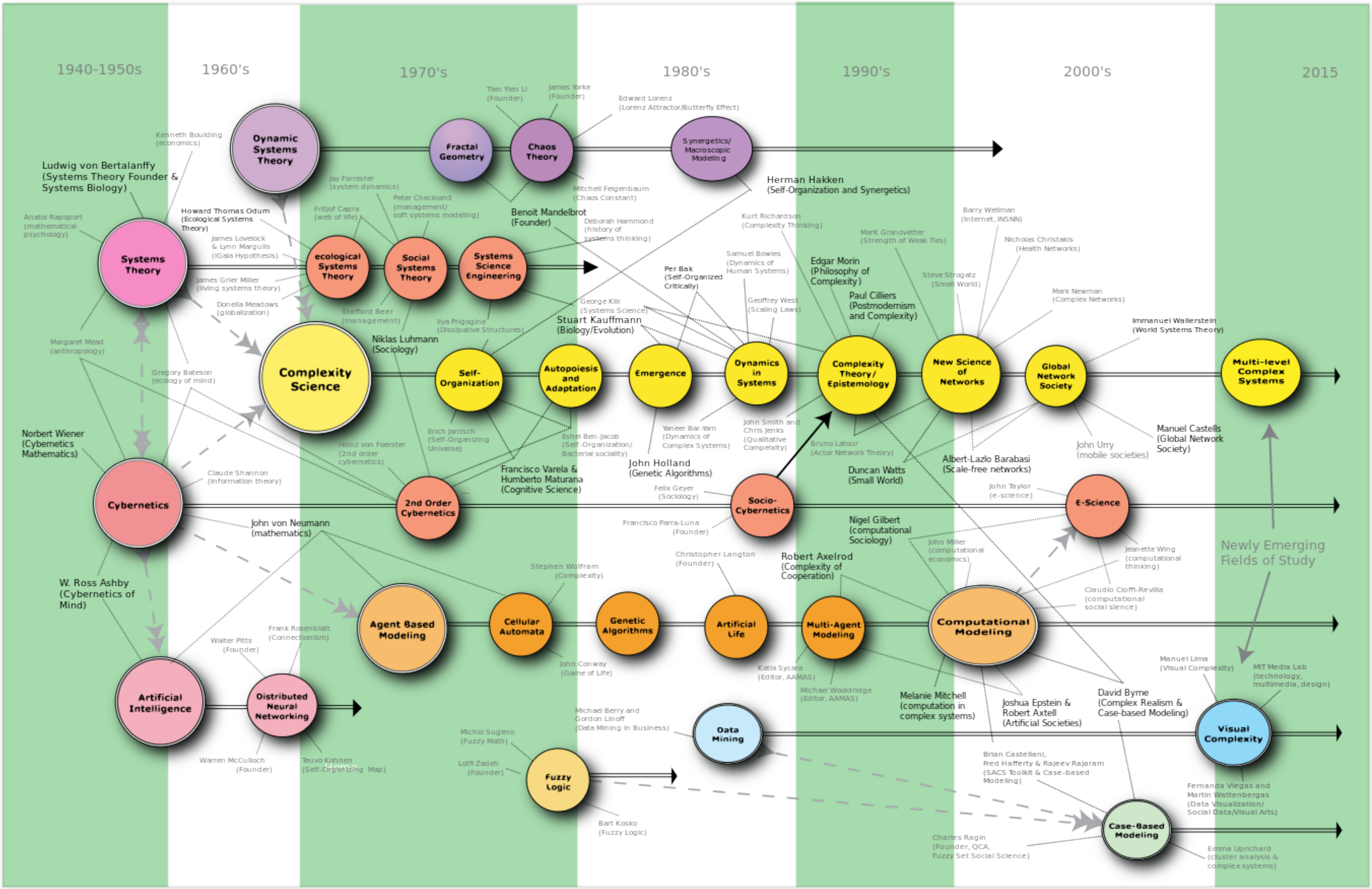
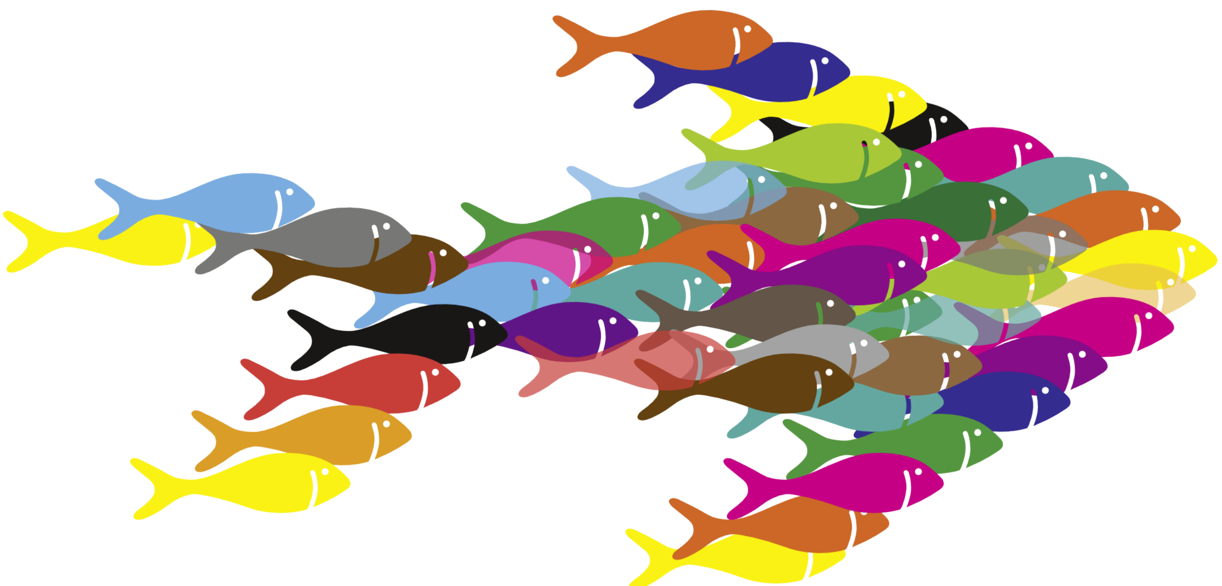


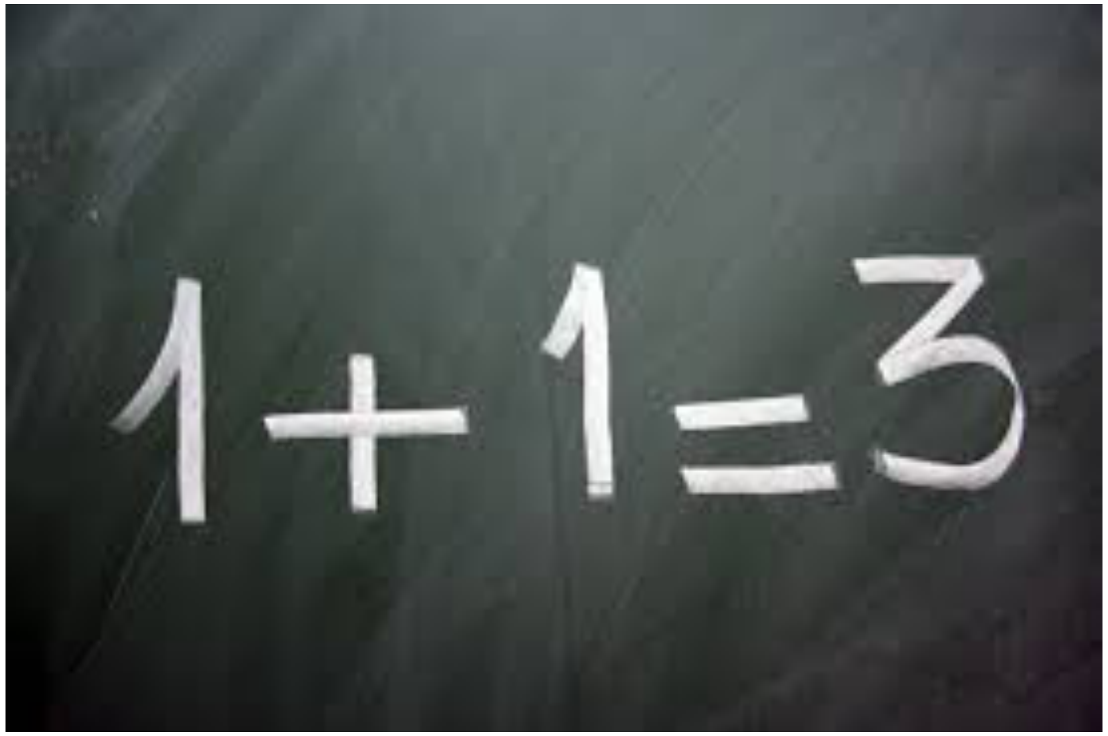
Figure from Brian Castellani

Emergent property

An emergent behavior or emergent property can appear when a number of simple entities (agents) interact in an environment, forming more complex behaviors as a collective.



A complex system (The whole) is more than the sum of its parts.
Aristotle



Spotlight

ARTWORK **Damián Ortega, *Cosmic Thing*,** 2002
Volkswagen Beetle 1983, stainless steel wire, acrylic
The Museum of Contemporary Art, Los Angeles

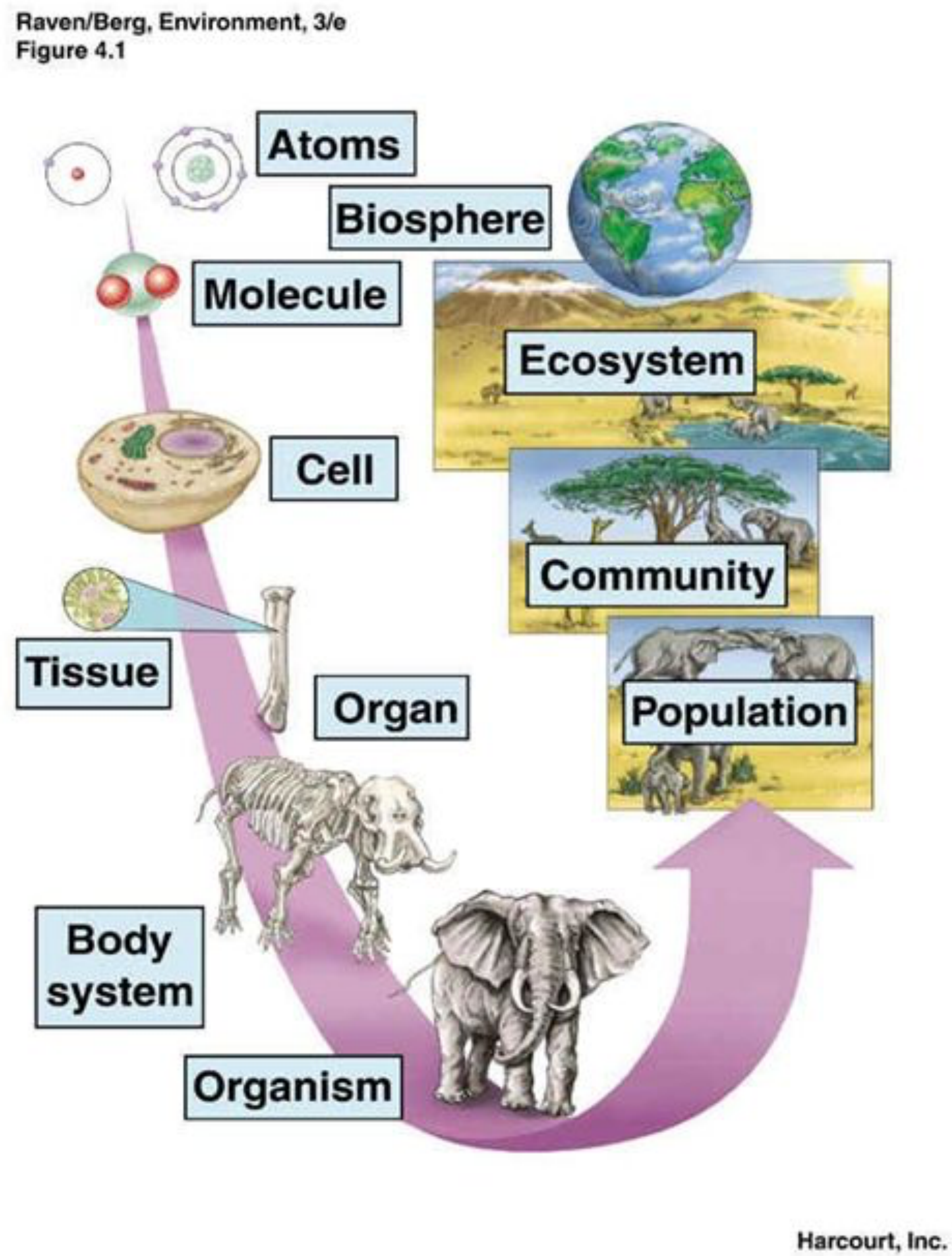
Installation view, *Damián Ortega: Do It Yourself*
The Institute of Contemporary Art/Boston
September 18, 2009–January 18, 2010

PHOTOGRAPHY: JOHN KENNARD

Reinvent Your Business Before It's Too Late

Watch Out for Those S Curves

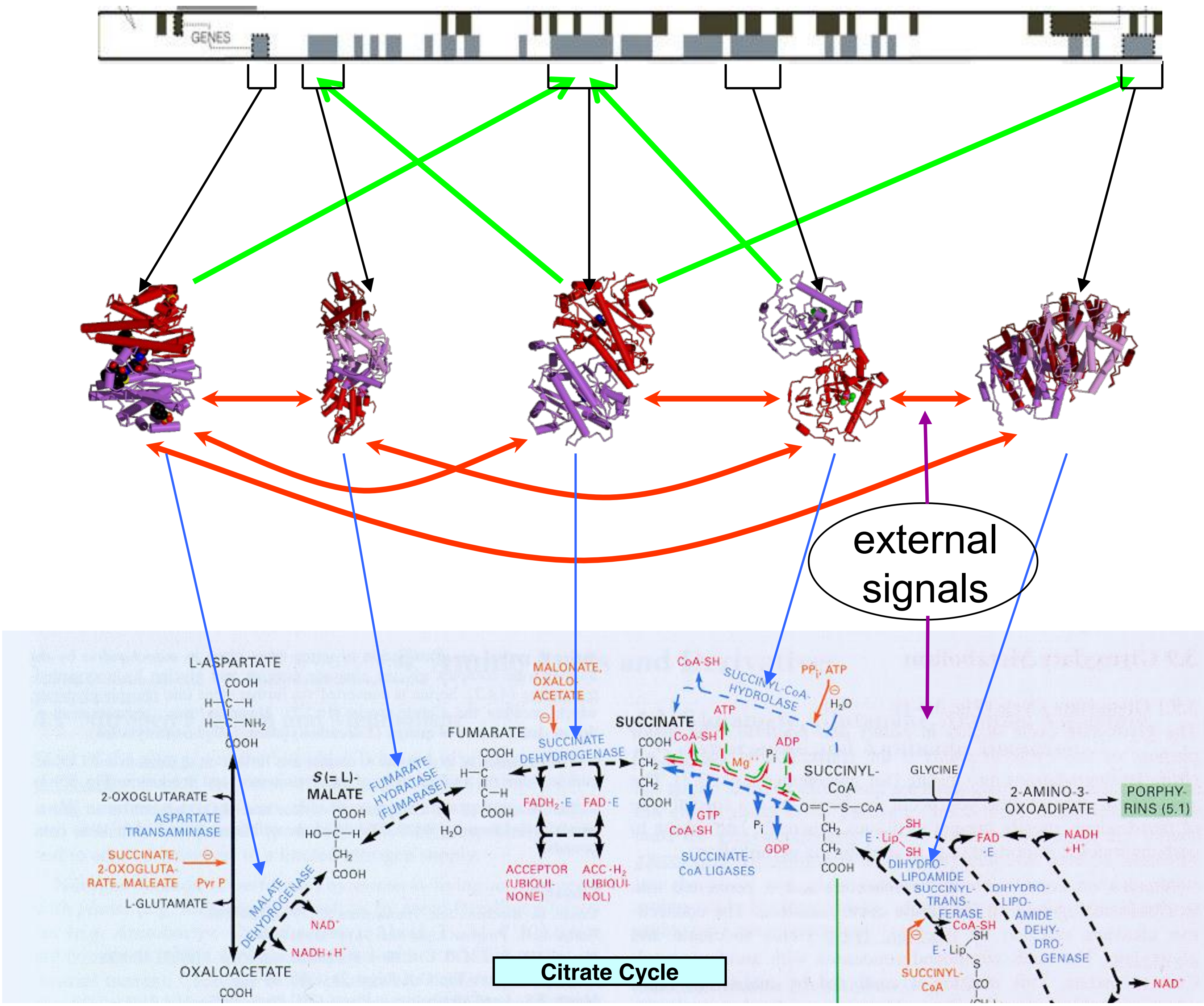
by Paul Nunes and Tim Breene



Each level of organization has **emergent properties.**

What the heck is an emergent property?

Quality that appears as biological complexity increases as it goes up a level and determined by interactions between individual parts



GENOME

gene regulation

PROTEOME

protein-protein interactions

signal transduction

METABOLISM

Bio-chemical reactions

Emergent properties: some examples

Emergent Properties of Networks of Biological Signaling Pathways

Upinder S. Bhalla and Ravi Iyengar*

Many distinct signaling pathways allow the cell to receive, process, and respond to information. Often, components of different pathways interact, resulting in signaling networks. Biochemical signaling networks were constructed with experimentally obtained constants and analyzed by computational methods to understand their role in complex biological processes. These networks exhibit emergent properties such as integration of signals across multiple time scales, generation of distinct outputs depending on input strength and duration, and self-sustaining feedback loops. Feedback can result in bistable behavior with discrete steady-state activities, well-defined input thresholds for transition between states and prolonged signal output, and signal modulation in response to transient stimuli. These properties of signaling networks raise the possibility that information for "learned behavior" of biological systems may be stored within intracellular biochemical reactions that comprise signaling pathways.

Studies on the cyclic adenosine monophosphate (cAMP) signaling pathway led to the identification of several general mechanisms of signal transfer, such as regulation by protein-protein interactions, protein phosphorylation, regulation of enzymatic activity, production of second messengers, and cell surface signal transduction systems (*1*). These mechanisms of signal transfer have subsequently been shown to occur in many pathways, including Ca^{2+} signaling pathways (*2*), tyrosine kinase pathways (*3*), and other protein kinase cascades, and recently in the intracellular protease cascades in apoptosis (*4*). Initially, signaling pathways were studied in a linear fashion, and it was shown that many important biological effects are obtained through linear information transfer. However, it has become increasingly clear that signaling pathways interact with one another and the final biological response is shaped by interaction between pathways. These interactions result in networks that are quite complex and may have properties that are nonintuitive. A systematic analysis of interactions between signaling pathways could be useful in understanding the properties of these networks. We developed models for simple net-

U. S. Bhalla, Department of Pharmacology, Mount Sinai School of Medicine, New York, NY 10029, USA, and National Centre for Biological Sciences, Post Office Box 1234, Bangalore 560012, India. R. Iyengar, Department of Pharmacology, Mount Sinai School of Medicine, New York, NY 10029, USA.

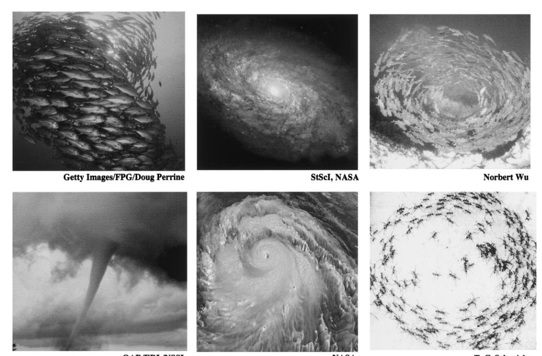
*To whom correspondence should be addressed.

www.sciencemag.org SCIENCE VOL 283 15 JANUAR

Self-Organized Fish Schools: An Examination of Emergent Properties

JULIA K. PARRISH[1,2,*], STEVEN V. VISCIDO[2], AND DANIEL GRÜNBAUM[3]

[1] *School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, Washington 98195-5020;* [2] *Zoology Department, University of Washington; and* [3] *School of Oceanography, University of Washington*



THE ROYAL SOCIETY biology letters

The emergent properties of a dolphin social network

David Lusseau†

Department of Zoology, University of Otago, PO Box 56, Dunedin, New Zealand (lusda563@student.otago.ac.nz)

Recd 29.05.03; *Accptd* 02.06.03; *Online* 04.07.03

Many complex networks, including human societies, the Internet, the World Wide Web and power grids, have surprising properties that allow vertices

Science REPORTS

Cite as: J. Liu *et al.*, *Science* 10.1126/science. aah4204 (2017).

Coupling between distant biofilms and emergence of nutrient time-sharing

Jintao Liu,[1] Rosa Martinez-Corral,[2]* Arthur Prindle,[1]* Dong-yeon D. Lee,[1] Joseph Larkin,[1] Marçal Gabalda-Sagarra,[2] Jordi Garcia-Ojalvo,[2] Gürol M. Süel[1,3,4]†

[1]Division of Biological Sciences, University of California, San Diego, CA 92093, USA. [2]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003 Barcelona, Spain. [3]San Diego Center for Systems Biology, University of California, San Diego, CA 92093, USA. [4]Center for Microbiome Innovation, University of California, San Diego, CA 92093, USA.

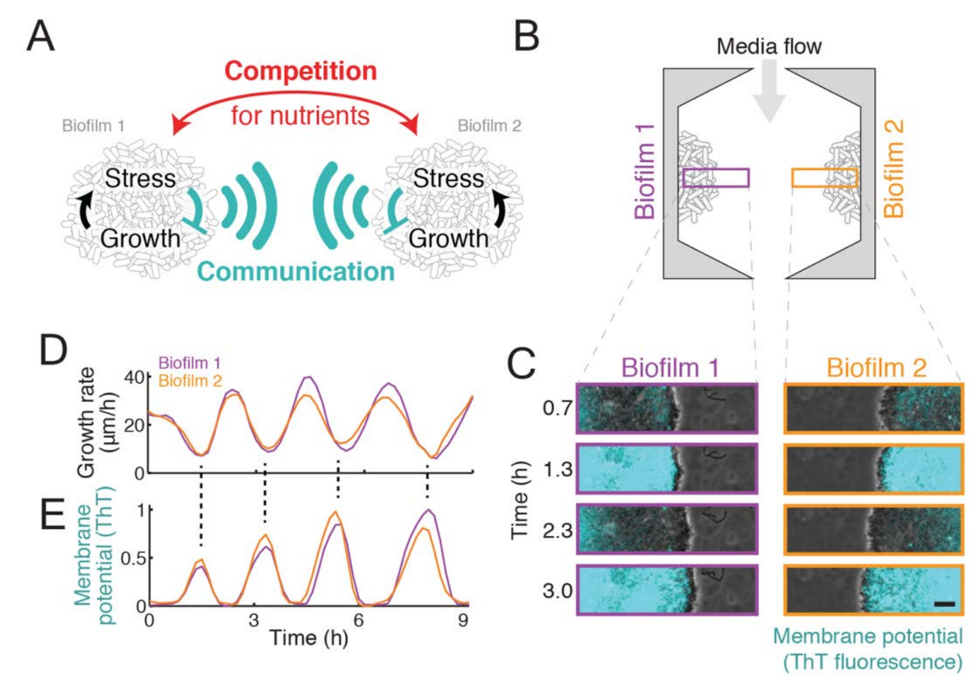
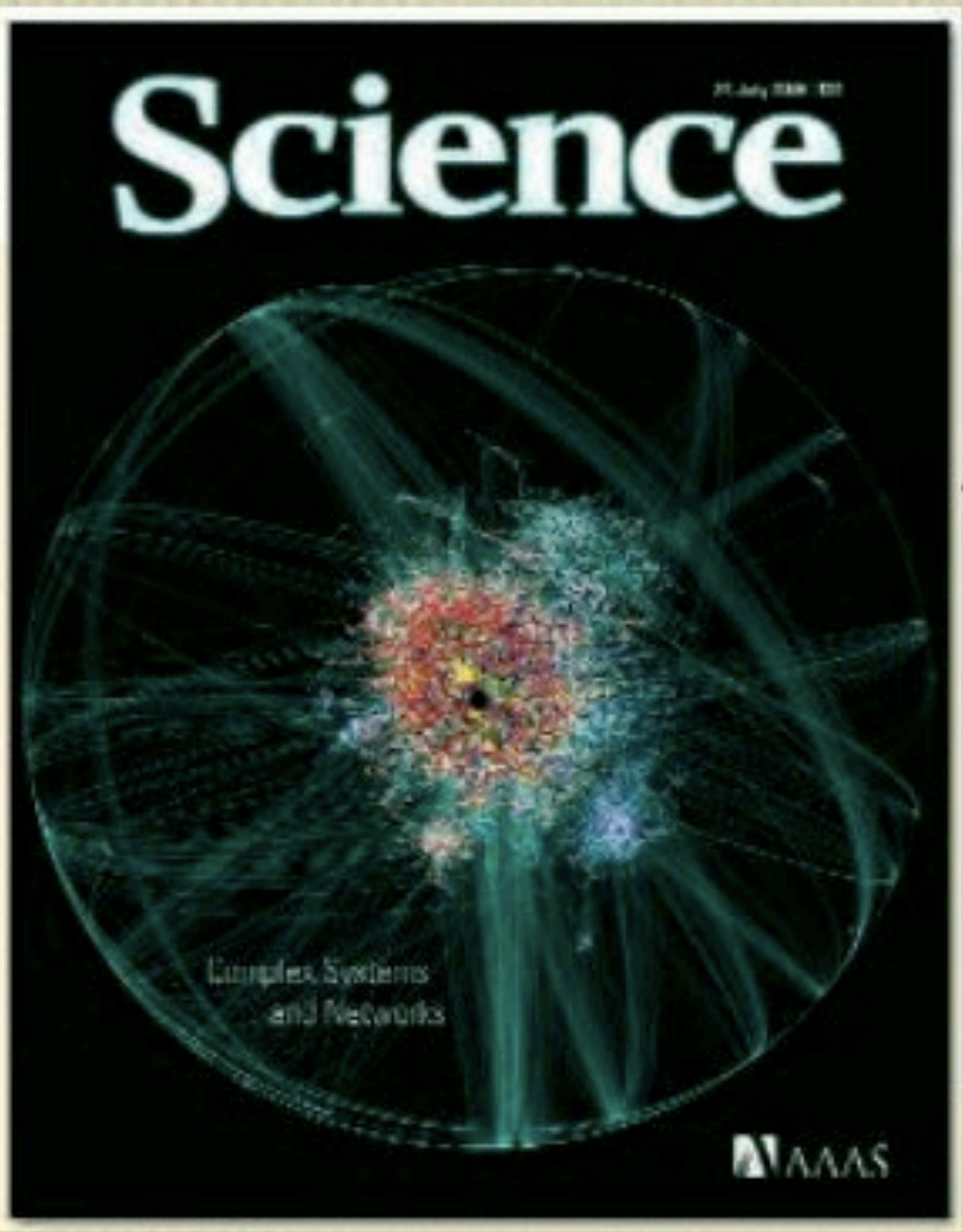


Fig. 1. Distant biofilms synchronize their growth dynamics. (A)



SPECIAL SECTION

Complex Systems and Networks

HAT TIP TO DREW CONWAY OF ZIA

Nicolas Perony: TED Talk

https://youtu.be/0Y8-IzP01lw?t=1m21s

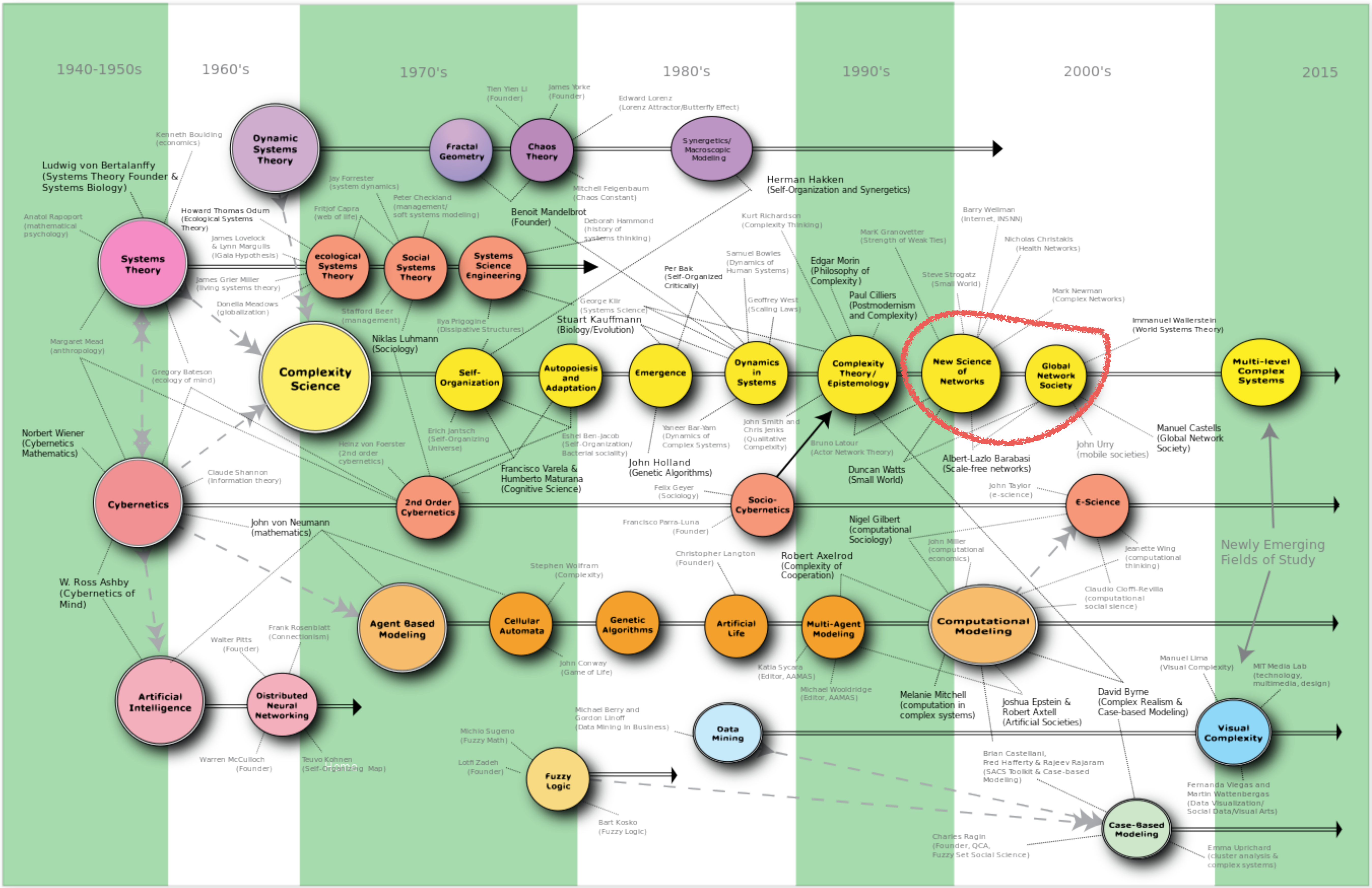


Figure from Brian Castellani

Networks: (a) tool for studying complex systems

- A big number of elements connected by some kind of interaction. For example cell networks: protein interactions, chemical reactions and regulatory relationships.

- Network’s properties provide relevant information of interactions

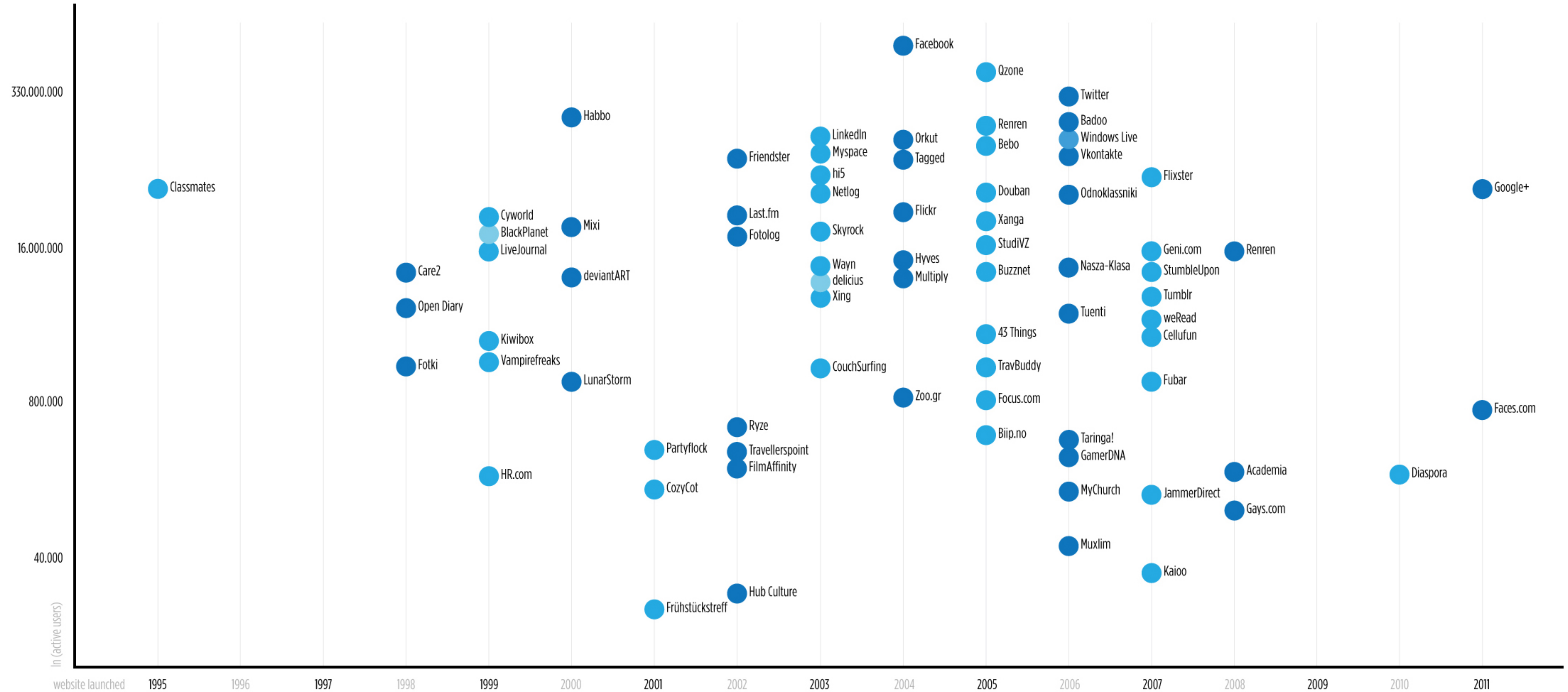
- Network’s architecture could provide a measure of robustness and an insight of the dynamically involve.

- Understand emergent properties: sync, cell differentiation, homeostasis.

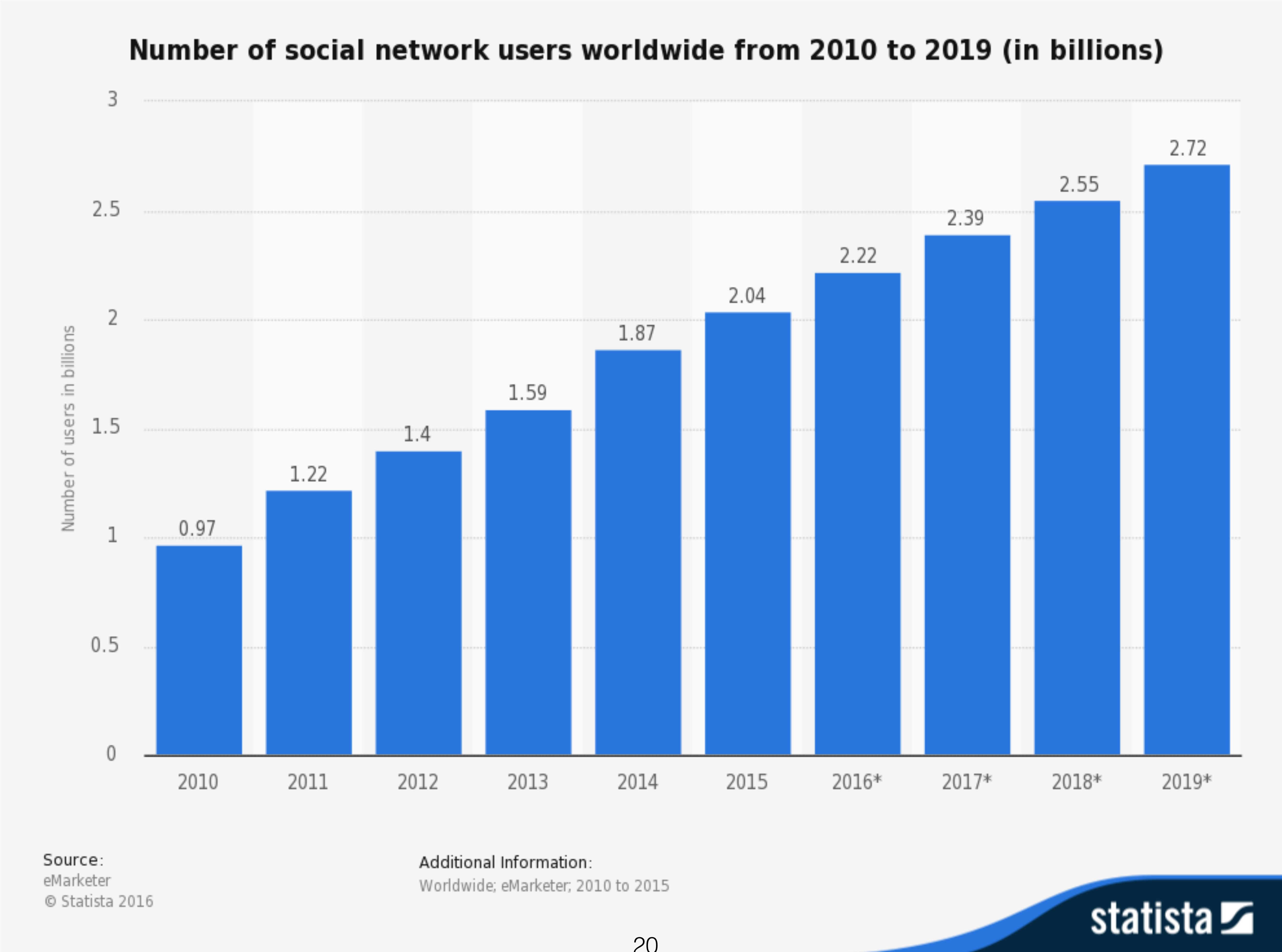
Tools for Studying Complex Systems
Cellular Automaton (Ulam, von Neumann 1940, Wolfram 1980)
Agent-Based Models (von Neumann, Ulam, Conway)
Complex Networks (Barabasi, Albert, Newman, Strogatz)
Machine Learning (Samuel)

Networks: Why now?

- Availability of Data
- Scientifically interesting topic
- Necessary for understanding our complex world



330.000.000
16.000.000
800.000
40.000
ln (active users)
website launched
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
Facebook
Qzone
Twitter
Habbo
LinkedIn
Myspace
Orkut
Renren
Badoo
Windows Live
Vkontakte
Tagged
Bebo
Friendster
hi5
Netlog
Flixster
Classmates
Douban
Odnoklassniki
Google+
Flickr
Cyworld
Mixi
Last.fm
Xanga
BlackPlanet
Fotolog
Skyrock
StudiVZ
LiveJournal
Geni.com
Renren
Hyves
Care2
Wayn
Nasza-Klasa
deviantART
Multiply
Buzznet
StumbleUpon
delicius
Xing
Tumblr
Open Diary
Tuenti
weRead
Cellufun
43 Things
Kiwibox
Vampirefreaks
Fotki
CouchSurfing
TravBuddy
LunarStorm
Fubar
Zoo.gr
Focus.com
Faces.com
Ryze
Biip.no
Taringa!
Partyflock
Travellerspoint
GamerDNA
FilmAffinity
Academia
Diaspora
HR.com
CozyCot
MyChurch
JammerDirect
Gays.com
Muxlim
Kaioo
Hub Culture
Frühstückstreff


Number of social network users worldwide from 2010 to 2019 (in billions)
Number of users in billions
3
2.5
2
1.5
1
0.5
0
0.97
1.22
1.4
1.59
1.87
2.04
2.22
2.39
2.55
2.72
2010
2011
2012
2013
2014
2015
2016*
2017*
2018*
2019*
Source:
eMarketer
© Statista 2016
Additional Information:
Worldwide; eMarketer; 2010 to 2015
statista

Two Forces that Helped Network Science

- The Emergence of Network Maps
- The Universality of Network Characteristics

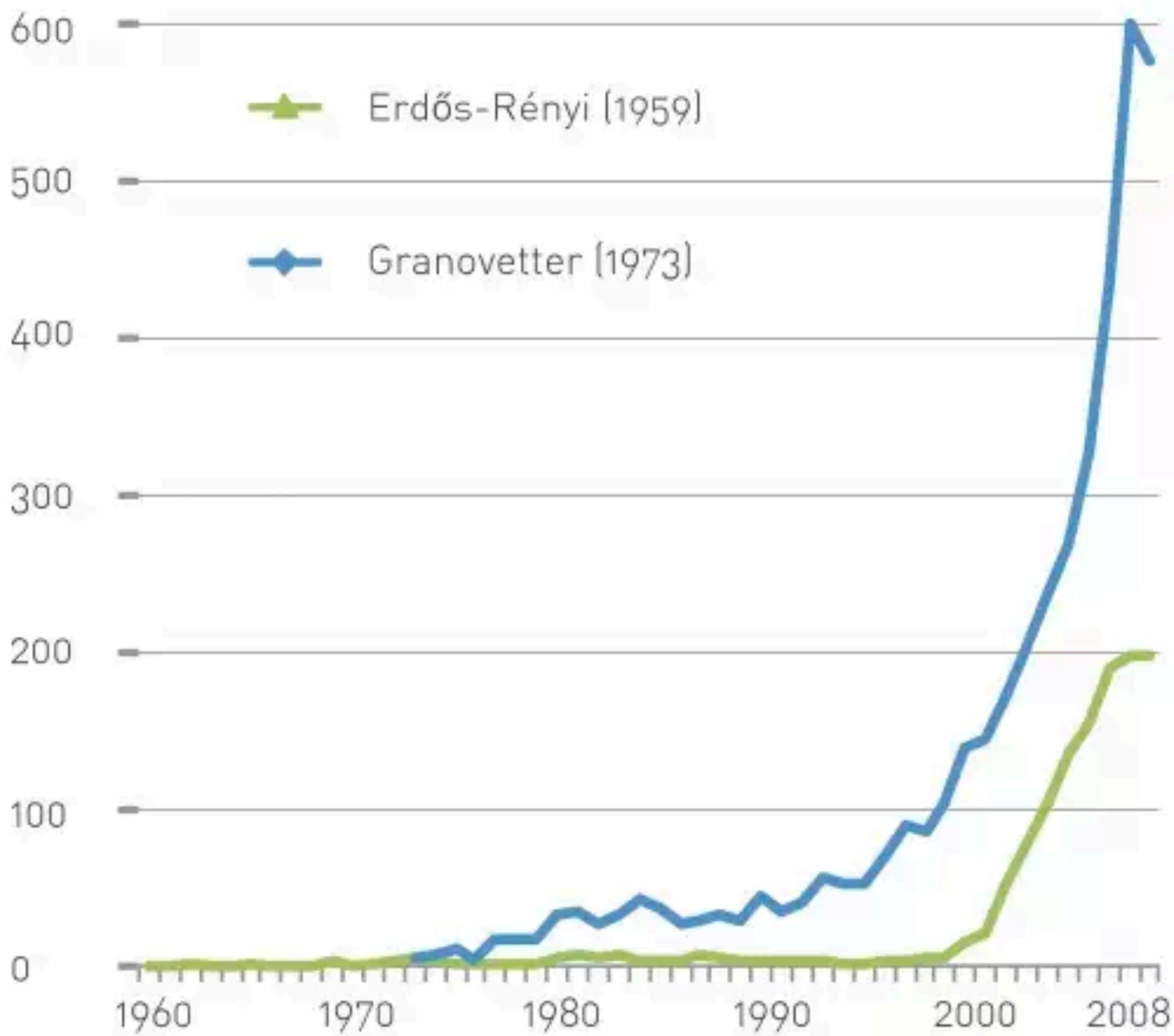
The Origins of Network Maps

A few of the maps studied today by network scientists were generated with the purpose of studying networks. Most are the byproduct of other projects and morphed into maps only in the hands of network scientists.

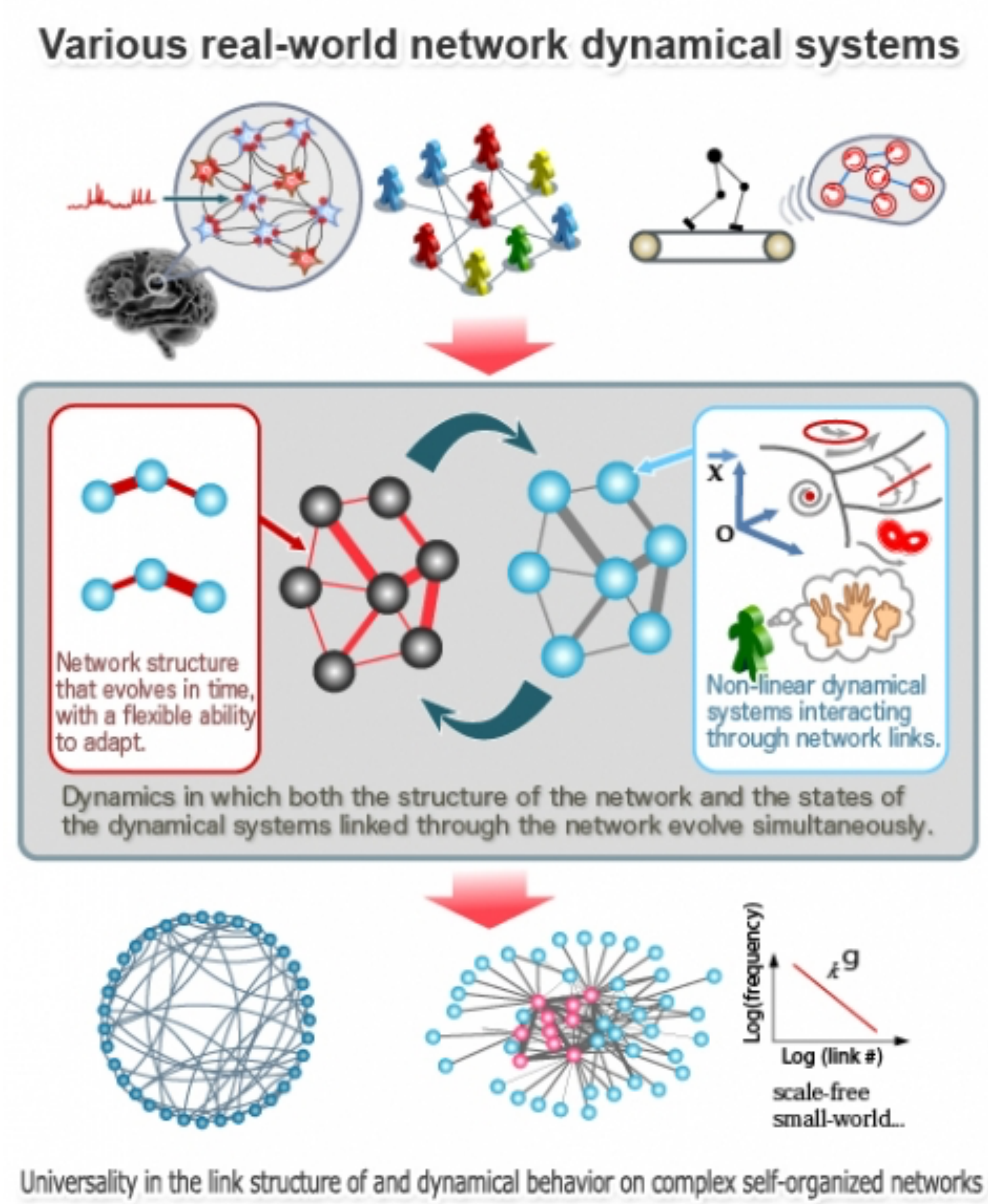
- ✤ The list of chemical reactions in a cell were discovered one by one over a 150 year period by biochemists. In the 1990s they were collected in central databases, offering the first chance to assemble the biochemical networks within a cell.
- ✤ The list of actors that play in each movie were traditionally scattered in newspapers, books and encyclopedias. With the advent of th Internet, these data were assembled into central databases, like imdb.com, feeding the curiosity of movie aficionados. The databas allowed network scientists to reconstruct the affiliation network behind Hollywood.
- ✤ The list of authors of millions of research papers were traditionally scattered in the table of content of thousands of journals. Recen Web of Science, Google Scholar, and other services have assembled them into comprehensive databases, allowing network scientists to reconstruct accurate maps of scientific collaboration networks.

Much of the early history of network science relied on the investigators' ingenuity to recognize and extract networks from preexisting databases. Network science changed that: Today well-funded research collaborations focus on map making, capturing accurate wiring diagrams of biological, communication and social systems.

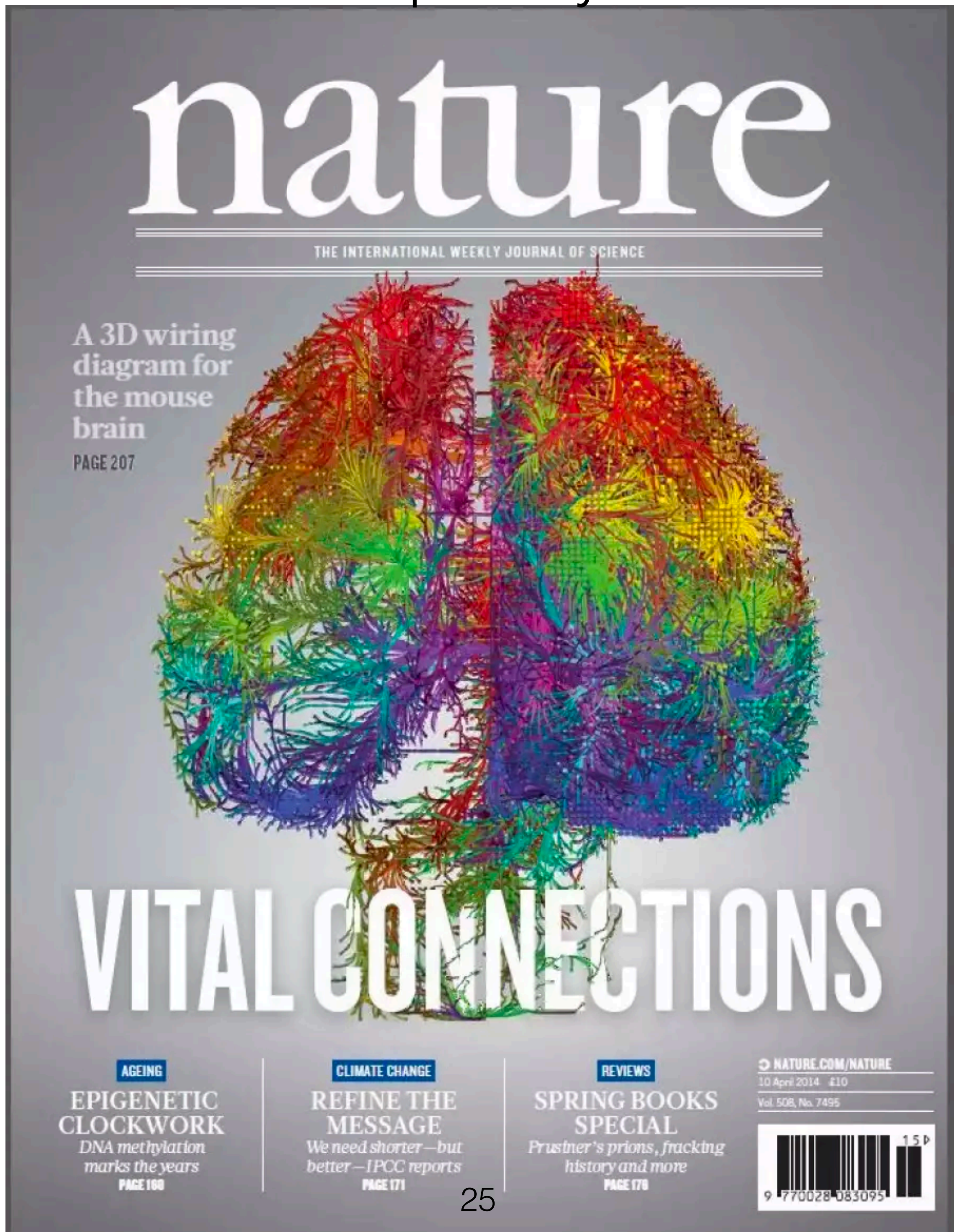
The Emergence of Network Science



The Universality of Network Characteristics



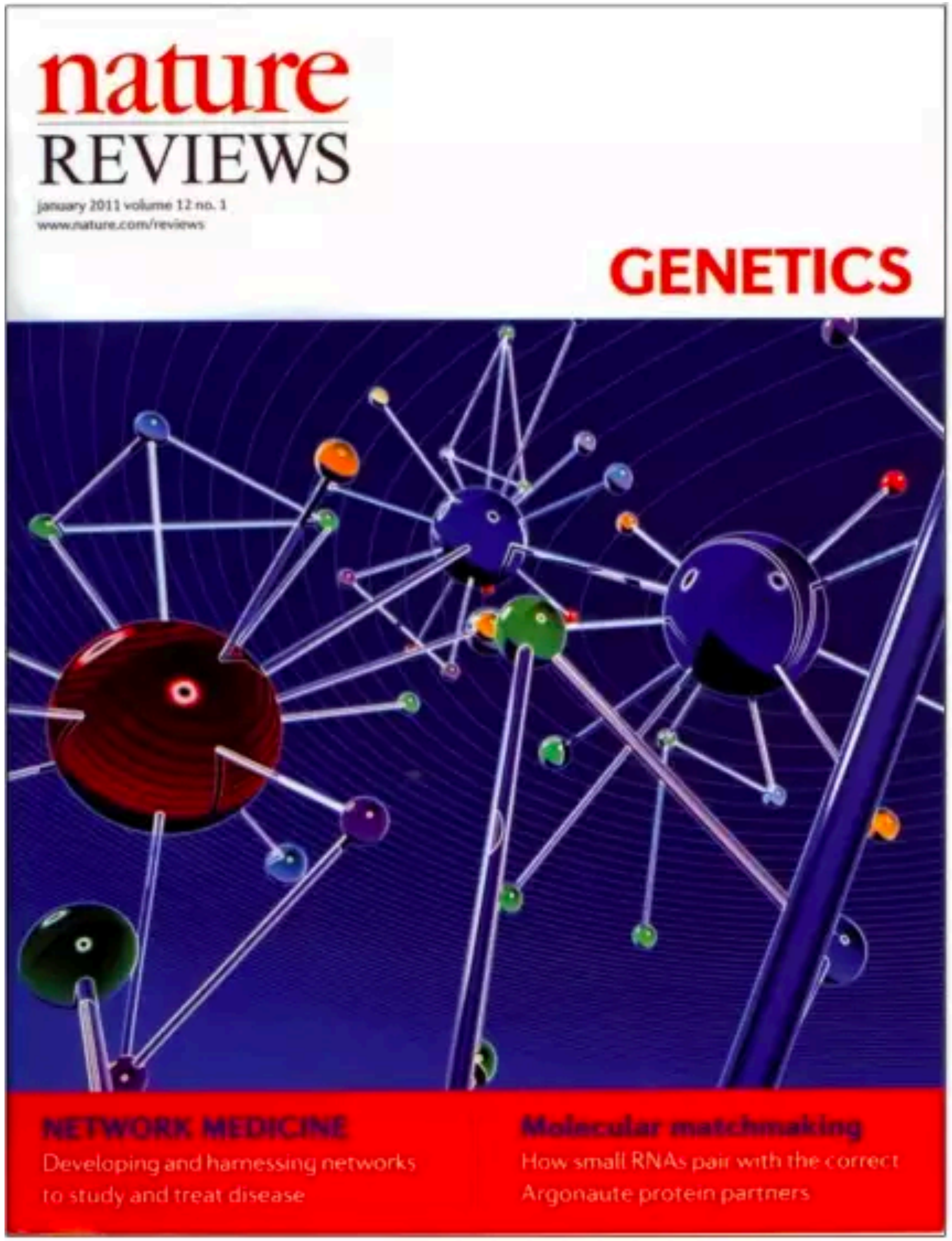
The characteristic of Network Science
Interdisciplinary Nature



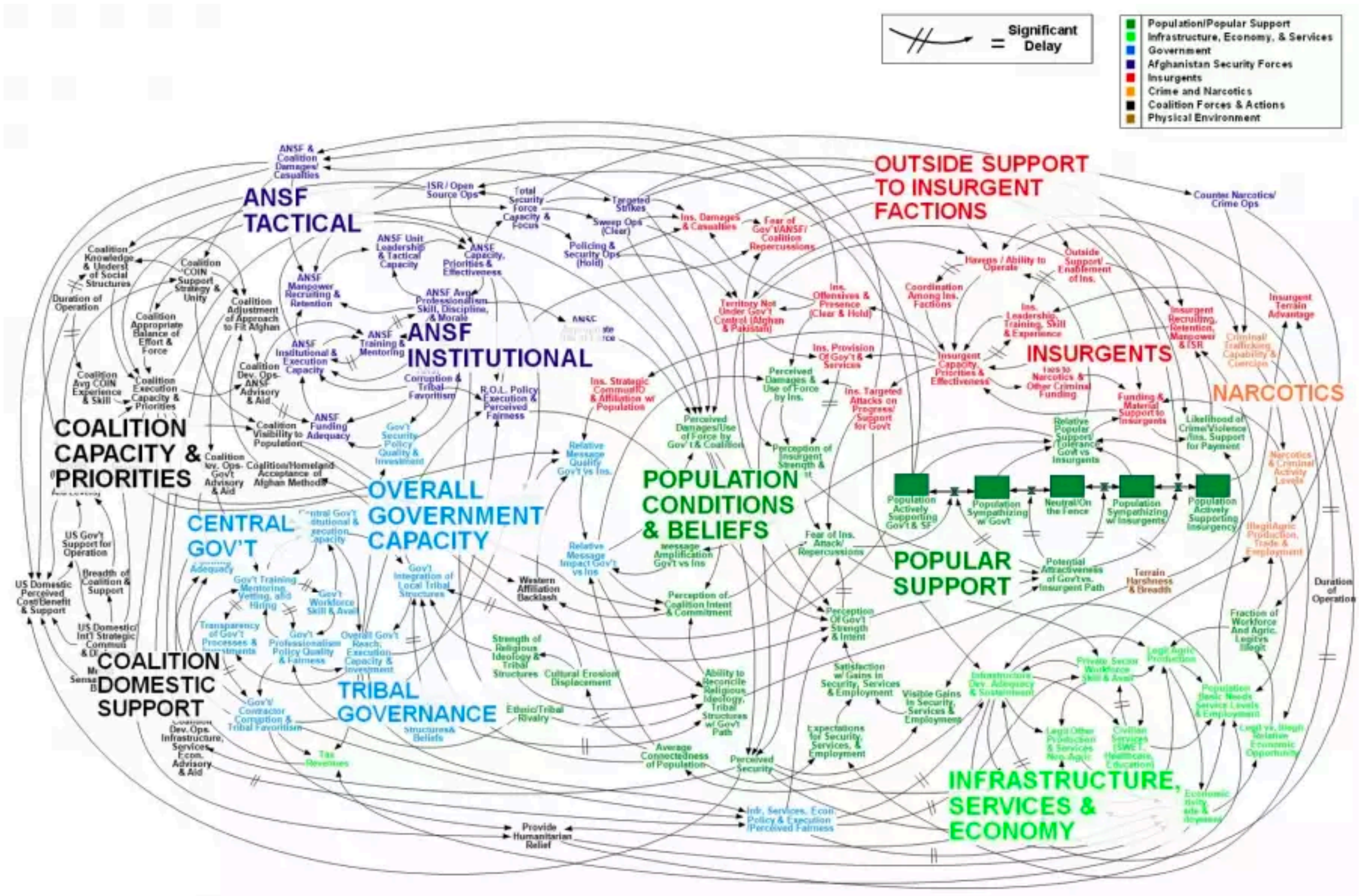
Societal Impact

- Economic Impact: From Web Search to Social Networking
- Health: From Drug Design to Metabolic Engineering
- Security: Fighting Terrorisms

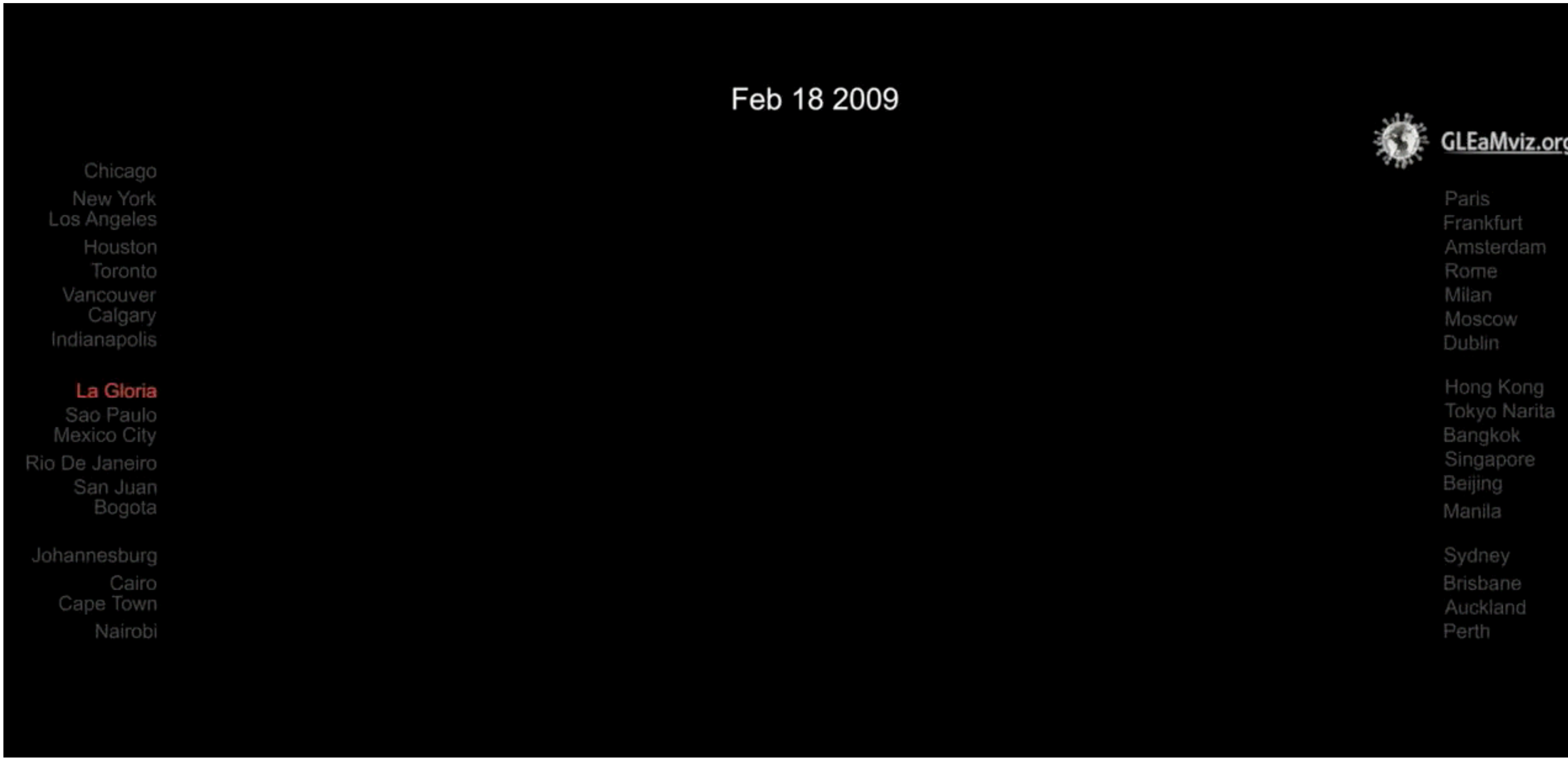
Network Biology and Medicine



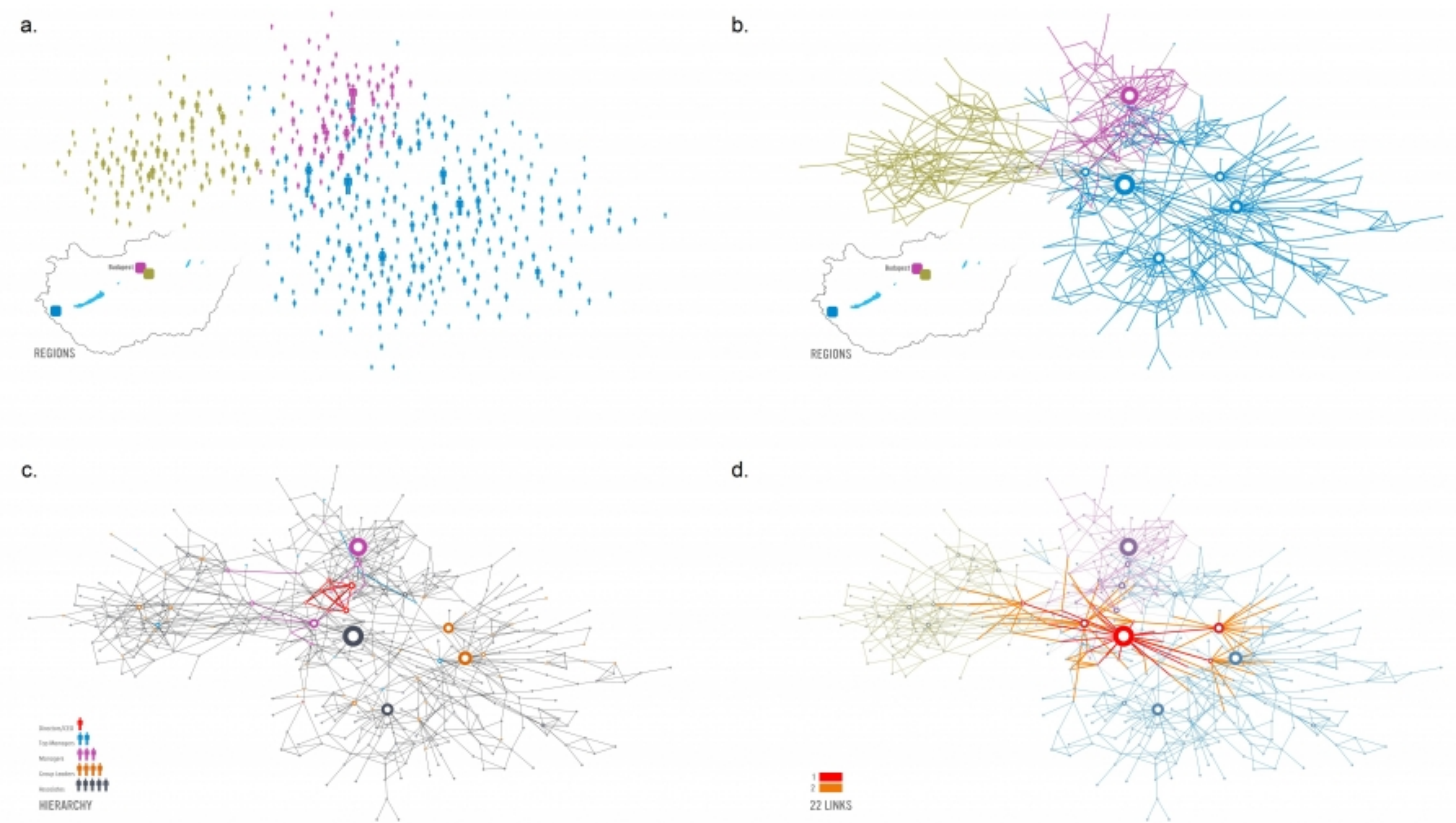
The Network Behind Military Engagement



Predicting the H1N1 Epidemic



Mapping Organizations



Complexity and Network Science

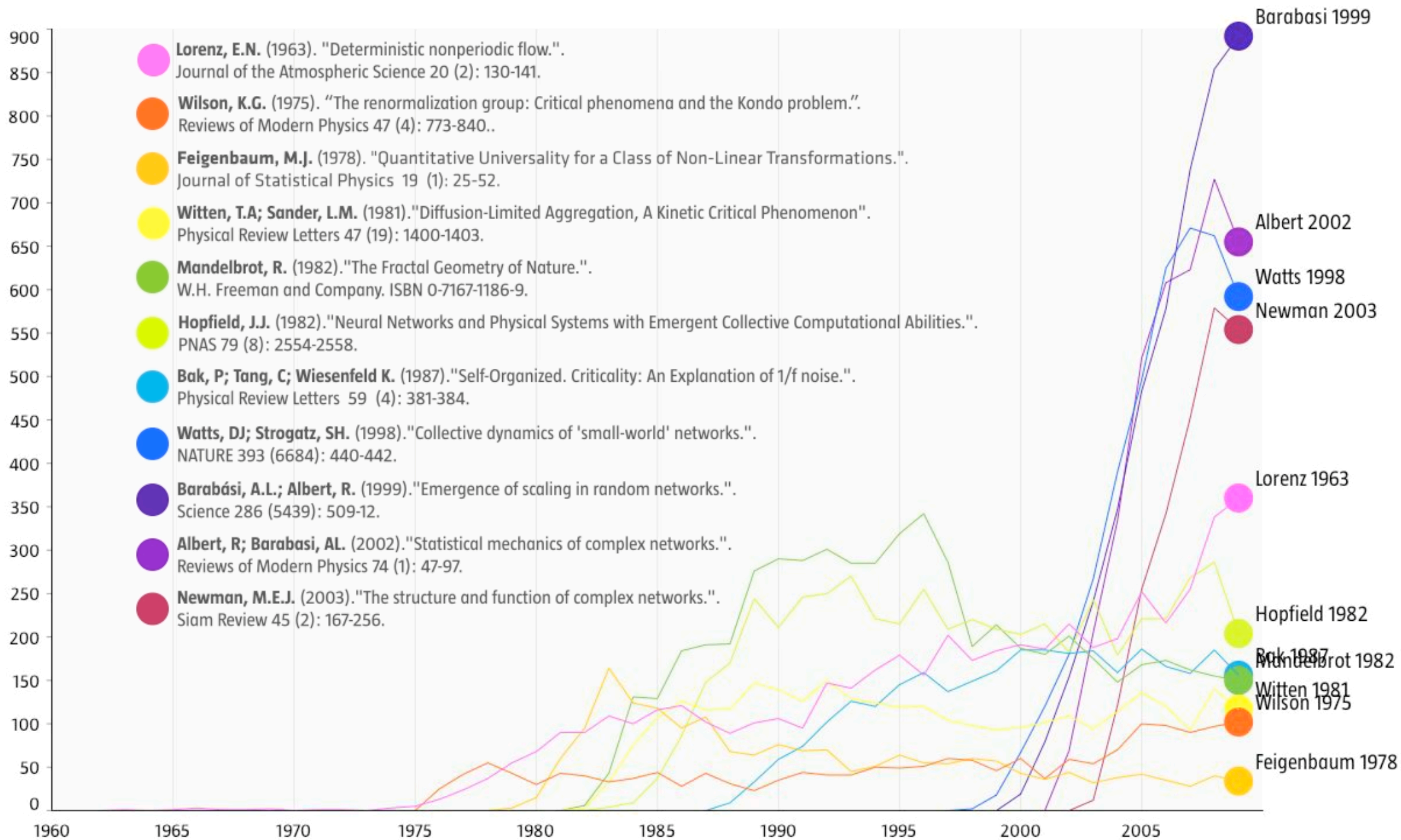
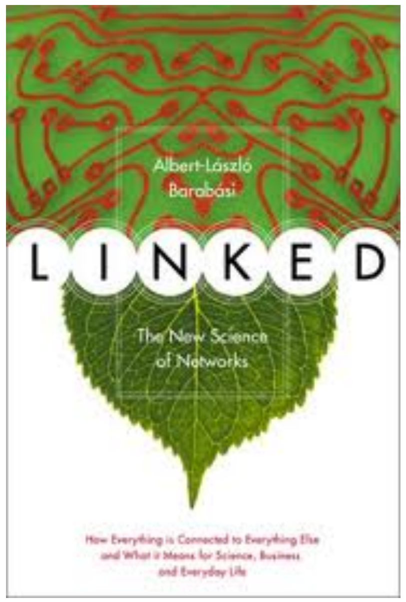
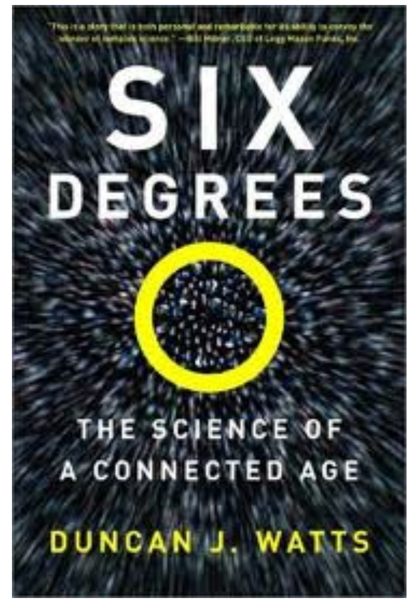
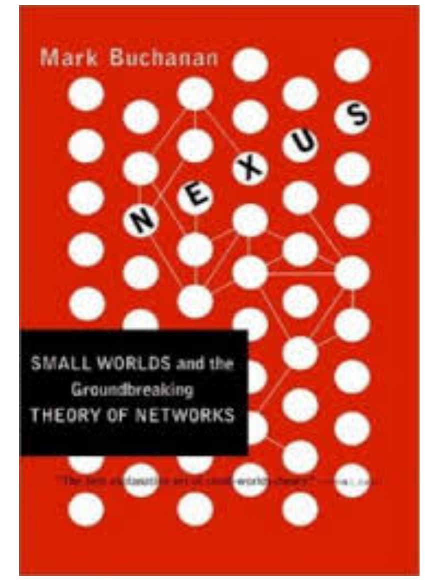


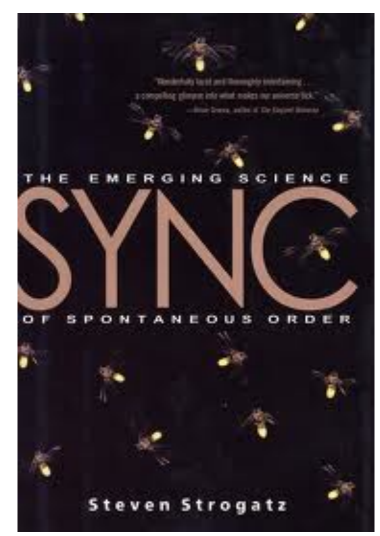
Figura: Barabasi et al. 2011

Popular Science Books

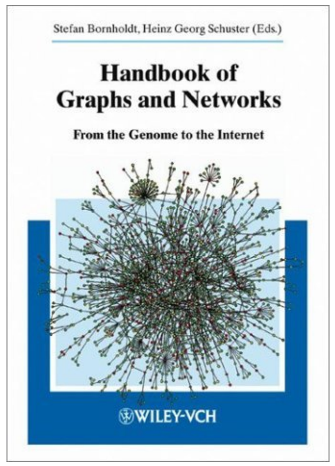




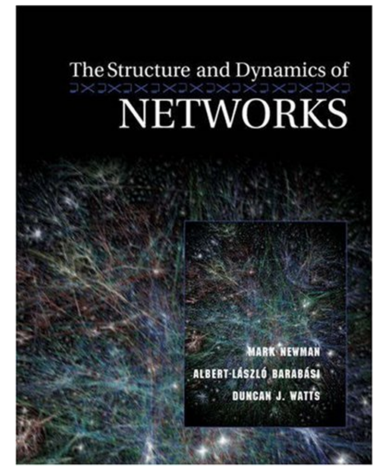




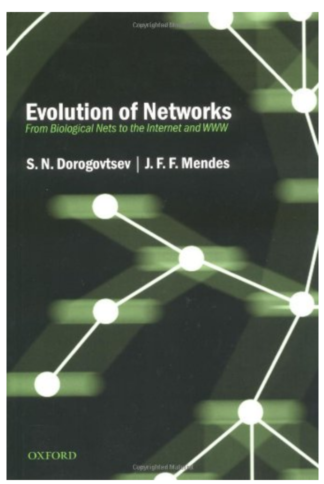
Textbook/Course Books



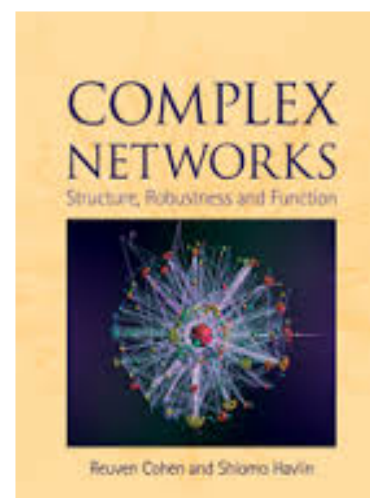
Handbook of Graphs and Networks: From the Genome to the Internet (Wiley-VCH, 2003).



M. Newman, A.-L. Barabasi, D. J. Watts, The Structure and Dynamics of Networks: (Princeton Studies in Complexity) (Princeton University Press, 2006)

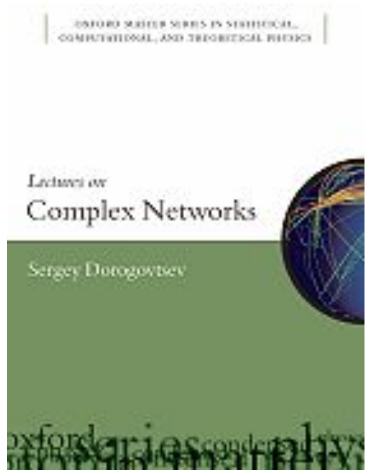


S. N. Dorogovtsev and J. F. F. Mendes, Evolution of Networks: From Biological Nets to the Internet and WWW (Oxford University Press, 2003).

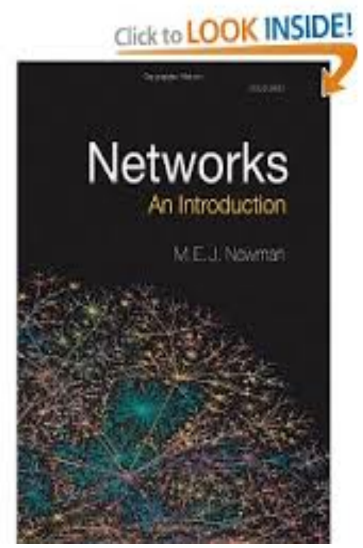


R. Cohen, S. Havlin, Complex Networks. Structure, Robustness and Function, (Cambridge University Press, 2010)

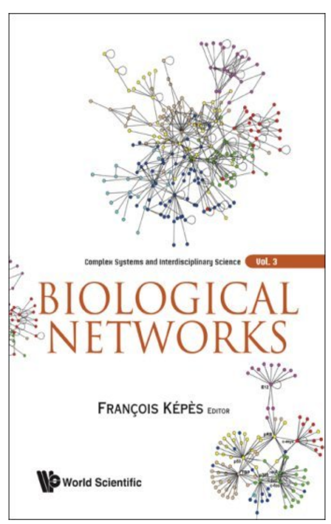
Textbook/Course Books



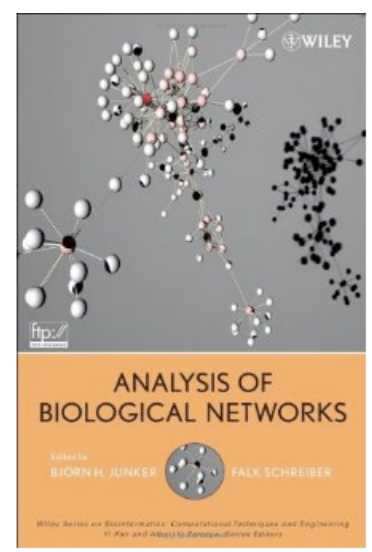
S. Dorogovtsev, Lectures on Complex Networks, Oxford Master Series in Physics, 2010



M.E.J. Newman, Networks. An Introduction, Oxford University Press,2010

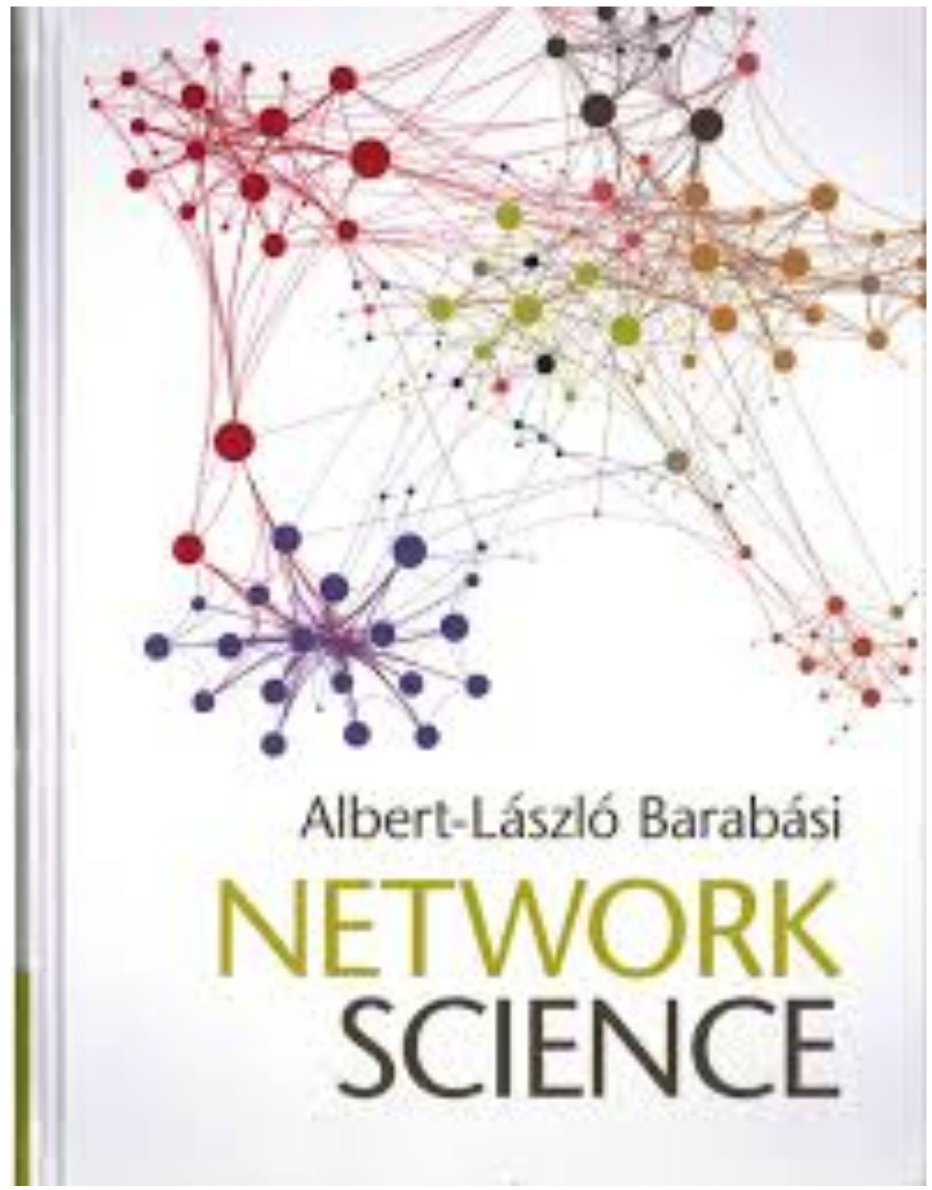


F. Kopos, Biological Networks (Complex Systems and Interdisciplinary Science) (World Scientic Publishing Company, 2007)



B. H. Junker, F. Schreiber, Analysis of Biological Networks (Wiley Series in Bioinformatics) (Wiley-Interscience, 2008).

Textbook/Course Books

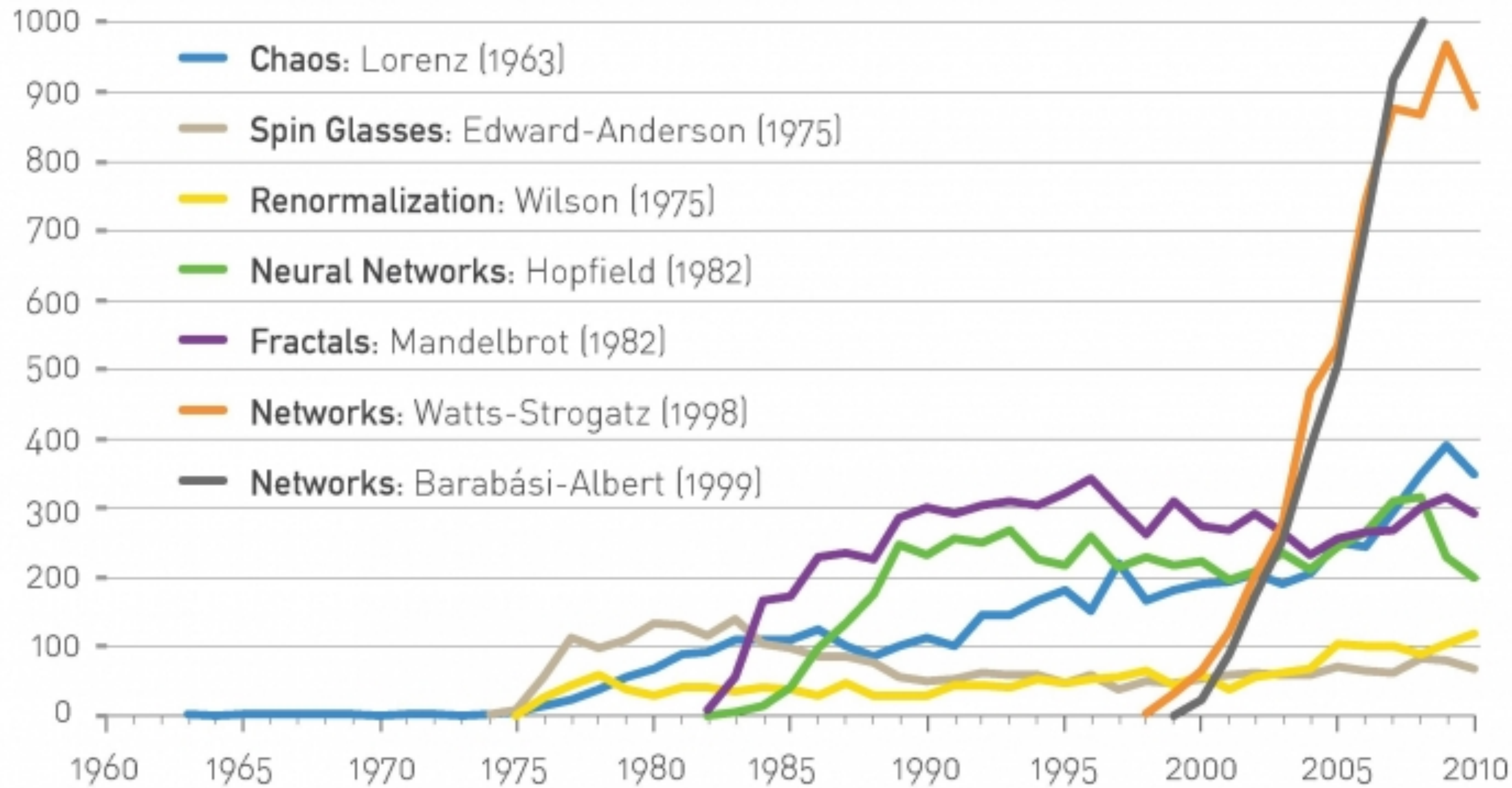


Some fundamental papers

- 1959: Erdös, P.; Rényi, A. “On Random Graphs. I”. Publicationes Mathematicae 6: 290–297
- 1998: Watts y Strogatz ”Collective dynamics of “small-world” networks. Nature 393 (6684): 440–442
- 1999: Barabasi y Albert “Emergence of scaling in random networks”. Science 286: pp. 509–512
- 2001: Pastor -Satorras y Vespignani “Epidemic spreading in scale-free networks” Physical Review Letters: 86 (14), 3200
- 2002: Girvan y Newman “Community structure in social and biological network”s, Proc. Natl. Acad. Sci. USA 99, 7821–7826.

Reviews

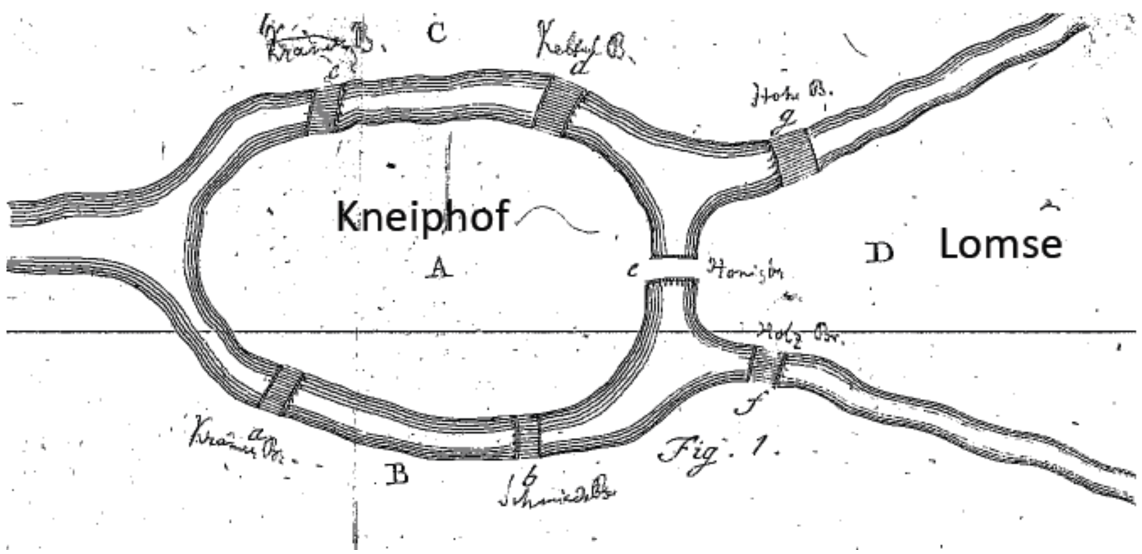
- R. Albert, A-L Barabasi, “Statistichal Mechanics of Complex Networks”, REviews of Modern Physics, 2001
- A-L Barabasi, ZN. Oltvai, “Network biology: understanding the cell’s functional organization”, Nature Reviews Genetics, 2004
- R. Albert, “Scale-free networks in cell biology”, Journal of cell science, 2005

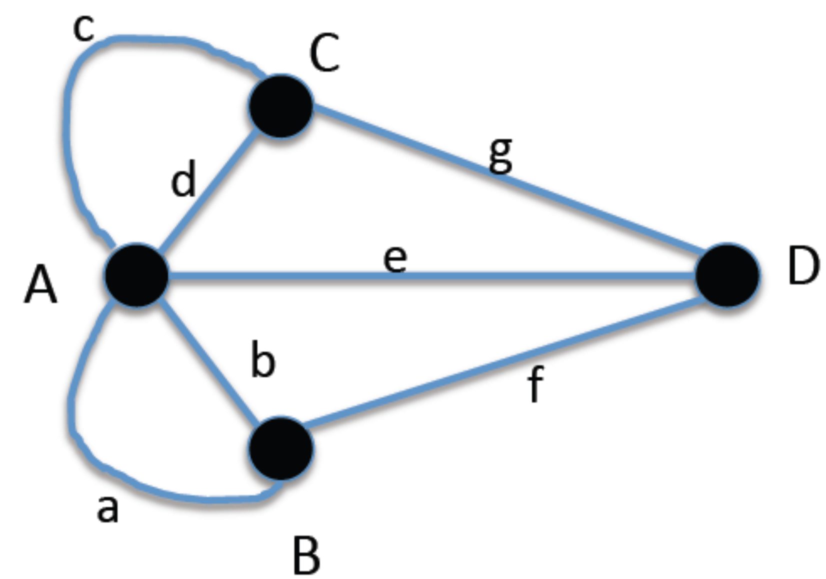


1000
900
800
700
600
500
400
300
200
100
0
Chaos: Lorenz (1963)
Spin Glasses: Edward-Anderson (1975)
Renormalization: Wilson (1975)
Neural Networks: Hopfield (1982)
Fractals: Mandelbrot (1982)
Networks: Watts-Strogatz (1998)
Networks: Barabási-Albert (1999)
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010

Connected

https://youtu.be/zK1Cb9qj3qQ?t=10





Graph Theory

Basic Concepts

Network: Dictionary Definition

a group or system of interconnected people or things: *a trade network*.
• a group of people who exchange information, contacts, and experience for professional or social purposes: *a support network*.
• a group of broadcasting stations that connect for the simultaneous broadcast of a program: *the introduction of a second TV network* | *[as modifier] : network television*.
• a number of interconnected computers, machines, or operations: *specialized computers that manage multiple outside connections to a network* | *a local cellular phone network*.
• a system of connected electrical conductors.

Network: Dictionary Definition

a group or system of interconnected people or things: *a trade network*.
• a group of people who exchange information, contacts, and experience for professional or social purposes: *a support network*.
• a group of broadcasting stations that connect for the simultaneous broadcast of a program: *the introduction of a second TV network* | *[as modifier] : network television*.
• a number of interconnected computers, machines, or operations: *specialized computers that manage multiple outside connections to a network* | *a local cellular phone network*.
• a system of connected electrical conductors.

Examples: **Internet** (Computers, physical connections), **WWW** (web pages, hyper links), **Proteins** (Physical Interactions), **Metabolic network** (reactants, chemical reactions) , **Friends** (people, friendship),

Network: Definition

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

$$\mathcal{V} = Vertex, nodes$$

$$\mathcal{E} = Edges, connections$$

Network: Definition

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

$$\mathcal{V} = Vertex, nodes$$

$$\mathcal{E} = Edges, connections$$

Vertices, nodes: people, species, metabolites, proteins, genes, neurons,…
Edges, connections: physical contact, prey-predator, chemical reaction, binding, co-expression/regulation, activation,…

Network: Definition

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

$$\mathcal{V} = Vertex, nodes$$

$$\mathcal{E} = Edges, connections$$

Transcriptional Network

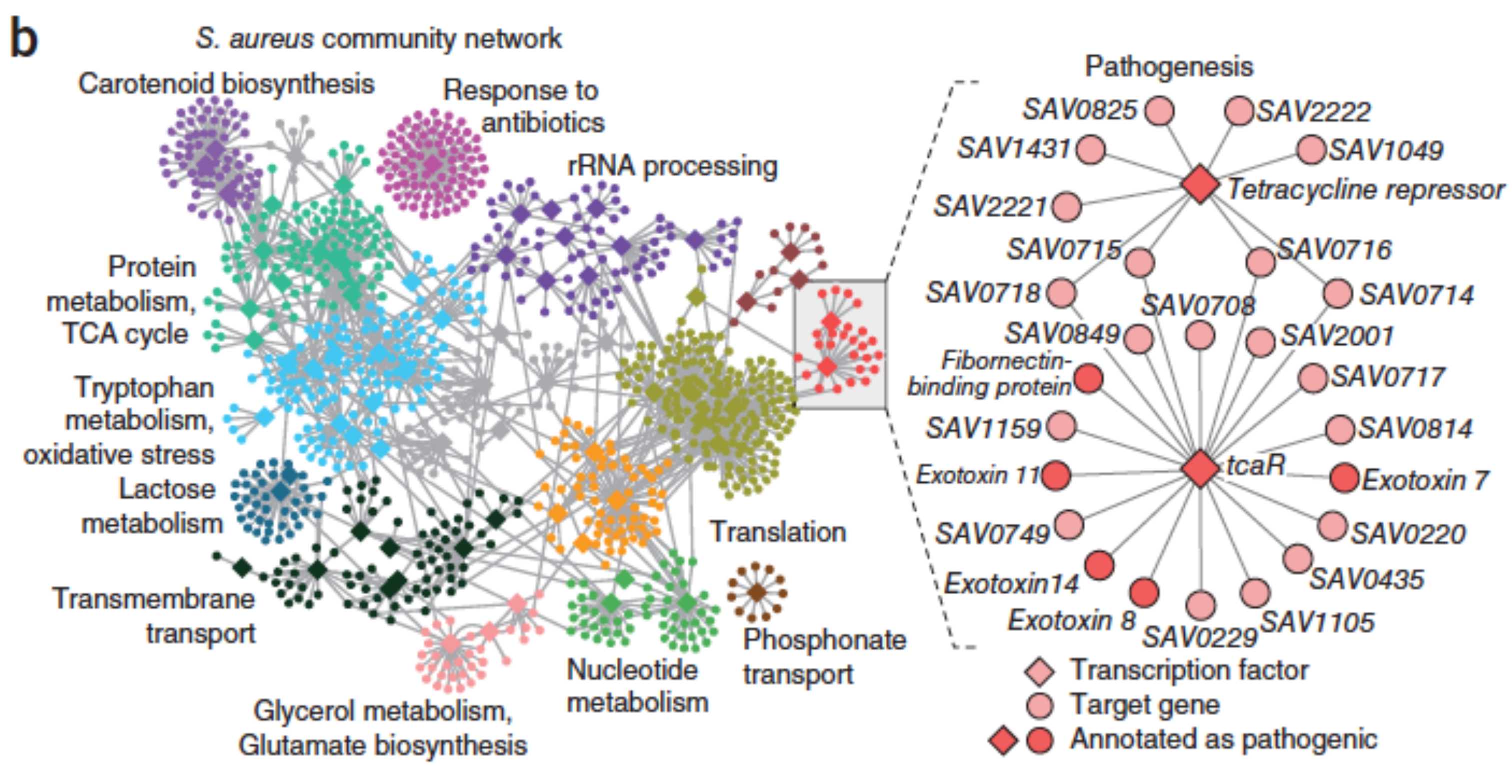


Figura: Redes de *E. coli* y *S. aureus*, $\sim 1,700$ interacciones transcripcionales (Marbach *et al.*, Nature Methods, 2012)

Protein-Protein Interaction Network

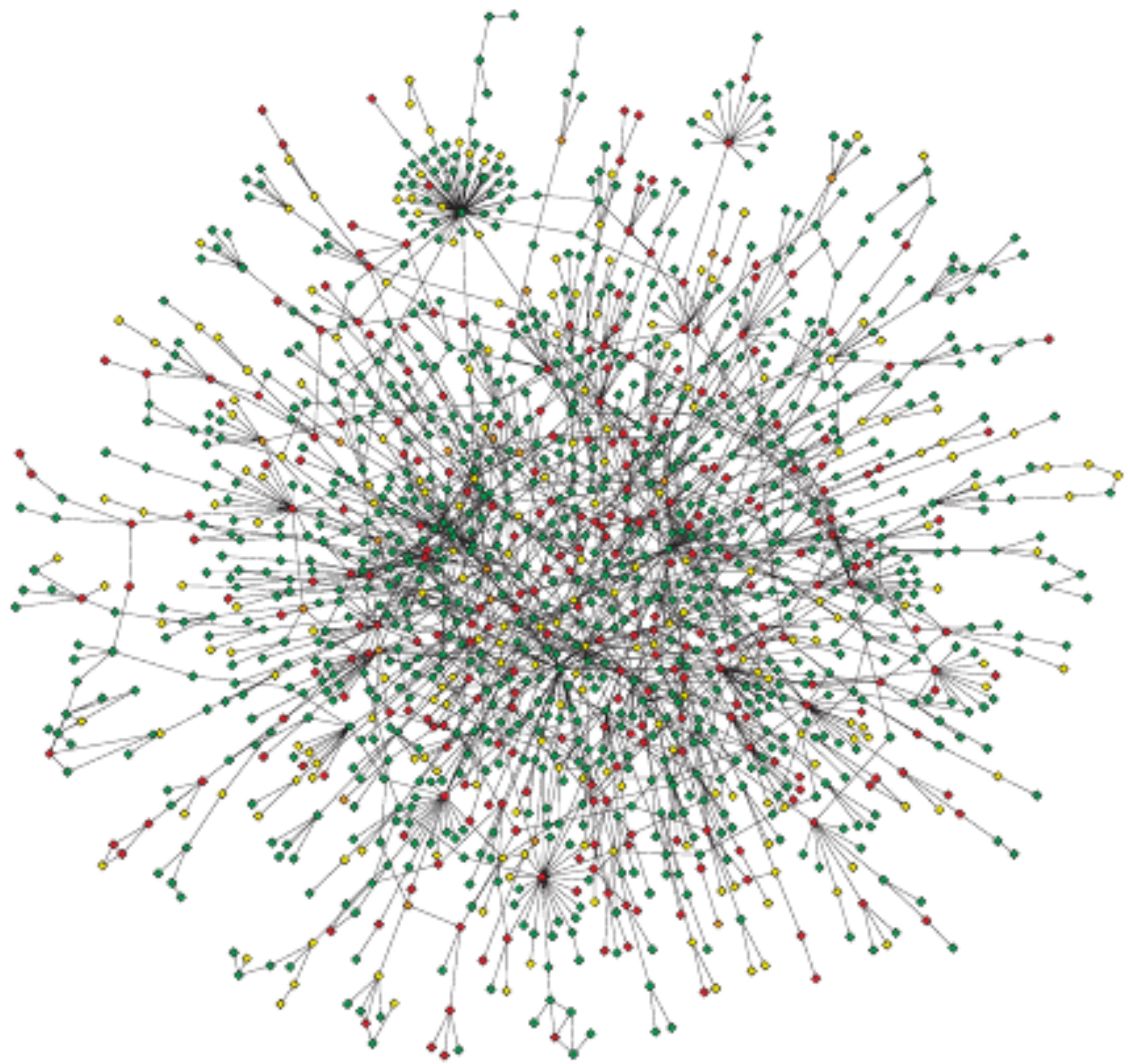
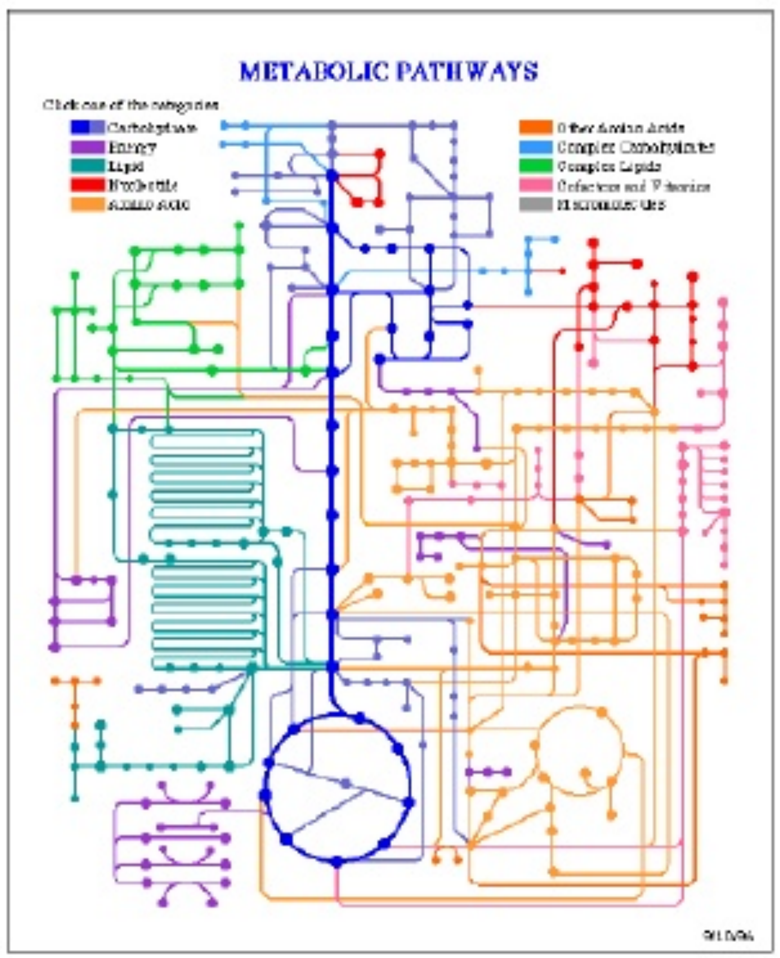


Figure 2 | **Yeast protein interaction network.** A map of protein–protein interactions[18] in *Saccharomyces cerevisiae*, which is based on early yeast two-hybrid measurements[23], illustrates that a few highly connected nodes (which are also known as hubs) hold the network together. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node indicates the phenotypic effect of removing the corresponding protein (red = lethal, green = non-lethal, orange = slow growth, yellow = unknown). Reproduced with permission from REF. 18 © Macmillan Magazines Ltd.

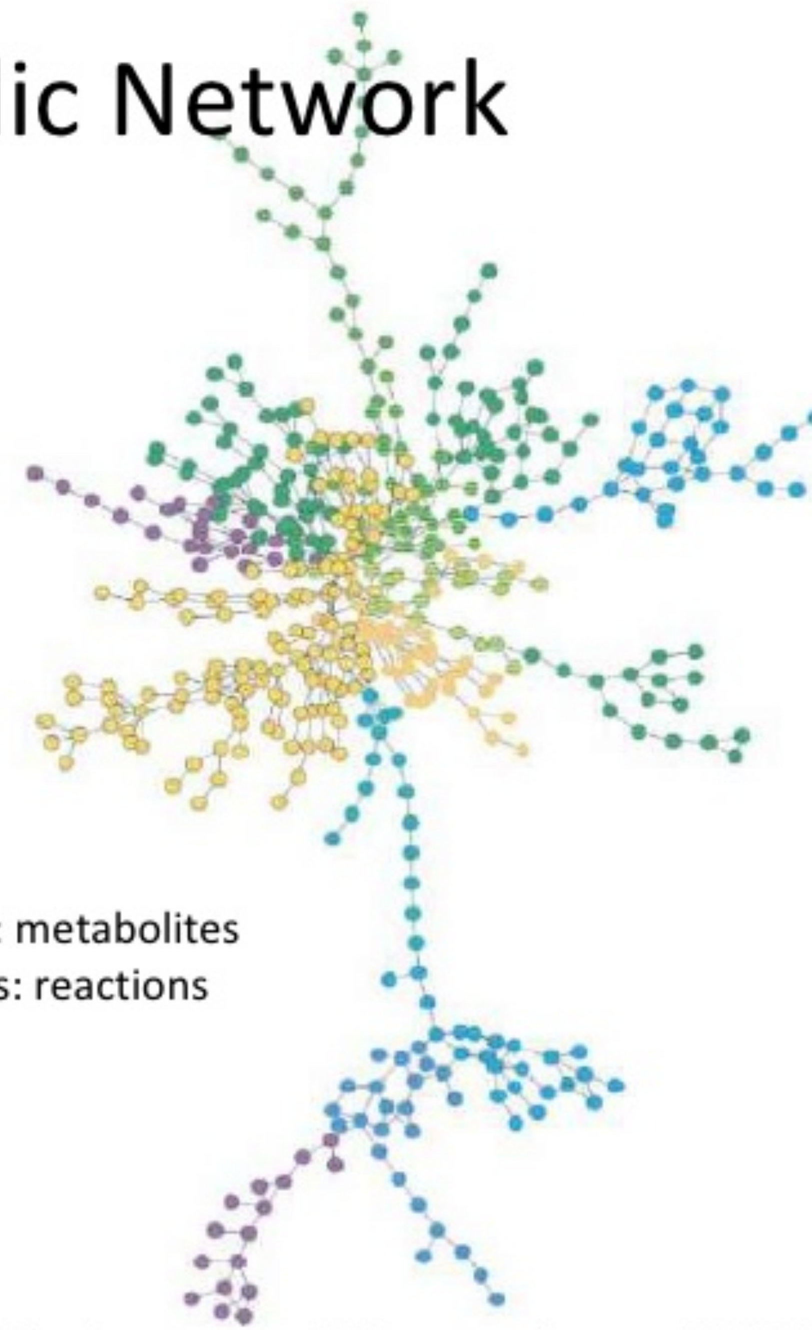
Metabolic Network

E. Coli Metabolic Network



Kegg, Wit, Biocyc, Bigg (UCSD)

Nodes: metabolites
Edges: reactions



Guimera and Nunes Amaral 2005

Brain Network

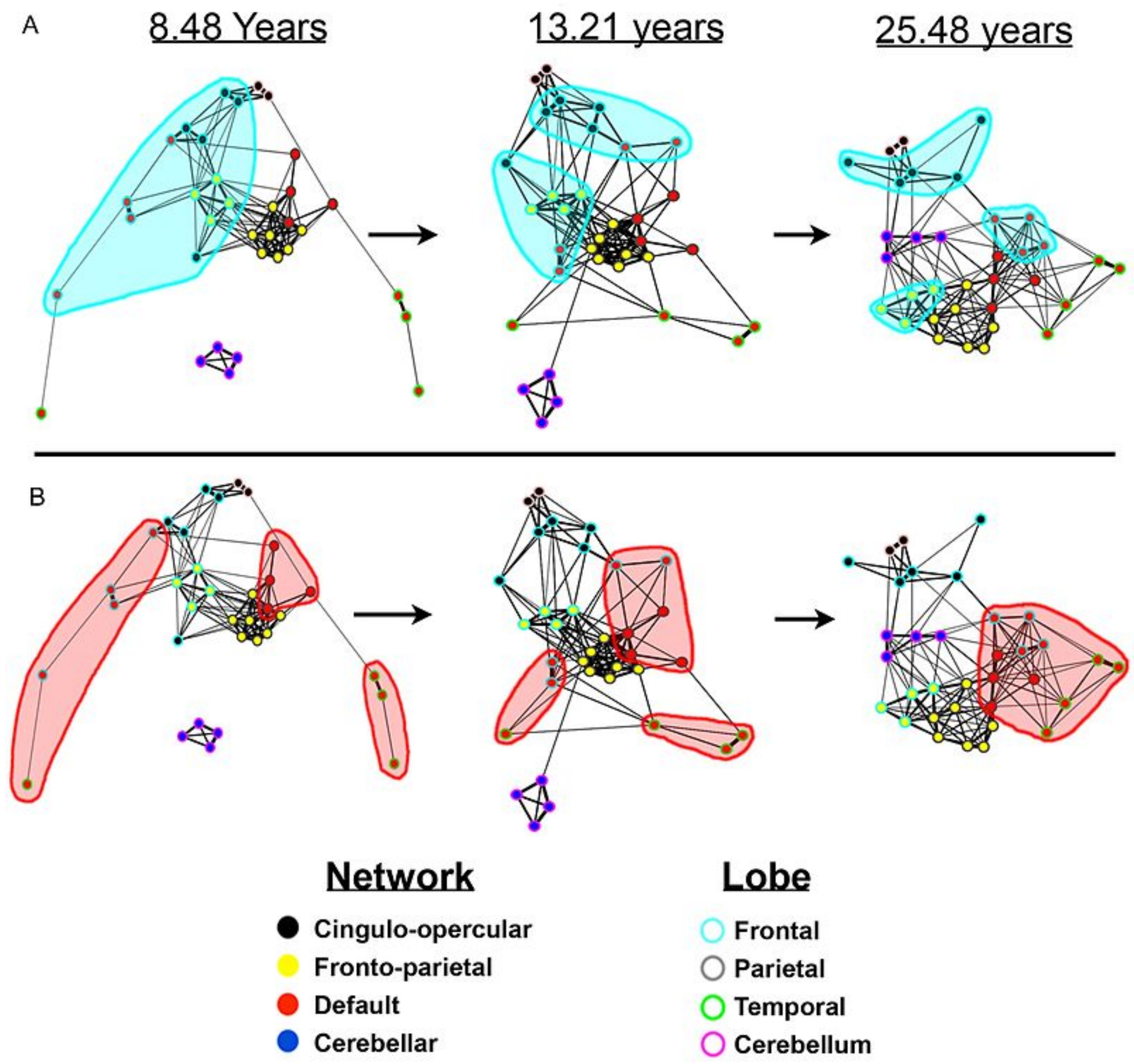
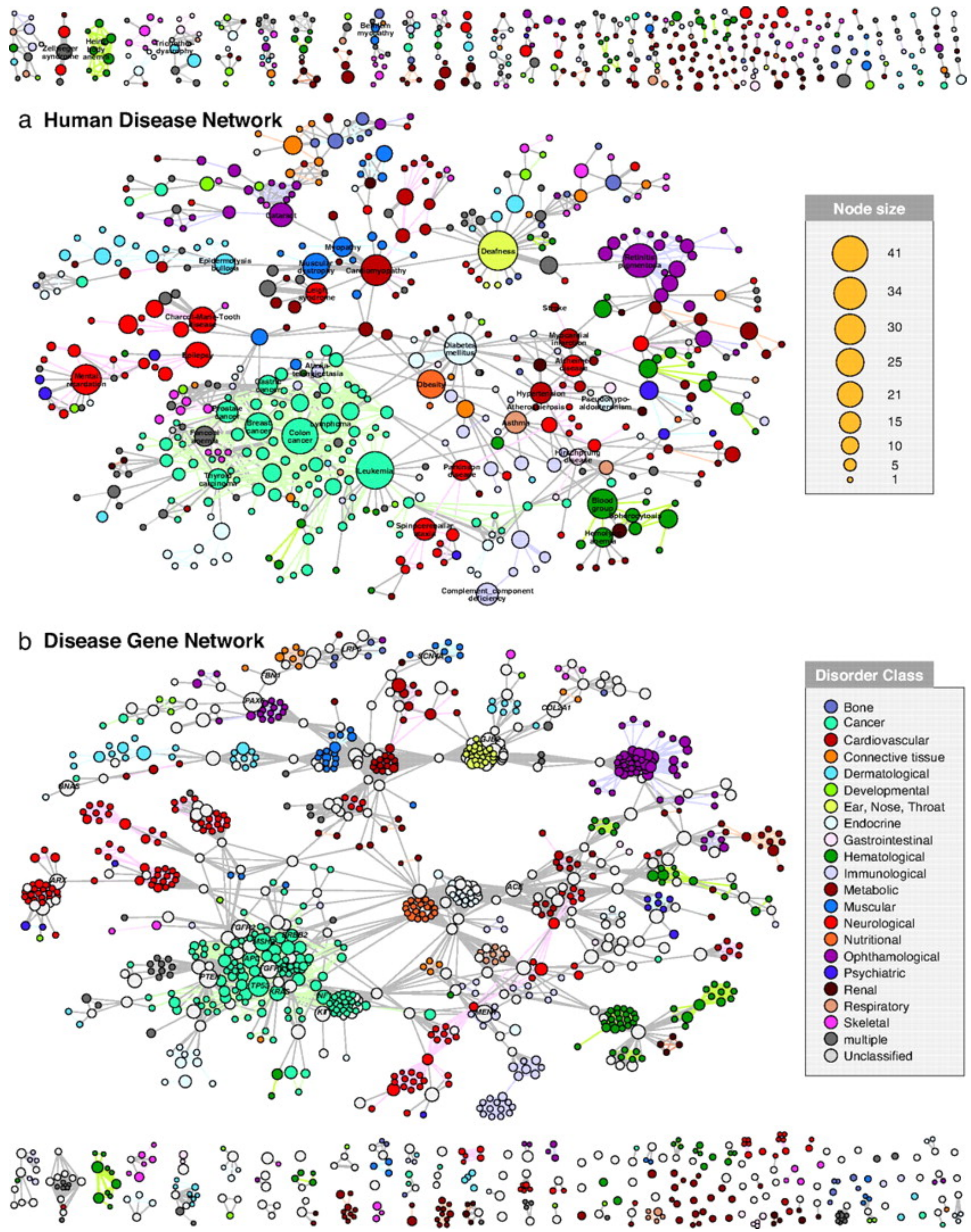


Figura: Fair, Damien A. *et al.*"Functional Brain Networks Develop from a 'Local to Distributed' Organization,PLoS Computational Biology 2009

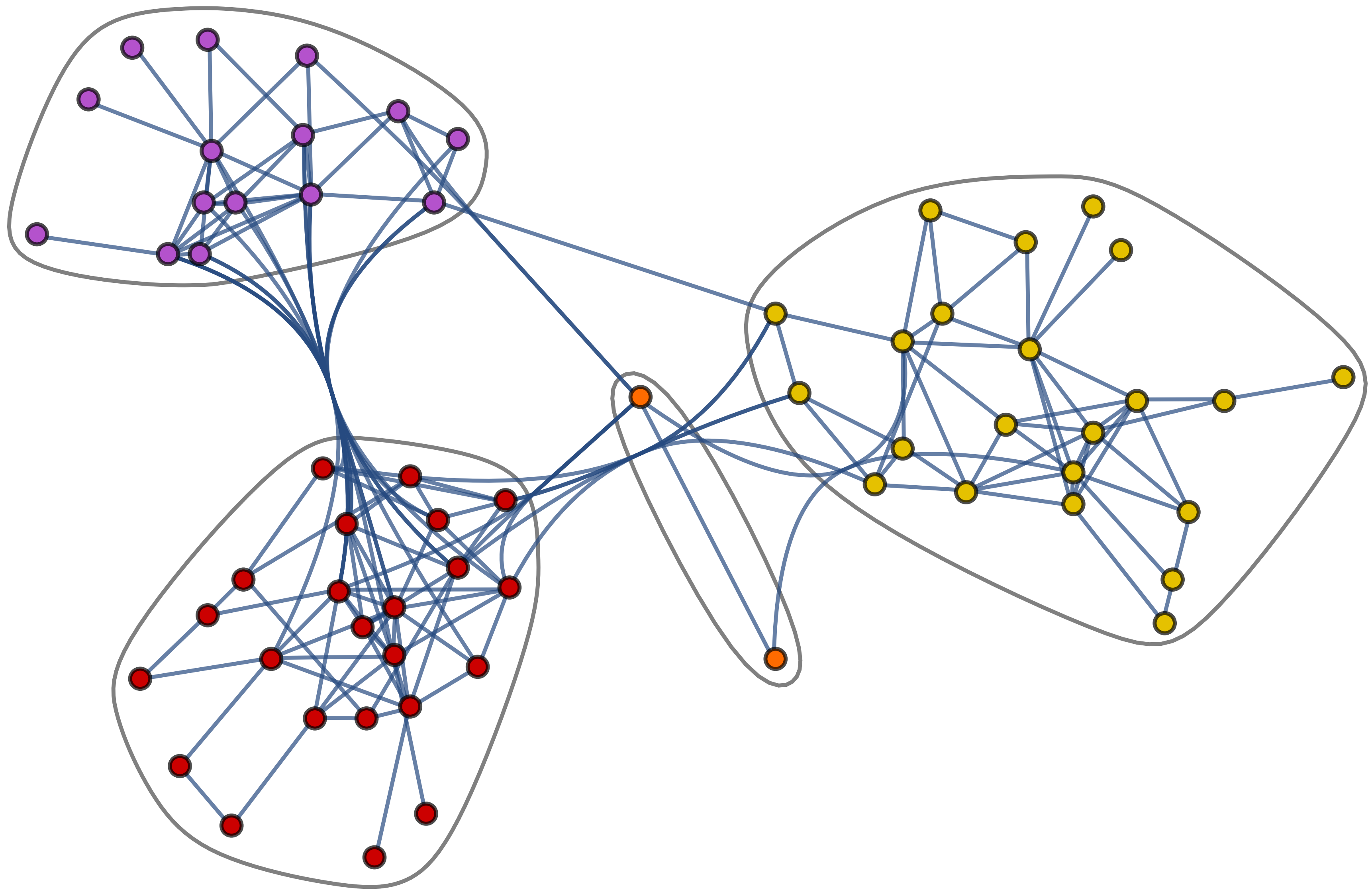
Disease Network



©2007 by National Academy of Sciences

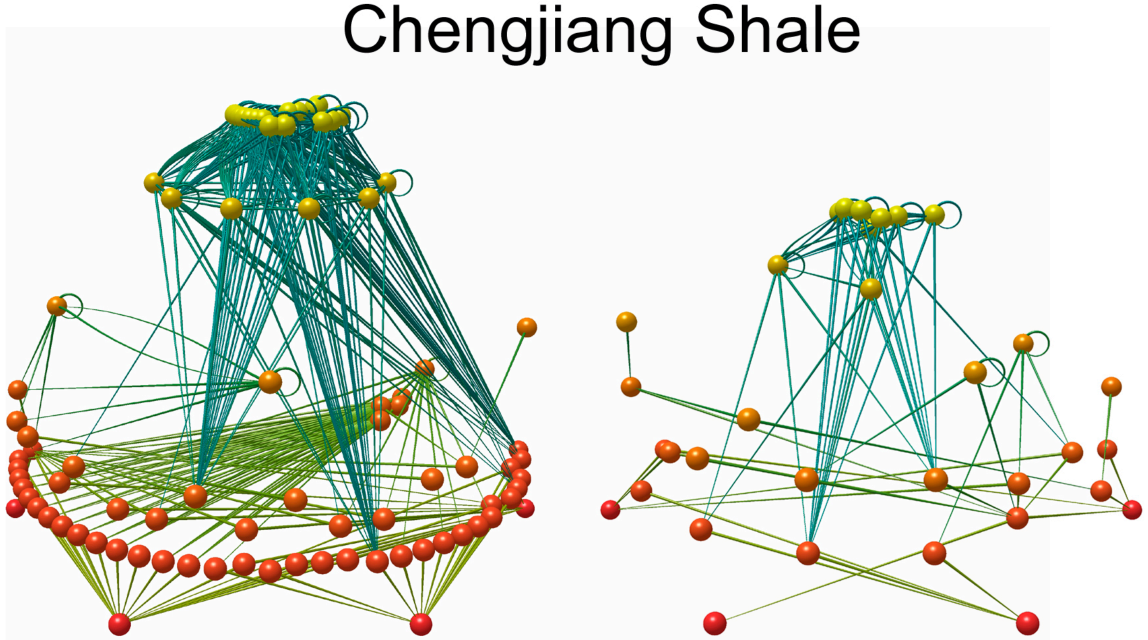
Kwang-Il Goh et al. PNAS 2007;104:8685-8690

Dolphin's Friendship Network



Food-web

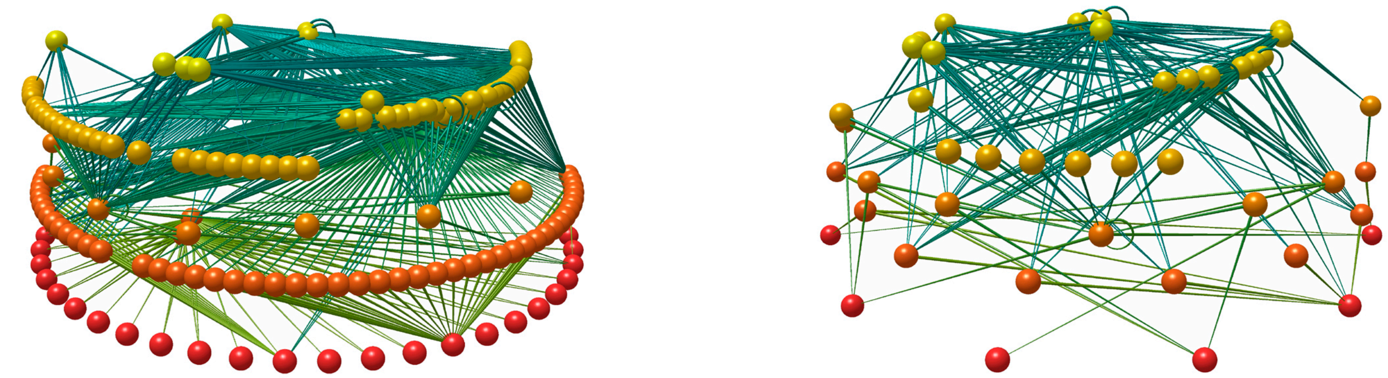
Chengjiang Shale



Original Species
S = 85, L = 559, C = 0.077
TL = 2.99, MaxTL = 5.15

Trophic Species
S = 33, L = 99, C = 0.091
TL = 2.84, MaxTL = 4.36

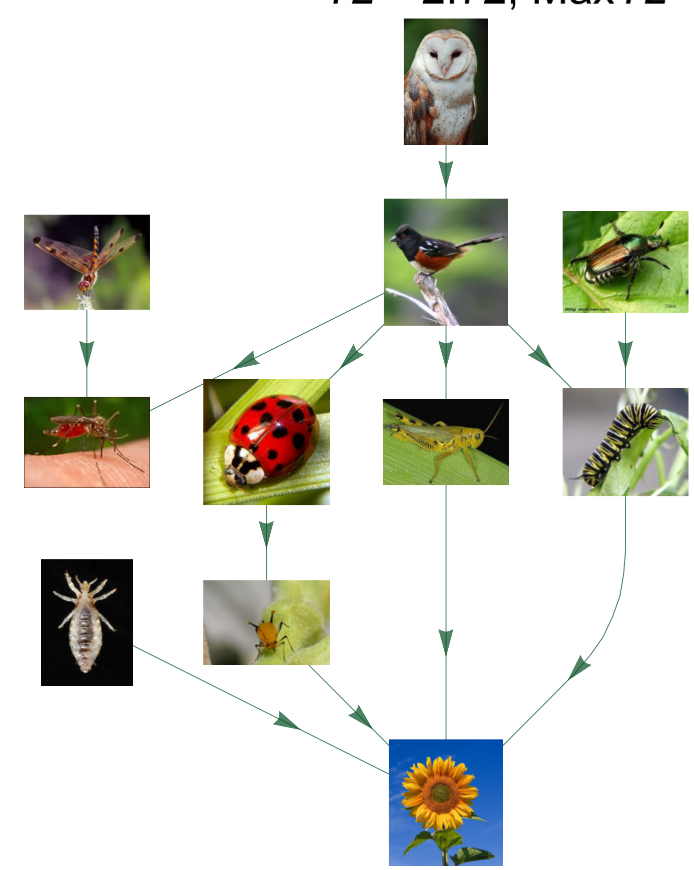
Burgess Shale



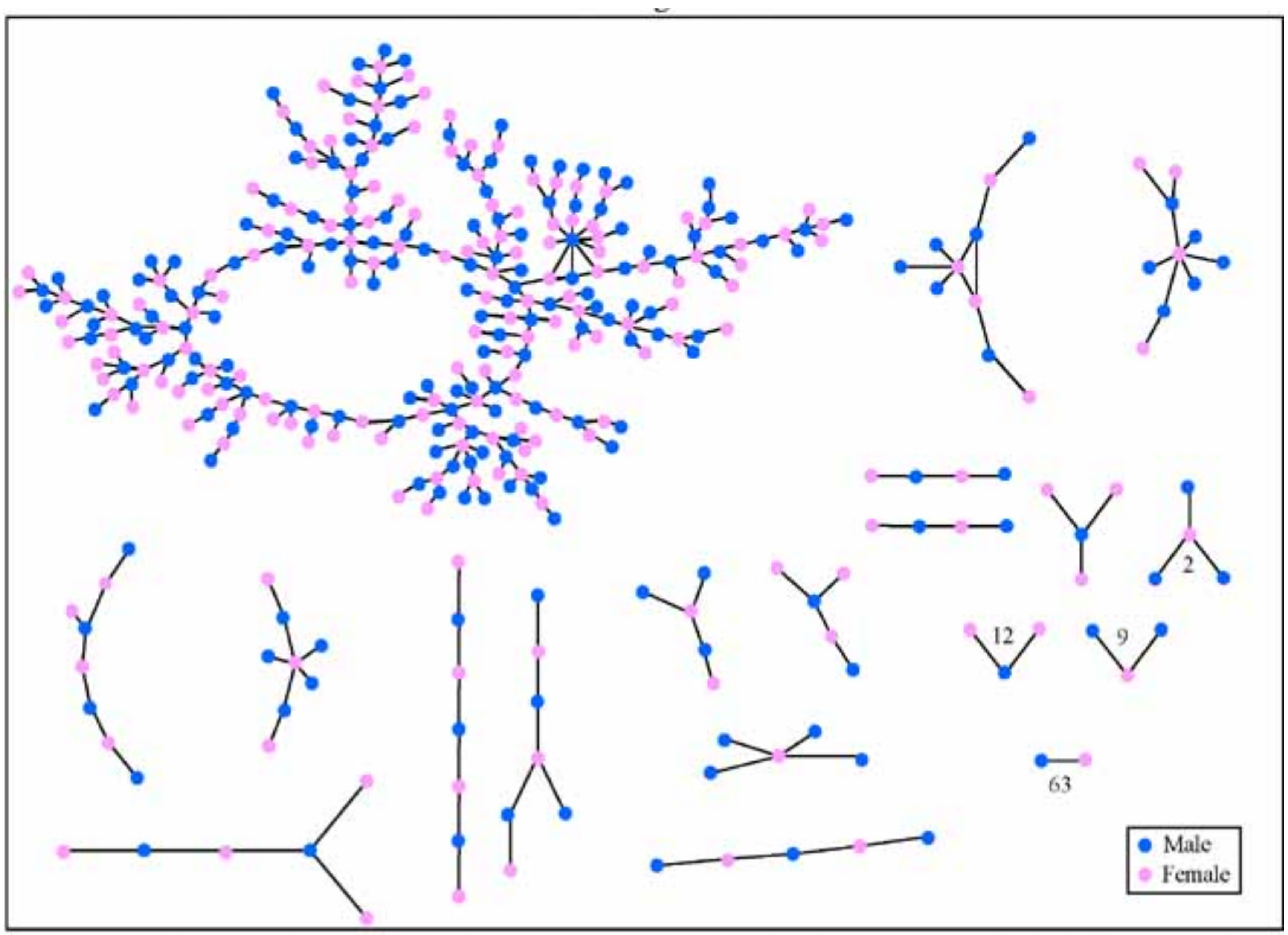
Original Species
S = 142, L = 771, C = 0.038
TL = 2.42, MaxTL = 3.67

Trophic Species
S = 48, L = 249, C = 0.108
TL = 2.72, MaxTL = 3.78

Compilation and Network Analyses of Cambrian Food Webs



Sexual Relationship Network



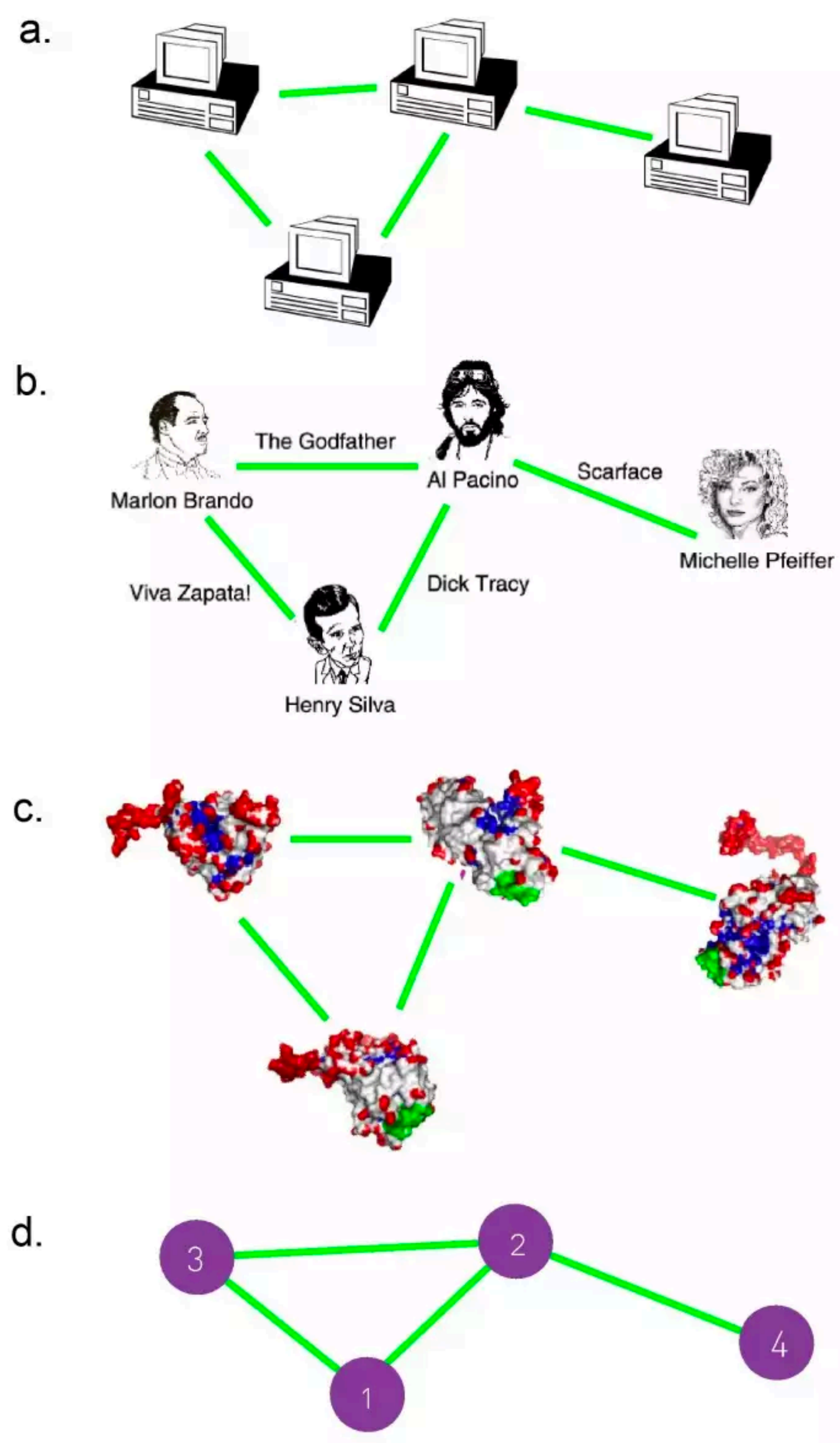
Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

Networks or Graphs?

Table 1

Network Science	Graph Teory
Network	Graph
Node	Vertex
Link	Edge

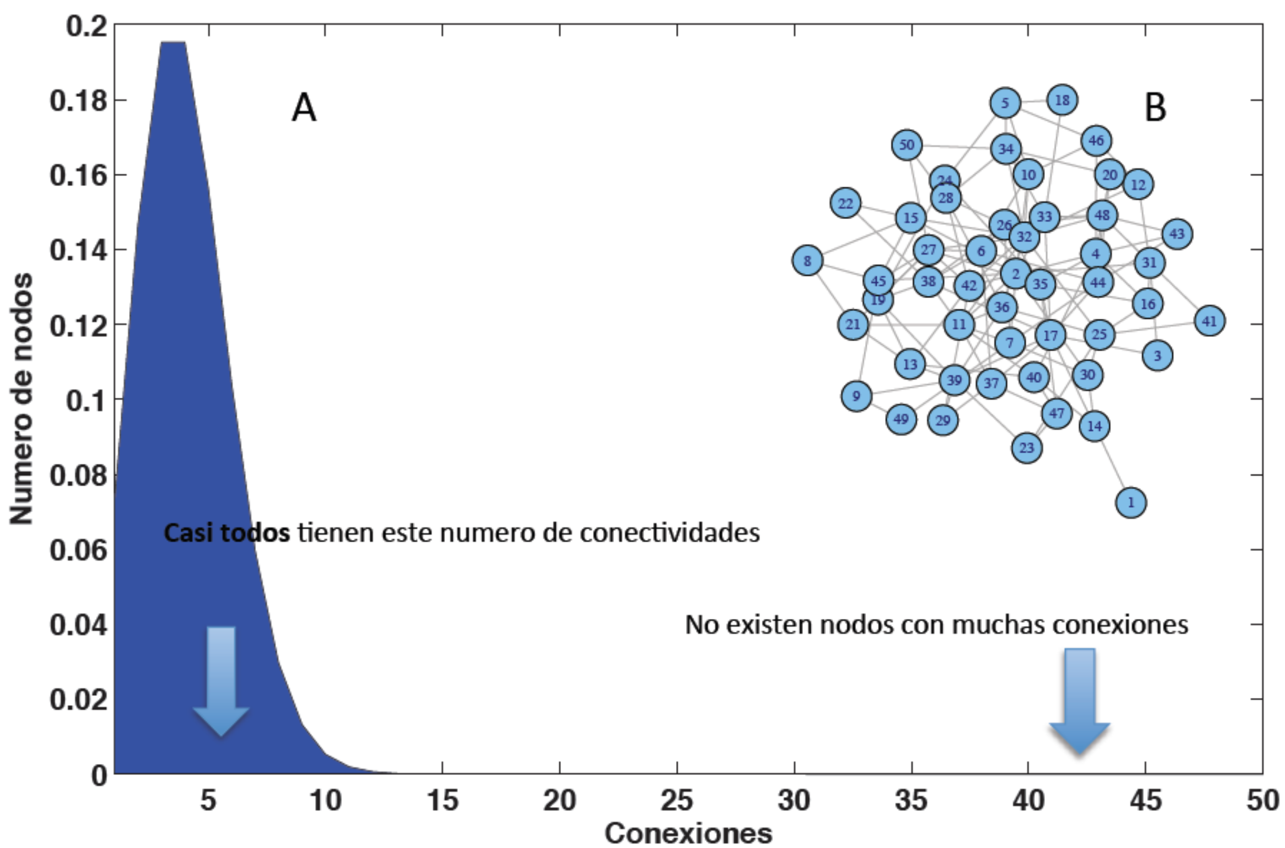
Different Networks, Same Graph

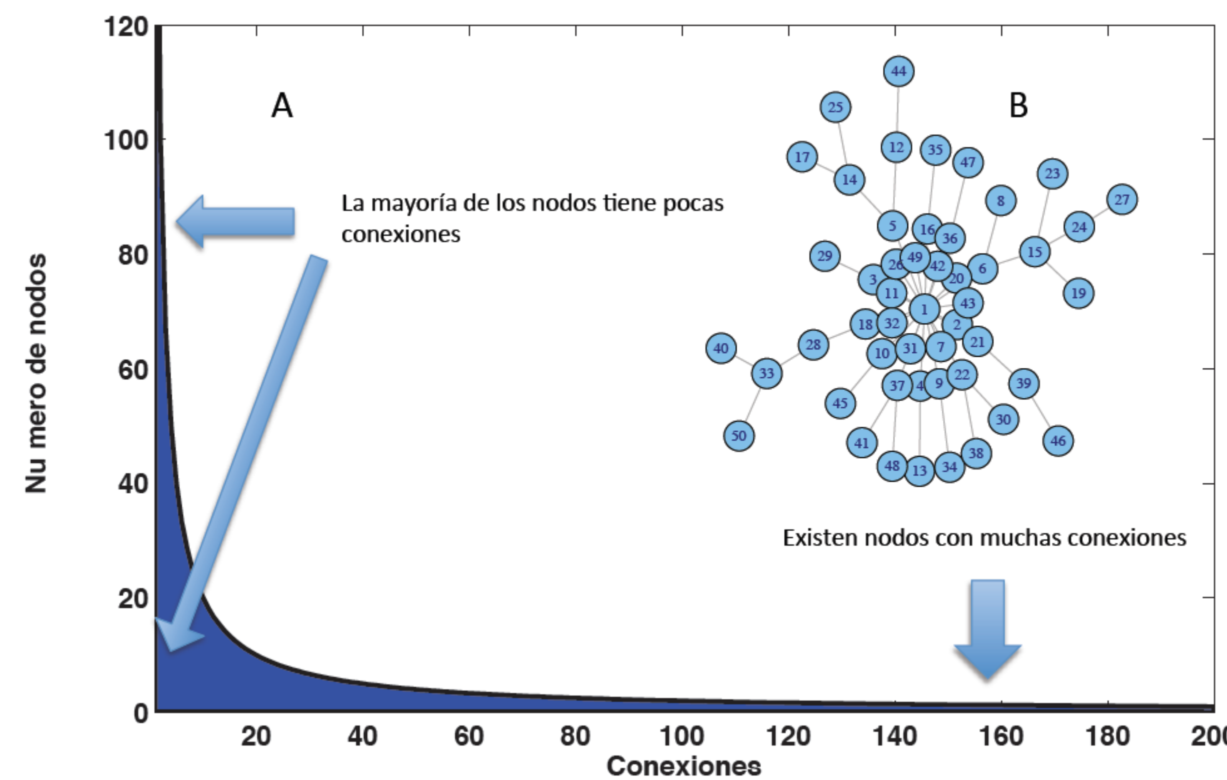


Network	Nodes	Links	Directed / Undirected	N	L	‹K›
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorships	Undirected	23,133	93,437	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Papers	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Canonical Network Maps
The basic characteristics of ten networks used throughout this book to illustrate the tools of network science. The table lists the nature of their nodes and links, indicating if links are directed or undirected, the number of nodes (N) and links (L), and the average degree for each network. For directed networks the average degree shown is the average in- or out-degrees $\langle k \rangle = \langle k_{in} \rangle = \langle k_{out} \rangle$

Degree, Average Degree and Degree Distribution





Brief Statistic Review

Average

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The *n*th moment

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

Brief Statistic Review

Standard deviation

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \langle x \rangle)^2}$$

Distribution of x

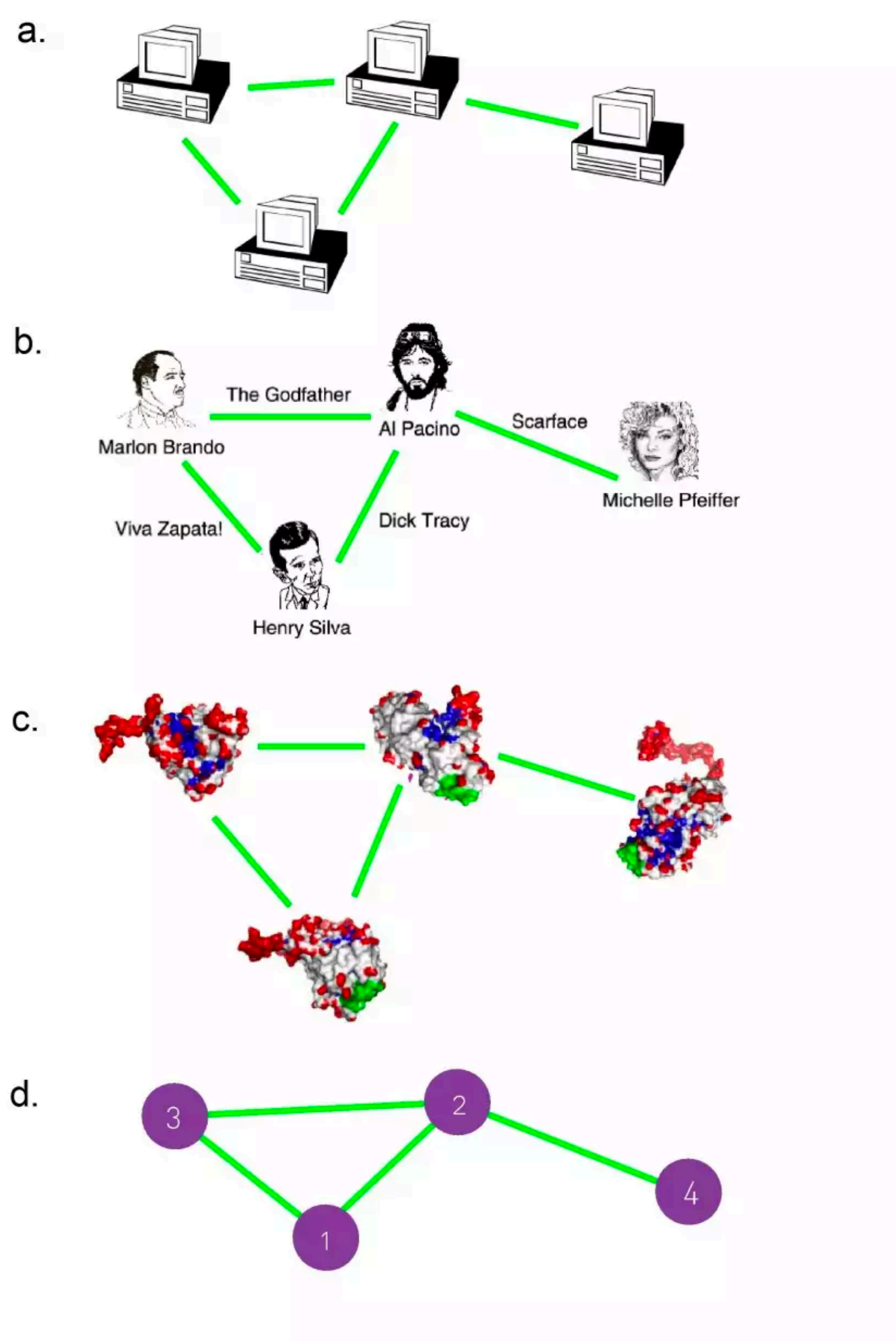
$$P_k = \frac{N_k}{N}$$

$$\sum_k^N P_k = 1$$

Degree

Definition: Let k_i denotes the degree of vertex i, which is the number of connections of that node.

Degree



$$L = \frac{1}{2}\sum_{i=1}^{N} k_i$$

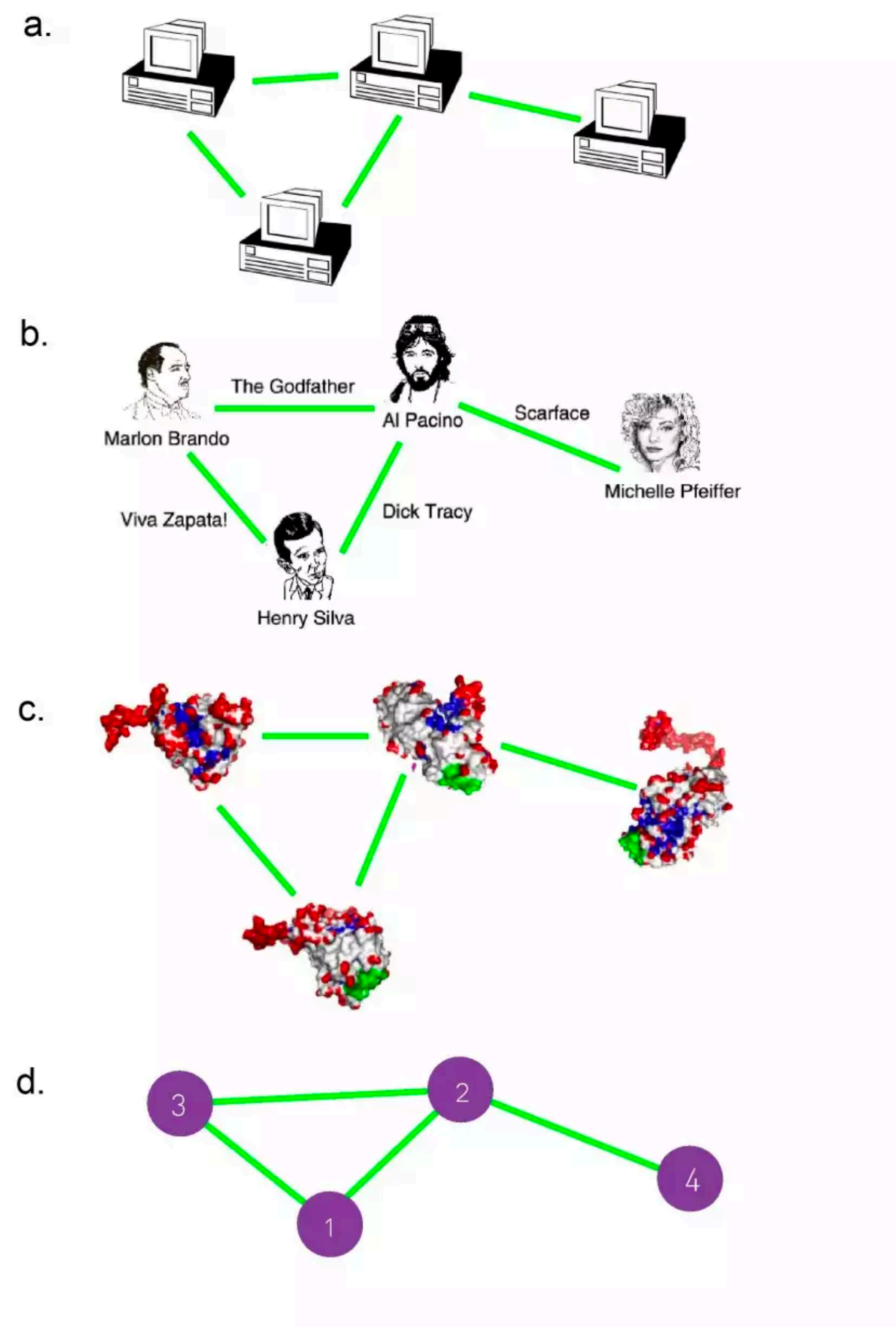
$$k_1 = 2$$

$$k_2 = 3$$

$$k_3 = 2$$

$$k_4 = 1$$

Average Degree

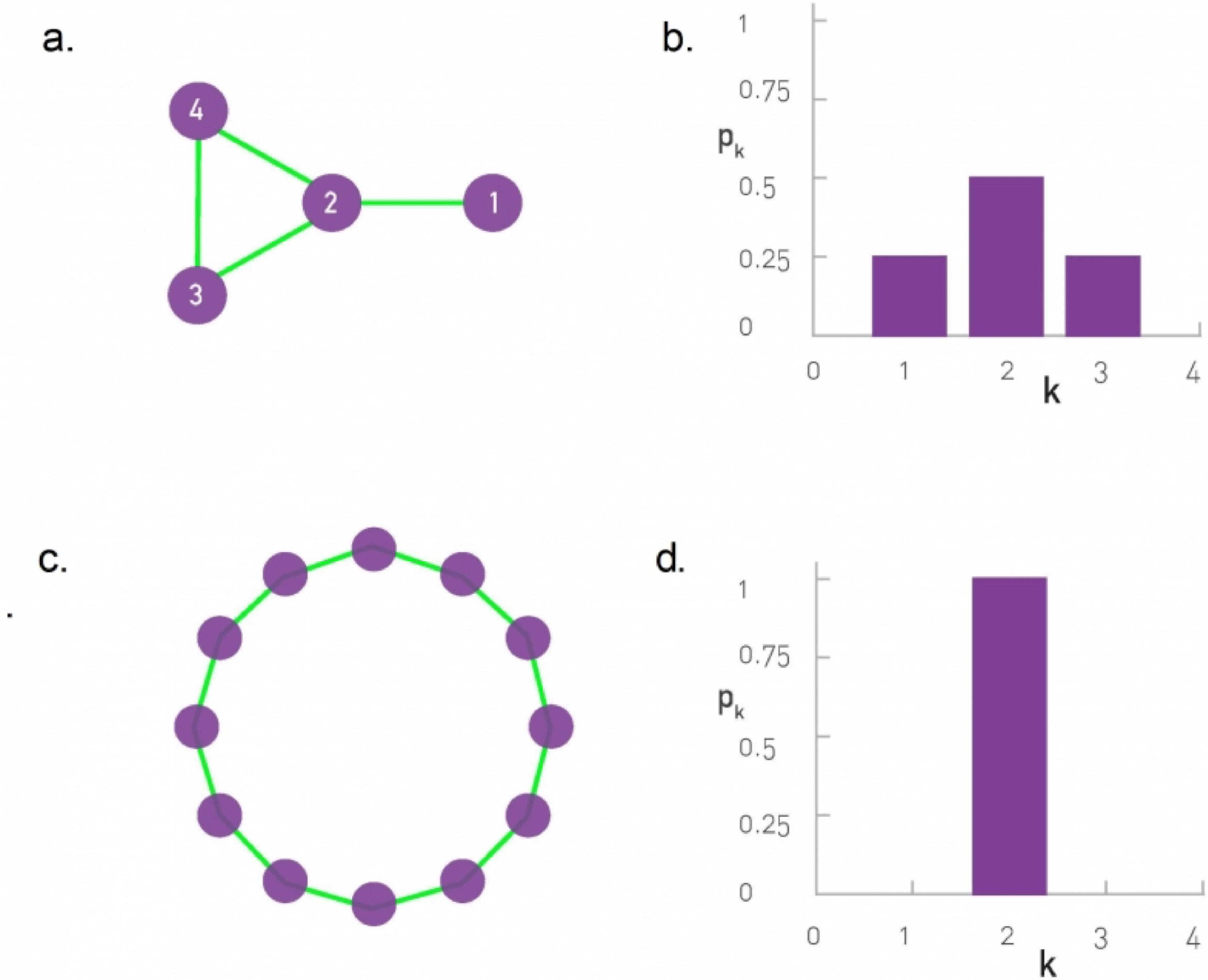


$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i$$

$$k_1 = 2$$
$$k_2 = 3$$
$$k_3 = 2$$
$$k_4 = 1$$

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}$$

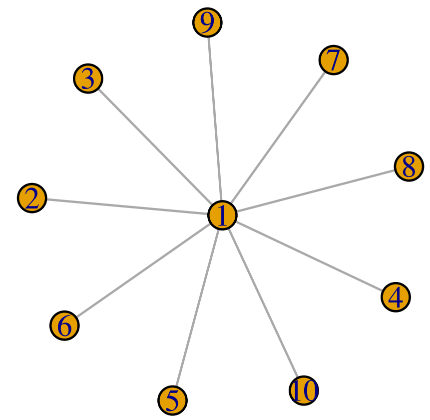
Degree Distribution

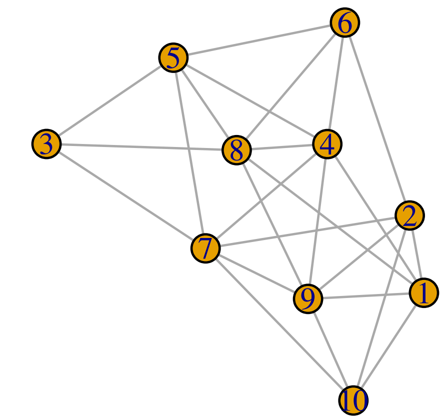


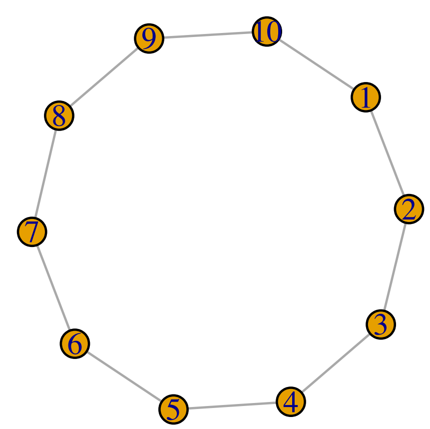
$$P_k = \frac{N_k}{N}$$

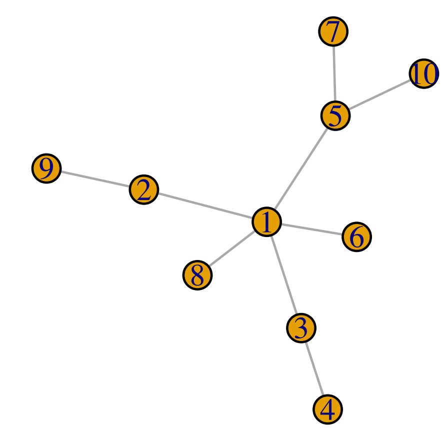
Exercise

Calculate the degree, average degree and degree distribution of the following networks

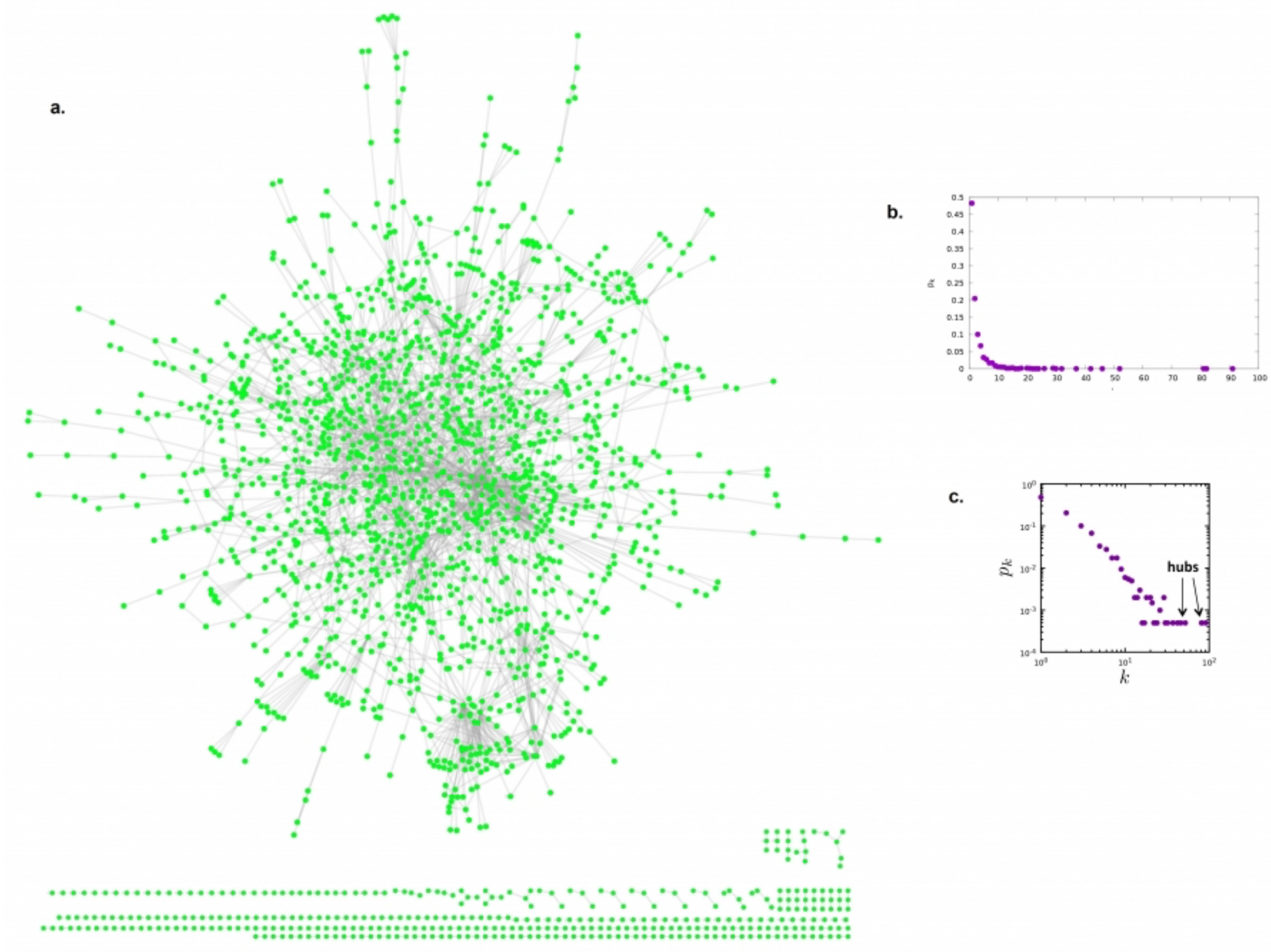






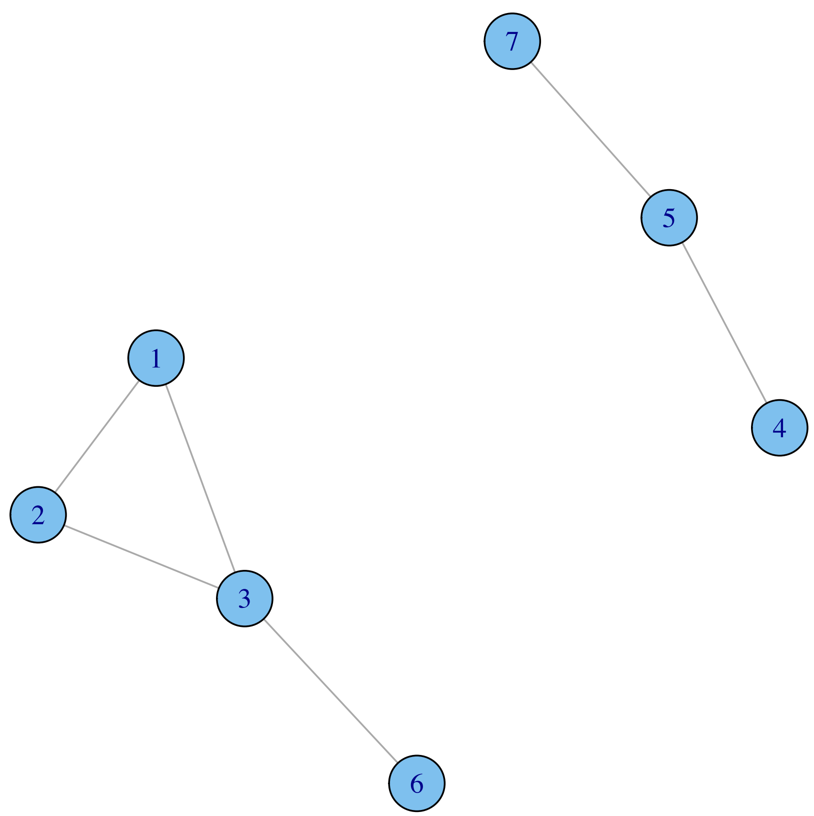


Degree Distributions of Real(Complex) Networks



Adjacency Matrix

$$M_{ij} = \begin{cases} 1 & \text{if node i is connected to node j} \\ 0 & \text{otherwise} \end{cases}$$



$$\begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

Undirected Network

$$M_{ij} = M_{ji}$$

Examples:

- Actor Networks(IMDB)
- Co-author Network
- Gene Co-expression Network
- PIP Network Protein-Protein Interaction

Directed Networks

Direction matters!

$$M_{ij} \neq M_{ji}$$

Examples:

- Twitter followers-following Network
- Citation Network
- Gene Regulatory Network (GRN)

Weighted Networks

Connection is not enough!

$$M_{ij} = w_{ji}$$

Examples:

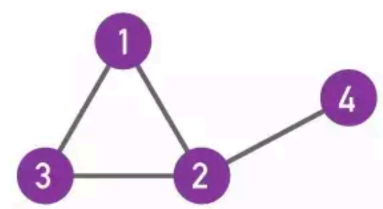
- Correlation Network
- Probabilistic Boolean Network

Adjacency Matrix

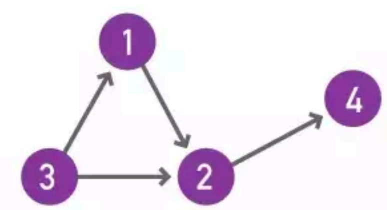
a. **Adjacency matrix**

$$A_{ij} = \begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$

b. **Undirected network**



c. **Directed network**



$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

$$A_{ij} = \begin{matrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

$$k_2 = \sum_{j=1}^{4} A_{2j} = \sum_{i=1}^{4} A_{i2} = 3$$

$$k_2^{\text{in}} = \sum_{j=1}^{4} A_{2j} = 2,\ k_2^{\text{out}} = \sum_{i=1}^{4} A_{i2} = 1$$

$$A_{ij} = A_{ji} \qquad A_{ii} = 0$$

$$A_{ij} \neq A_{ji} \qquad A_{ii} = 0$$

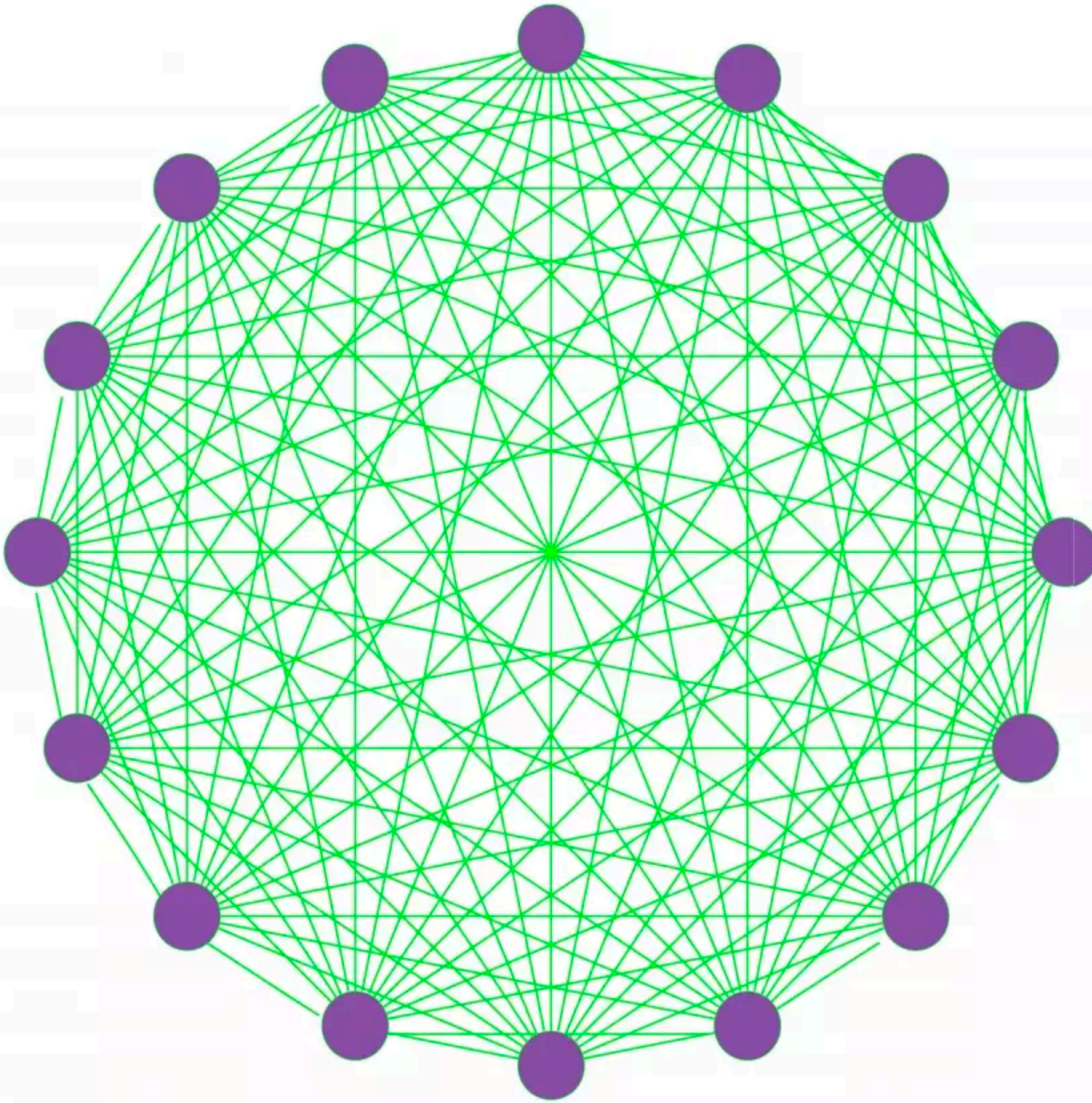
$$L = \frac{1}{2} \sum_{i\; i=1}^{N} A_{ij}$$

$$L = \sum_{i,j=1}^{N} A_{ij}$$

$$\langle k \rangle = \frac{2L}{N}$$

$$\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle = \frac{L}{N}$$

Real Networks are Sparse



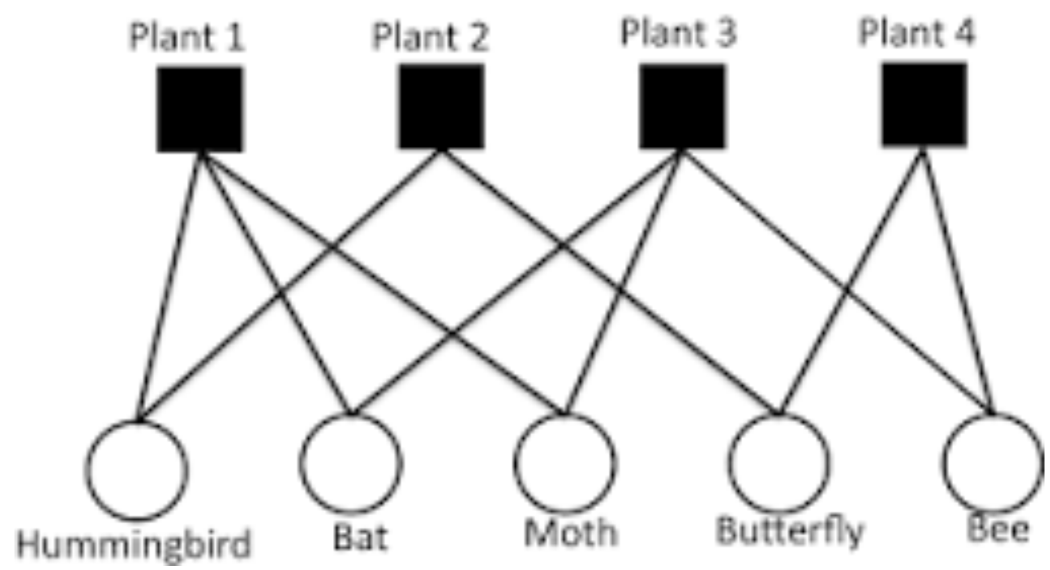
Complete connected network

$$L = N(N-1)/2$$

N: # nodes

Bipartite Networks

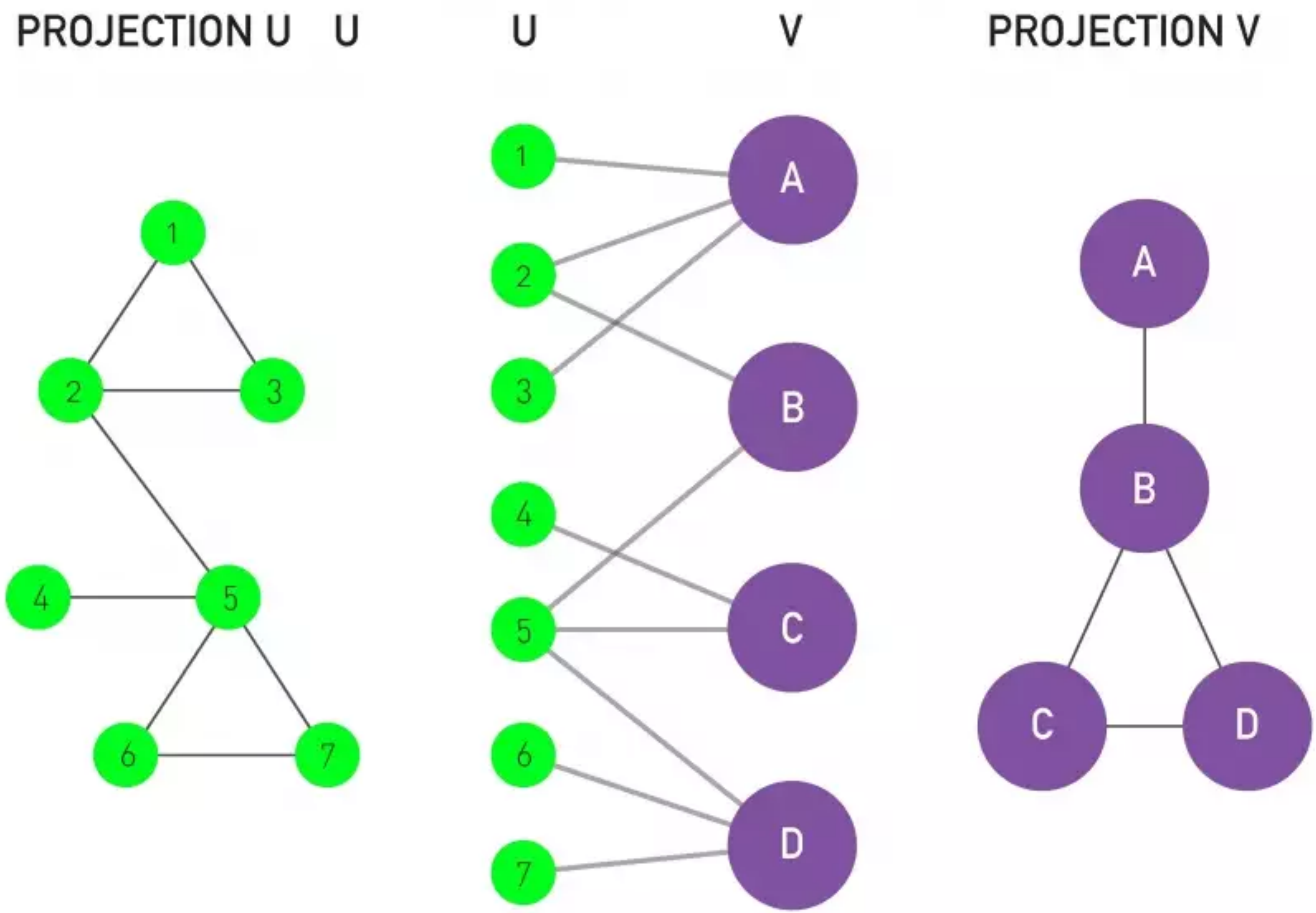
Two classes of nodes/vertex



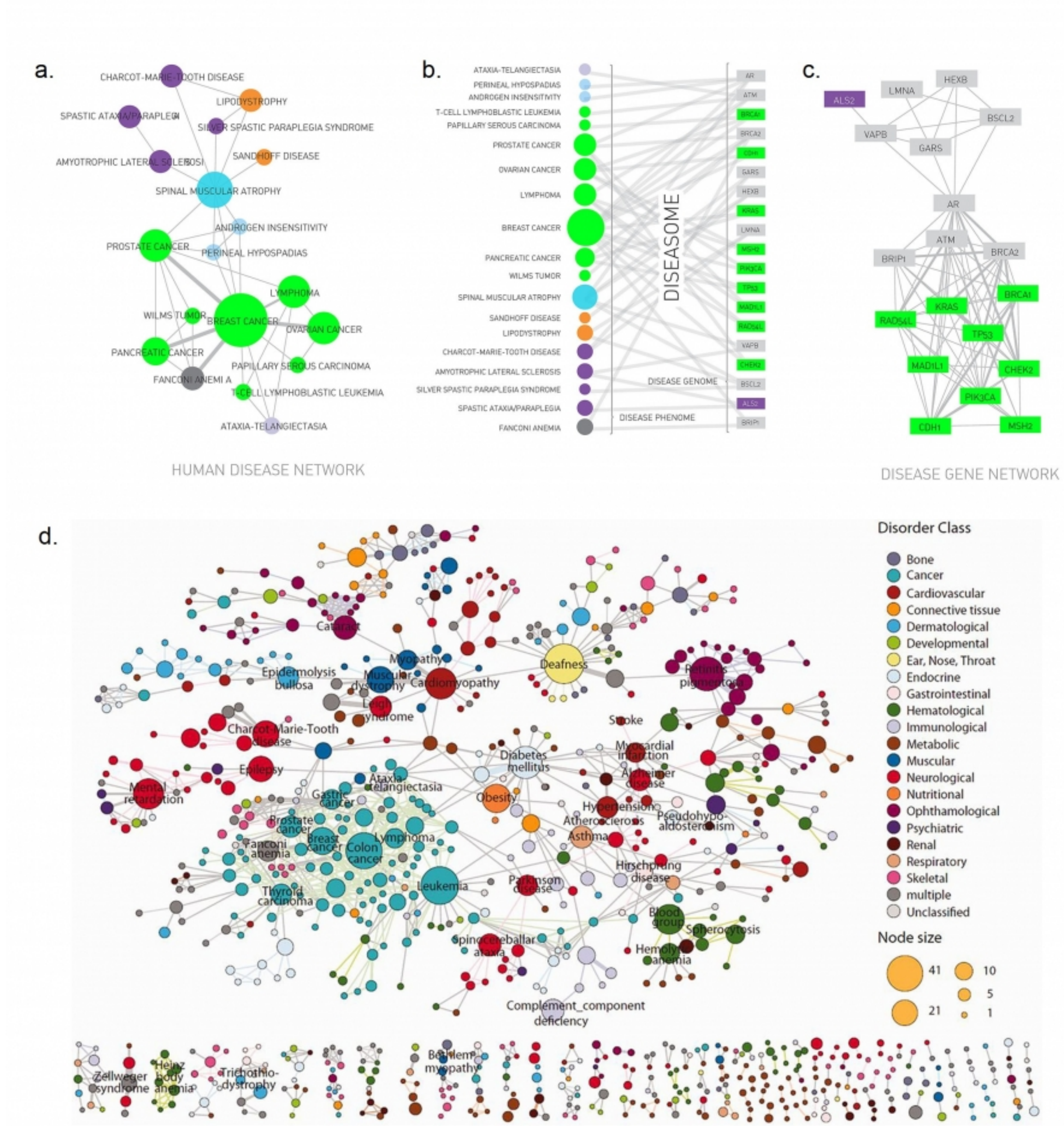
Examples:

- microRNA-genes Regulatory Network
- Disease Network

Bipartite Networks



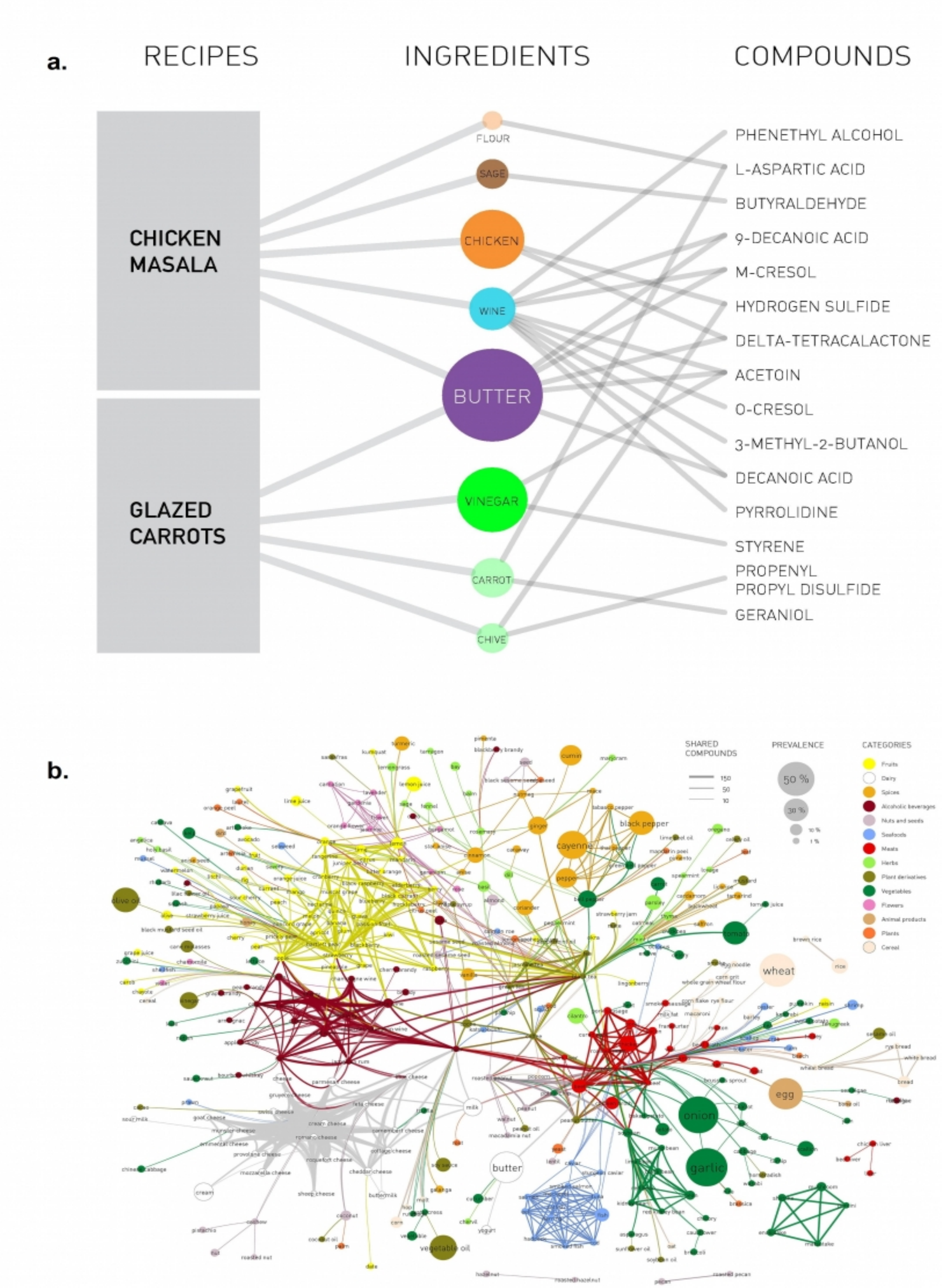
Human Disease Network



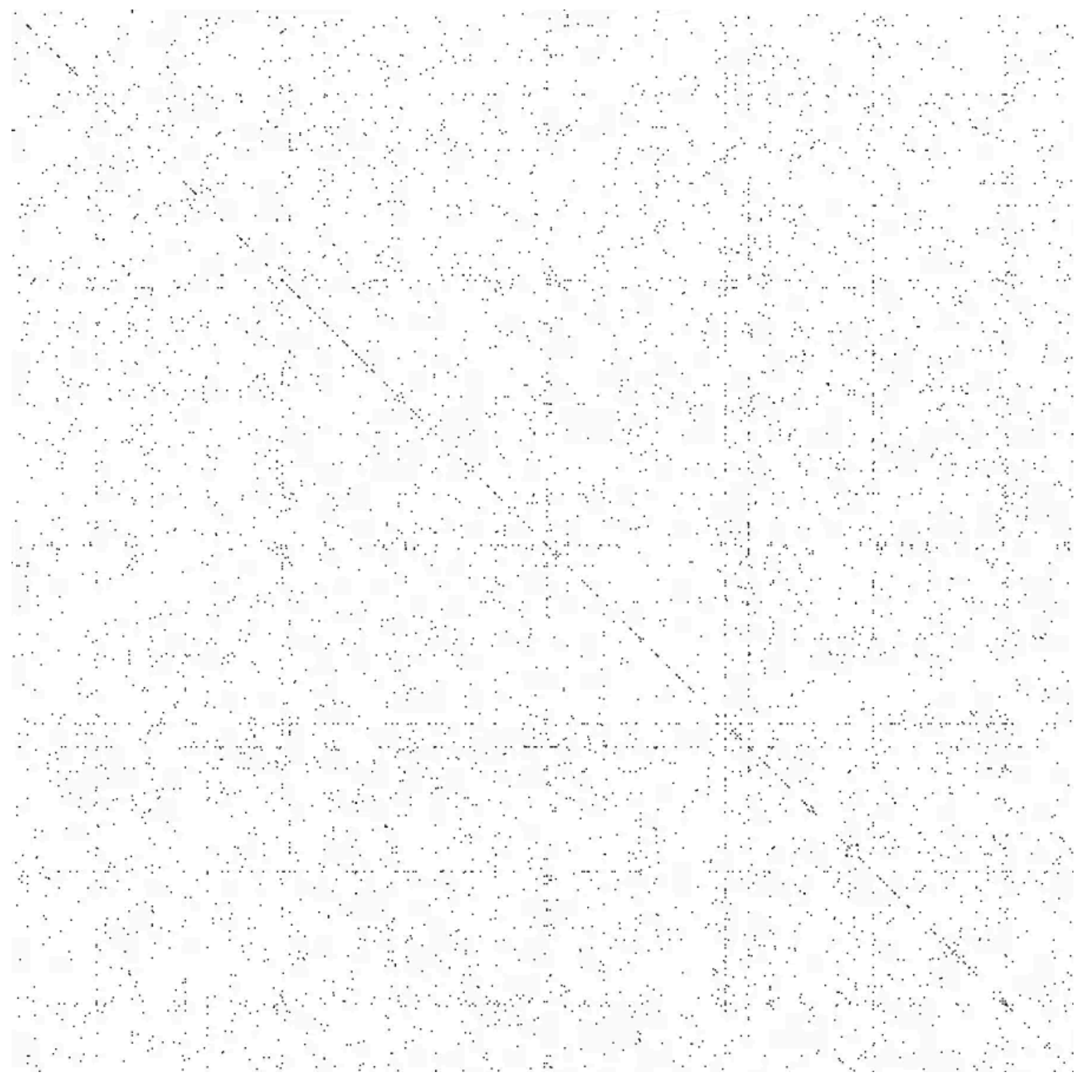
Human Disease Network

Interactive Disease NYT

Tripartite Network



The Adjacency Matrix is Sparse

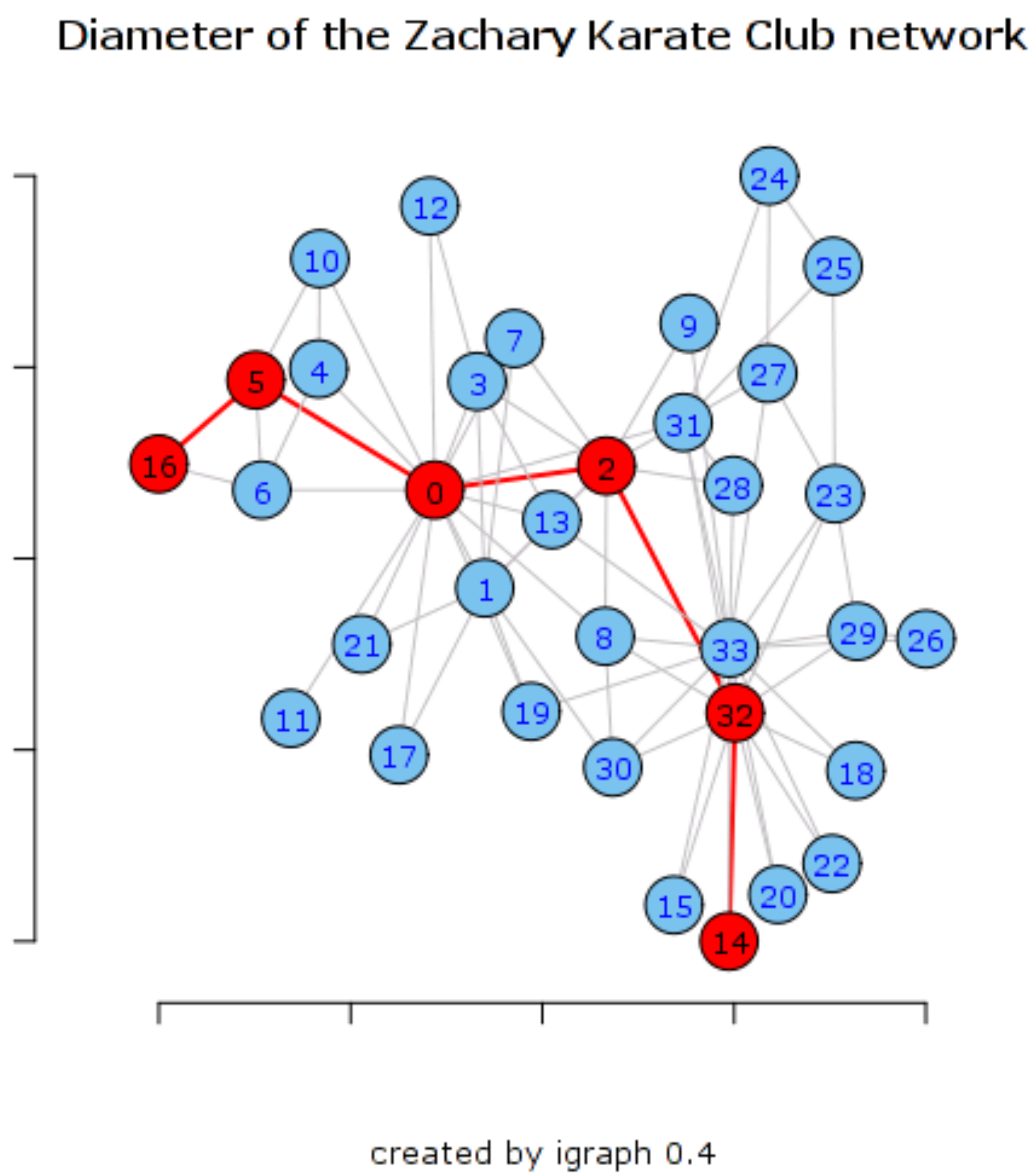


Yeast protein-protein interaction network. 2018 proteins

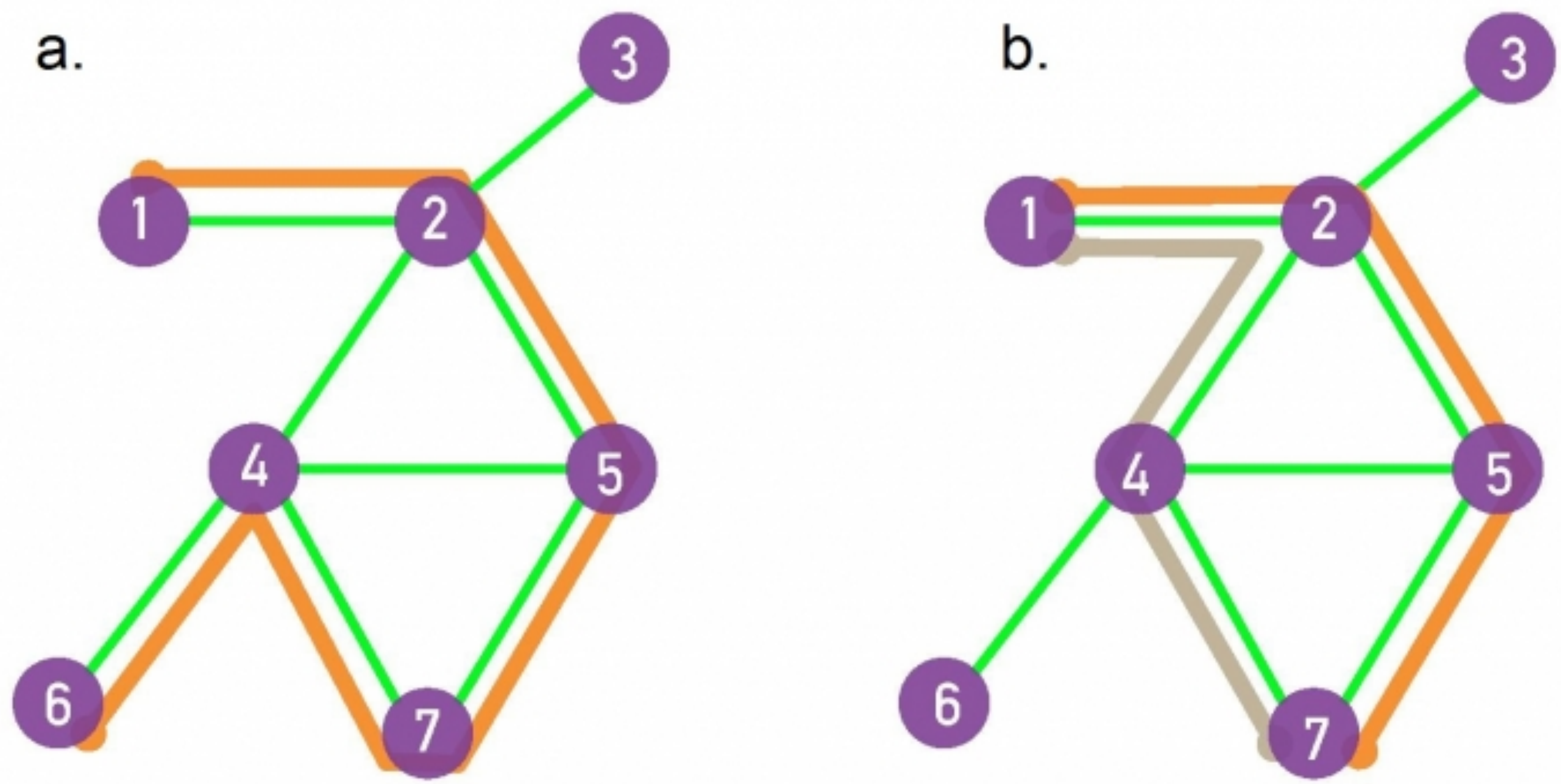
Exercises 1

Graph Theory: Basic Concepts

Paths and Distances

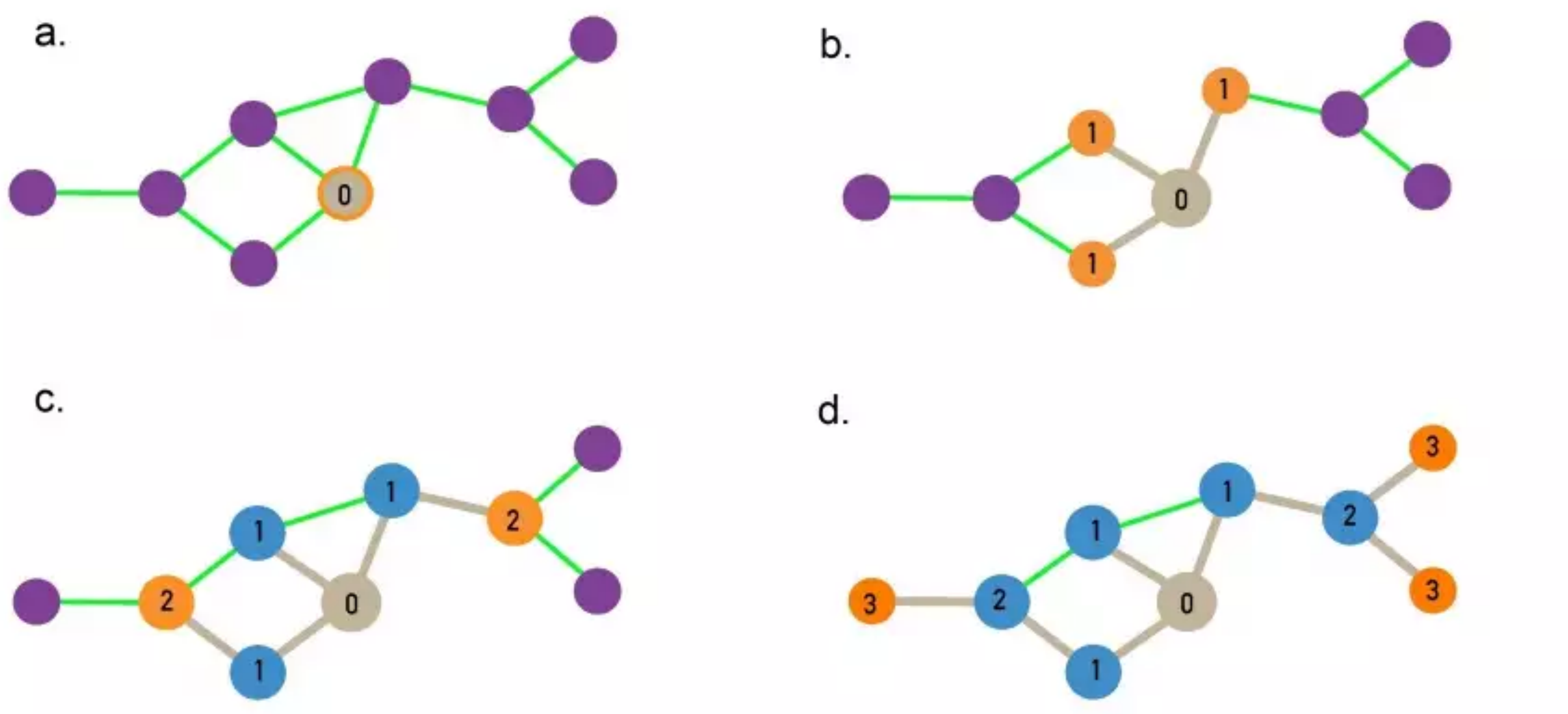


Paths in Networks

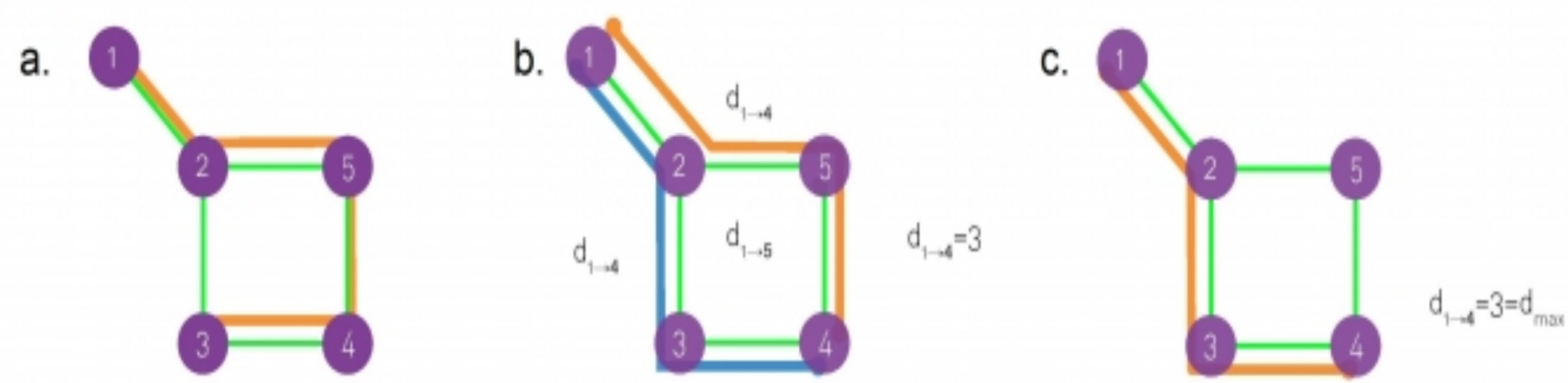


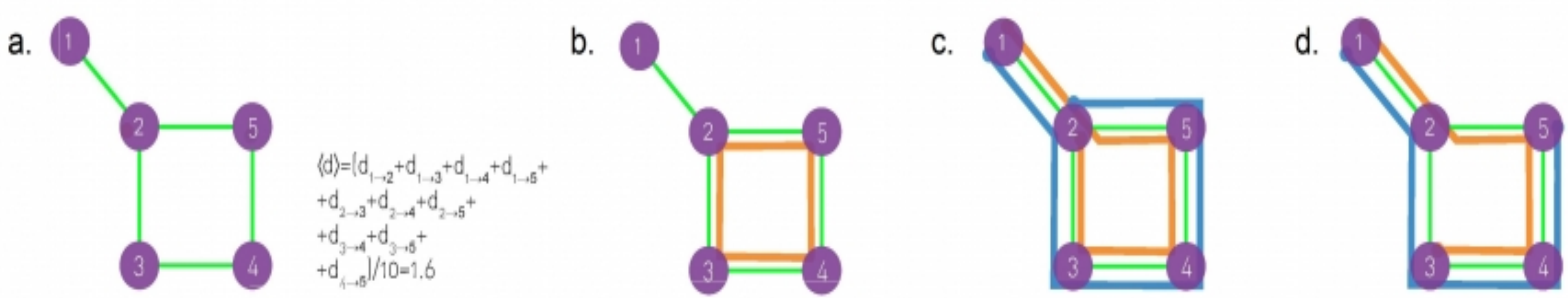
1. A path between nodes i_0 and i_n is an ordered list of n links P = {(i_0, i_1), (i_1, i_2), (i_2, i_3), ... ,(i_{n-1}, i_n)}. The length of this path is n. The path shown in orange in (a) follows the route 1→2→5→7→4→6, hence its length is n = 5.
2. The shortest paths between nodes 1 and 7, or the distance d_{17}, correspond to the path with the fewest number of links that connect nodes 1 to 7. There can be multiple paths of the same length, as illustrated by the two paths shown in orange and grey. The network diameter is the largest distance in the network, being d_{max} = 3 here.

Shortest paths (Distances)



Distances and average distances





Number of Shortest Paths Between Two Nodes

The number of shortest paths, N_{ij}, and the distance d_{ij} between nodes *i* and *j* can be calculated directly from the adjacency matrix A_{ij}.

- d_{ij} = 1: If there is a direct link between *i* and *j*, then $A_{ij} = 1$ ($A_{ij} = 0$ otherwise).

- d_{ij} = 2: If there is a path of length two between *i* and *j*, then $A_{ik} A_{kj} = 1$ ($A_{ik} A_{kj} = 0$ otherwise). The number of $d_{ij} = 2$ paths between *i* and *j* is

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik} A_{kj} = A_{ij}^{2}$$

- d_{ij} = d: If there is a path of length *d* between *i* and *j*, then $A_{ik} \ldots A_{lj} = 1$ ($A_{ik} \ldots A_{lj} = 0$ otherwise). The number of paths of length *d* between *i* and *j* is *N*(*d*)*ij*=*Adij*

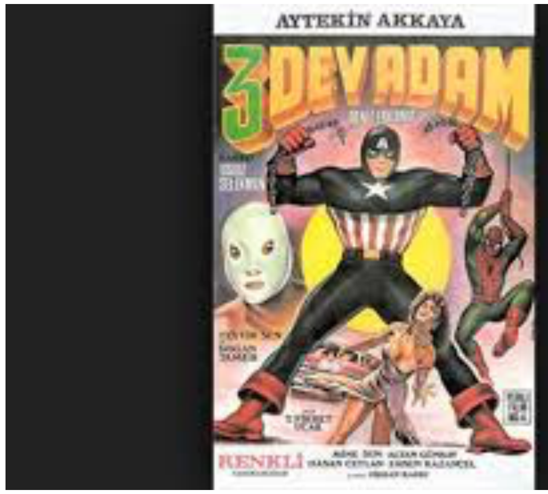
These equations hold for directed and undirected networks. The *distance* between nodes *i* and *j* is the path with the smallest *d* for which $N_{ij}^{(d)} > 0$.

Network Diameter

Definition: The *diameter* of a network, denoted by d_{max}, is the maximum shortest path in the network. In other words, it is the largest distance recorded between *any* pair of nodes.

Average Path Length

Definition: The *average path length*, denoted by $\langle d \rangle$, is the average distance between all pairs of nodes in the network.



Small-world

Google

bacon number el santo

Web Images More Search tools

About 54,000,000 results (0.75 seconds)

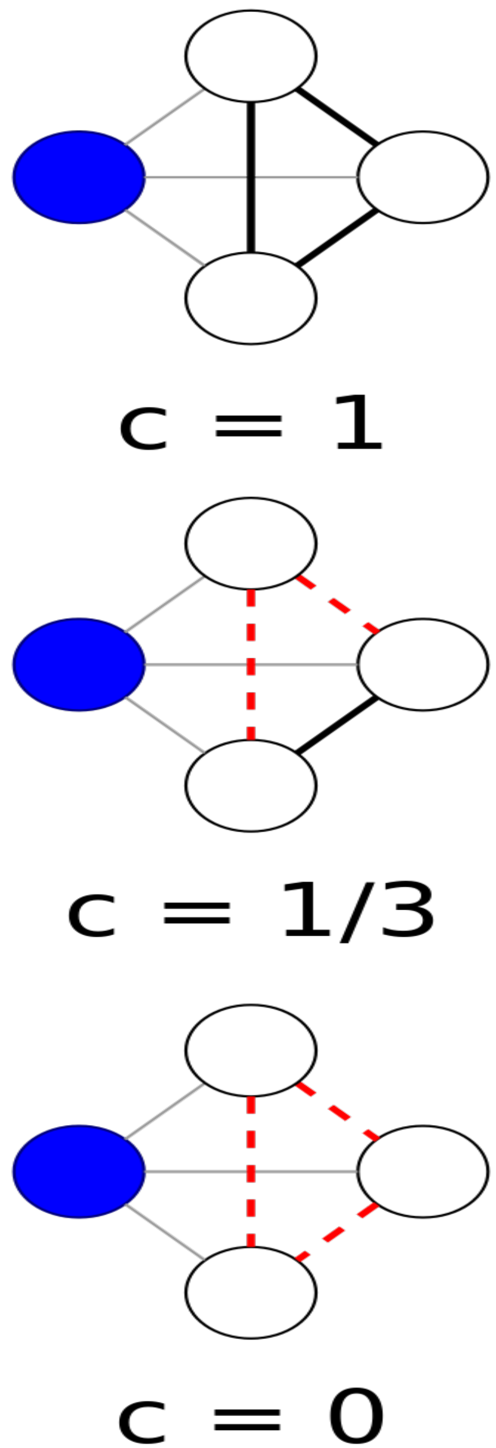
Santo's Bacon number is 3

Santo and Armando Silvestre appeared in Santo Contra los Zombis.

Armando Silvestre and Glenn Ford appeared in Rage.

Glenn Ford and Kevin Bacon appeared in The Gift.

Clustering coefficient



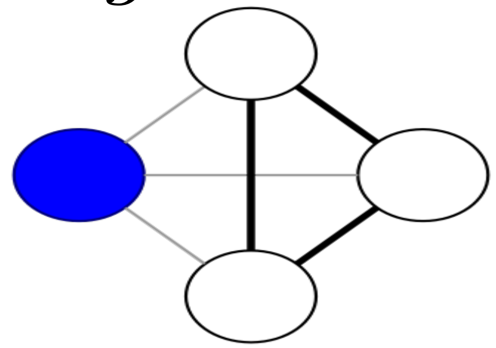
C_i = 0 if none of the neighbors of node *i* link to each other.

C_i = 1 if the neighbors of node *i* form a complete graph, i.e. they all link to each other.

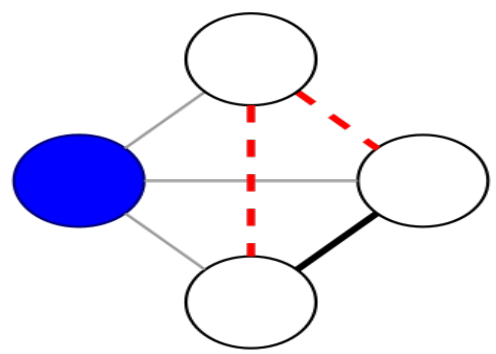
C_i is the probability that two neighbors of a node link to each other. Consequently *C* = 0.5 implies that there is a 50% chance that two neighbors of a node are linked.

Clustering coefficient

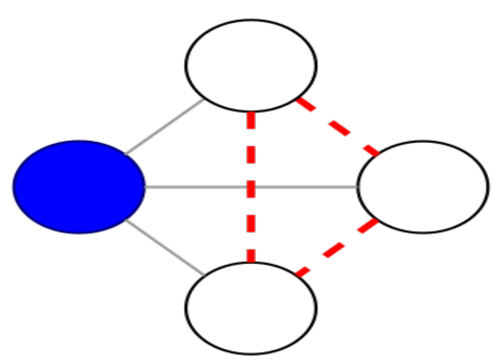
$$Neighbors = 3$$



c = 1



c = 1/3



c = 0

Transitivity

$$C_i = \frac{2(Number\ of\ connections\ between\ neighbors)}{k_i(k_i - 1)}$$

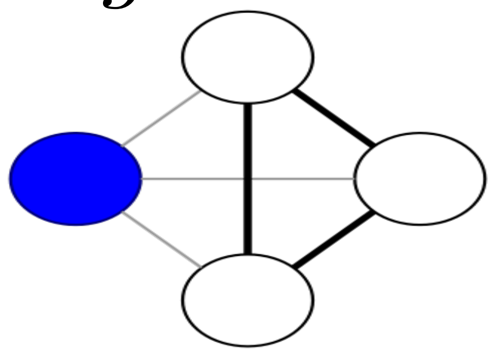
$$Number\ of\ connections\ between\ neighbors = 3$$

$$k_i = 3$$

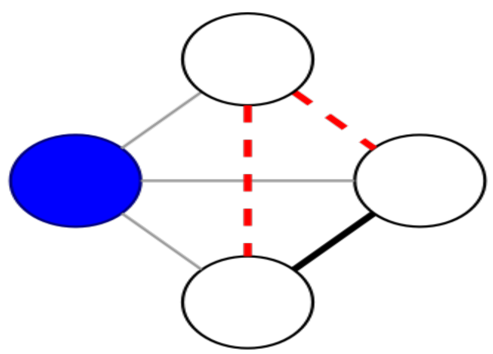
$$C_i = \frac{2 * 3}{3 * 2} = 1$$

Clustering coefficient

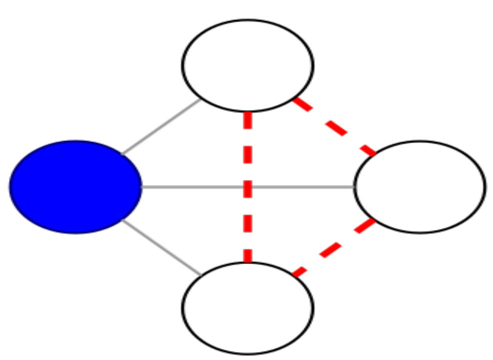
$$Neighbors = 3$$



c = 1



c = 1/3



c = 0

Transitivity

$$C_i = \frac{2(Number\ of\ connections\ between\ neighbors)}{k_i(k_i - 1)}$$

$$Number\ of\ connections\ between\ neighbors = 3$$

$$k_i = 3$$

$$C_i = \frac{2*3}{3*2} = 1$$

$$Number\ of\ connections\ between\ neighbors = 1$$

$$k_i = 3$$

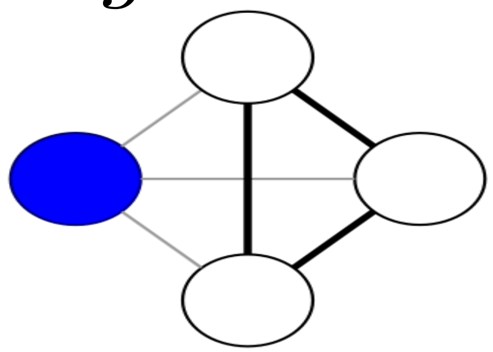
$$C_i = \frac{2*1}{3*2} = \frac{1}{3}$$

Clustering coefficient

$Neighbors = 3$

Transitivity

$$C_i = \frac{2(Number\ of\ connections\ between\ neighbors)}{k_i(k_i - 1)}$$

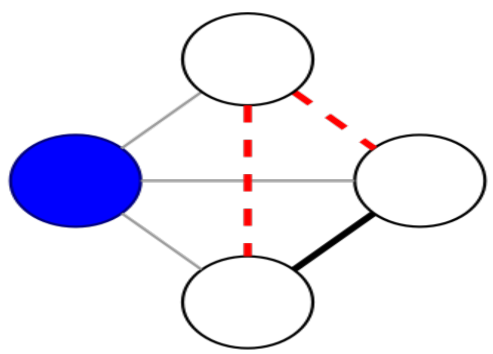


c = 1

$$Number\ of\ connections\ between\ neighbors = 3$$

$$k_i = 3$$

$$C_i = \frac{2 * 3}{3 * 2} = 1$$

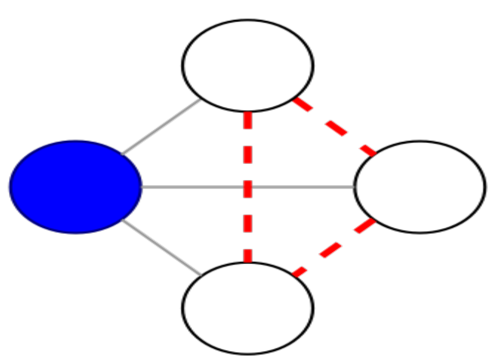


c = 1/3

$$Number\ of\ connections\ between\ neighbors = 1$$

$$k_i = 3$$

$$C_i = \frac{2 * 1}{3 * 2} = \frac{1}{3}$$

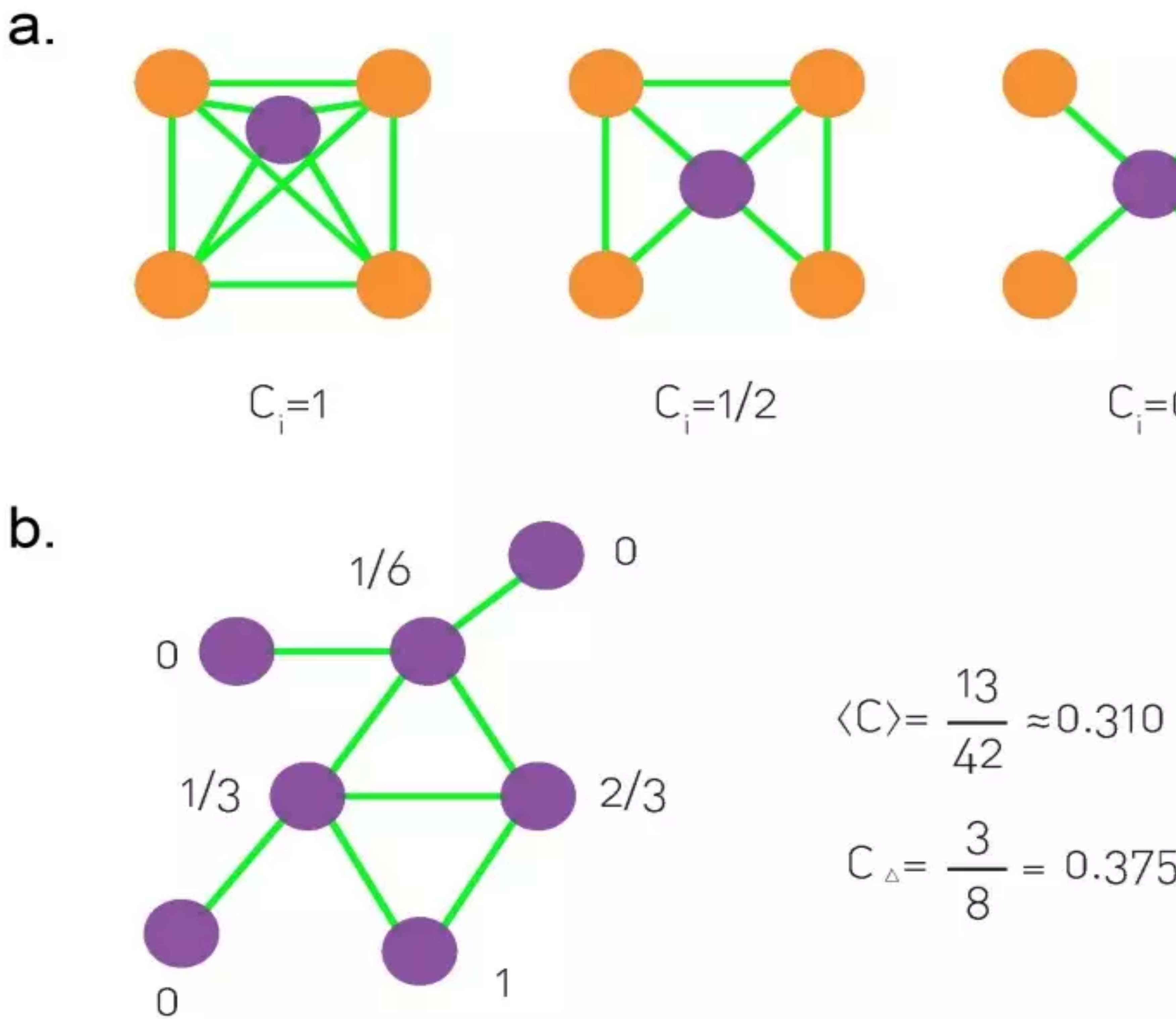


c = 0

$$Number\ of\ connections\ between\ neighbors = 0$$

$$k_i = 3$$

$$C_i = \frac{2 * 0}{3 * 2} = \frac{1}{3}$$

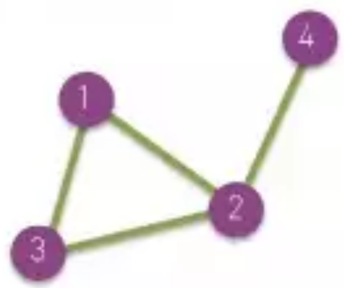


Clustering Coefficient

1. The local clustering coefficient, C_i , of the central node with degree k_i = 4 for three different configurations of its neighborhood. The local clustering coefficient measures the local density of links in a node's vicinity.
2. A small network, with the local clustering coefficient of each nodes shown next to it. We also list the network's average clustering coefficient $\langle C \rangle$, according to (2.16), and its global clustering coefficient C_Δ, Note that for nodes with degrees k_i = 0,1, the clustering coefficient is zero.

Summary

a. Undirected

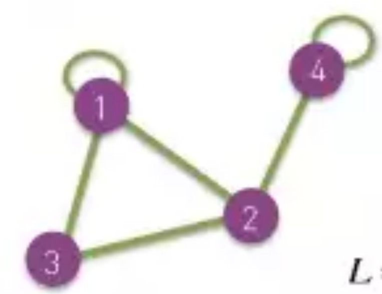


$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

b. Self-loops

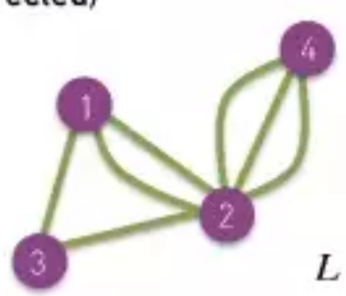


$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$\exists i, A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1, i\neq j}^{N} A_{ij} + \sum_{i=1}^{N} A_{ii} \qquad ?$$

c. Multigraph
(undirected)

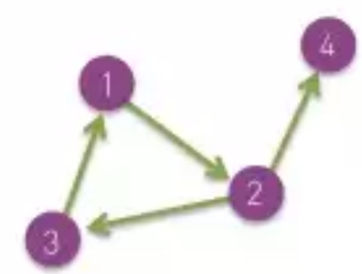


$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

d. Directed

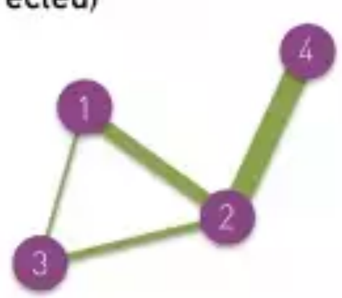


$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{L}{N}$$

e. Weighted
(undirected)

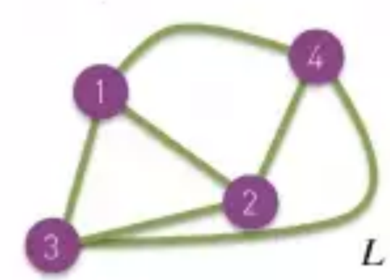


$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$<k> = \frac{2L}{N}$$

f. Complete Graph
(undirected)



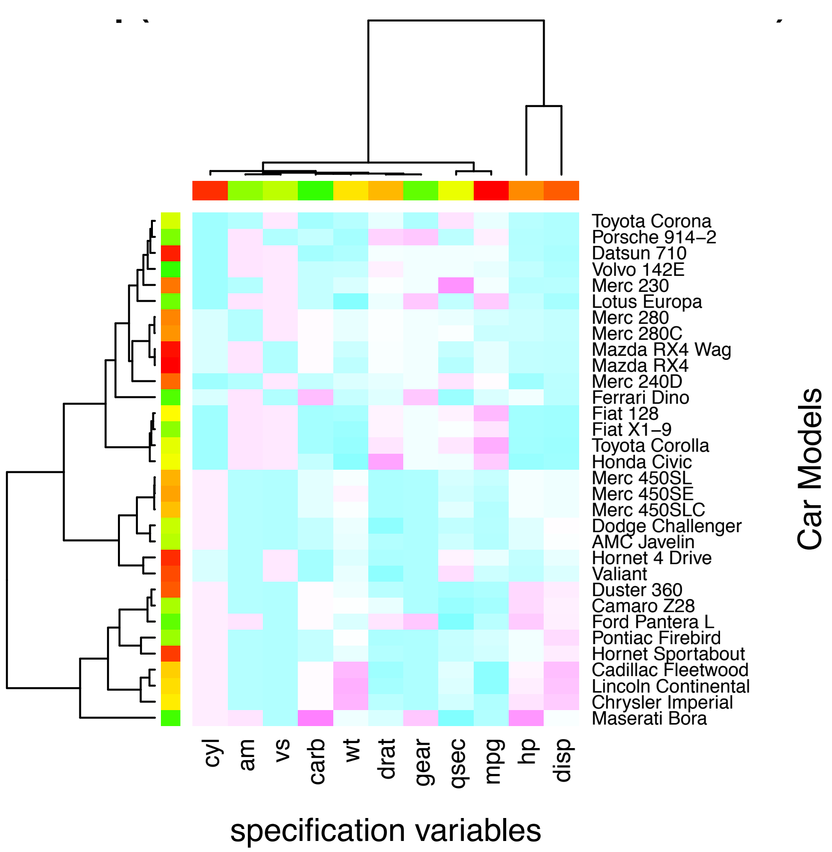
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{i\neq j} = 1$$

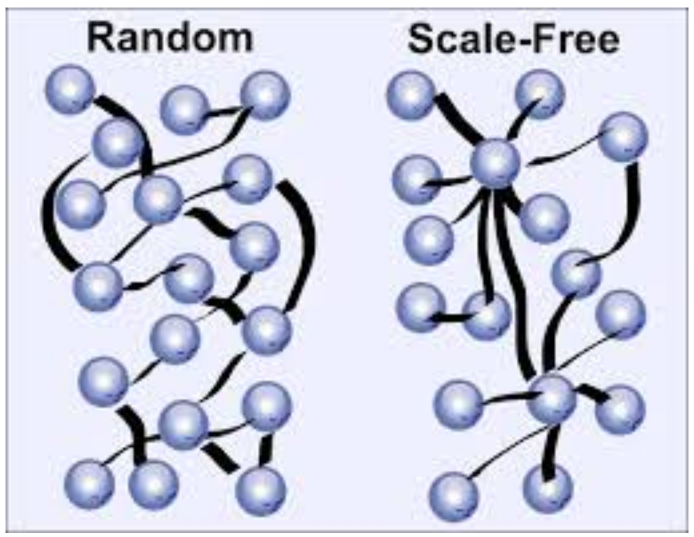
$$L = L_{\max} = \frac{N(N-1)}{2} \qquad <k> = N-1$$

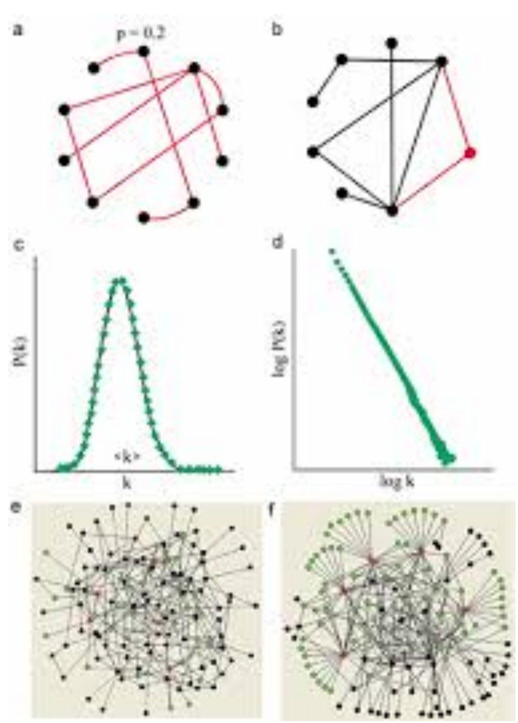
Exercises 2

Exercises 2 Distances and Clustering Coefficient



3. The Scale-Free property



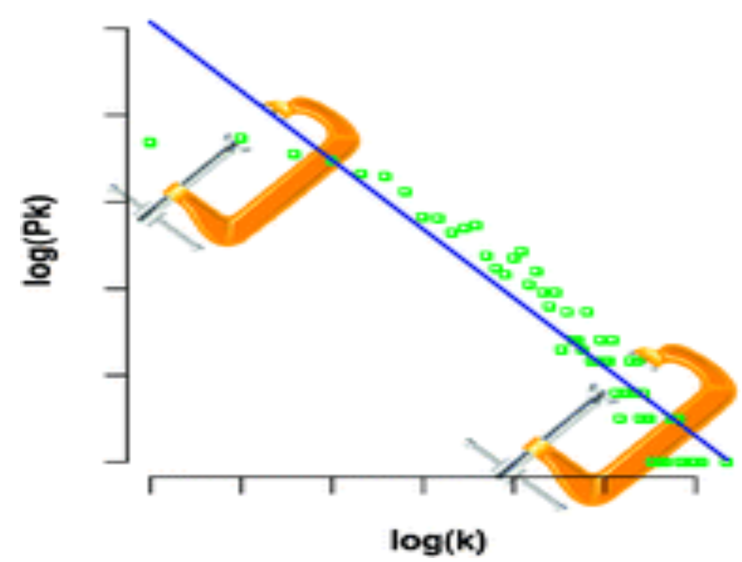


REVIEW www.rsc.org/molecularbiosystems | Molecular BioSystems

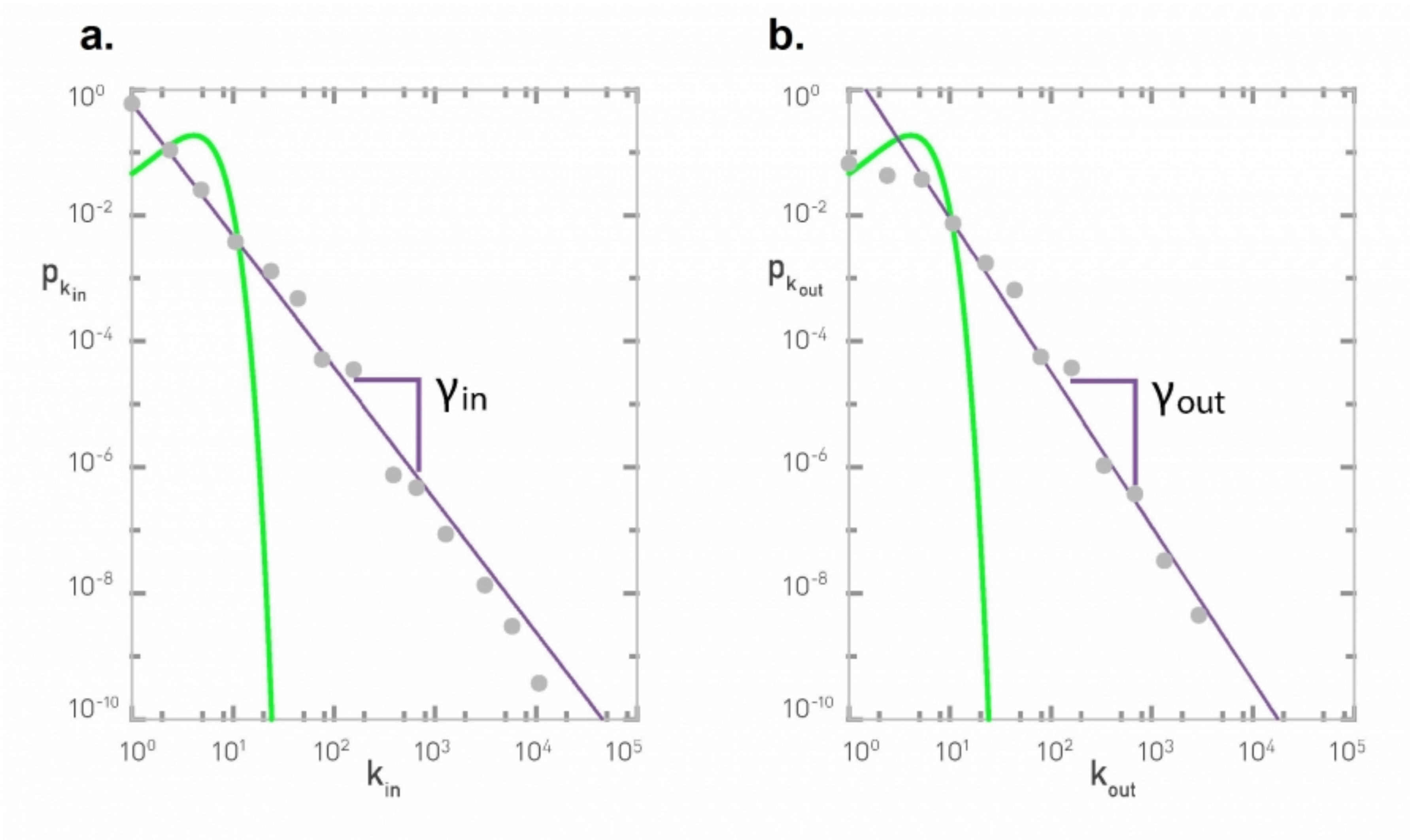
The powerful law of the power law and other myths in network biology†

Gipsi Lima-Mendez* and Jacques van Helden*

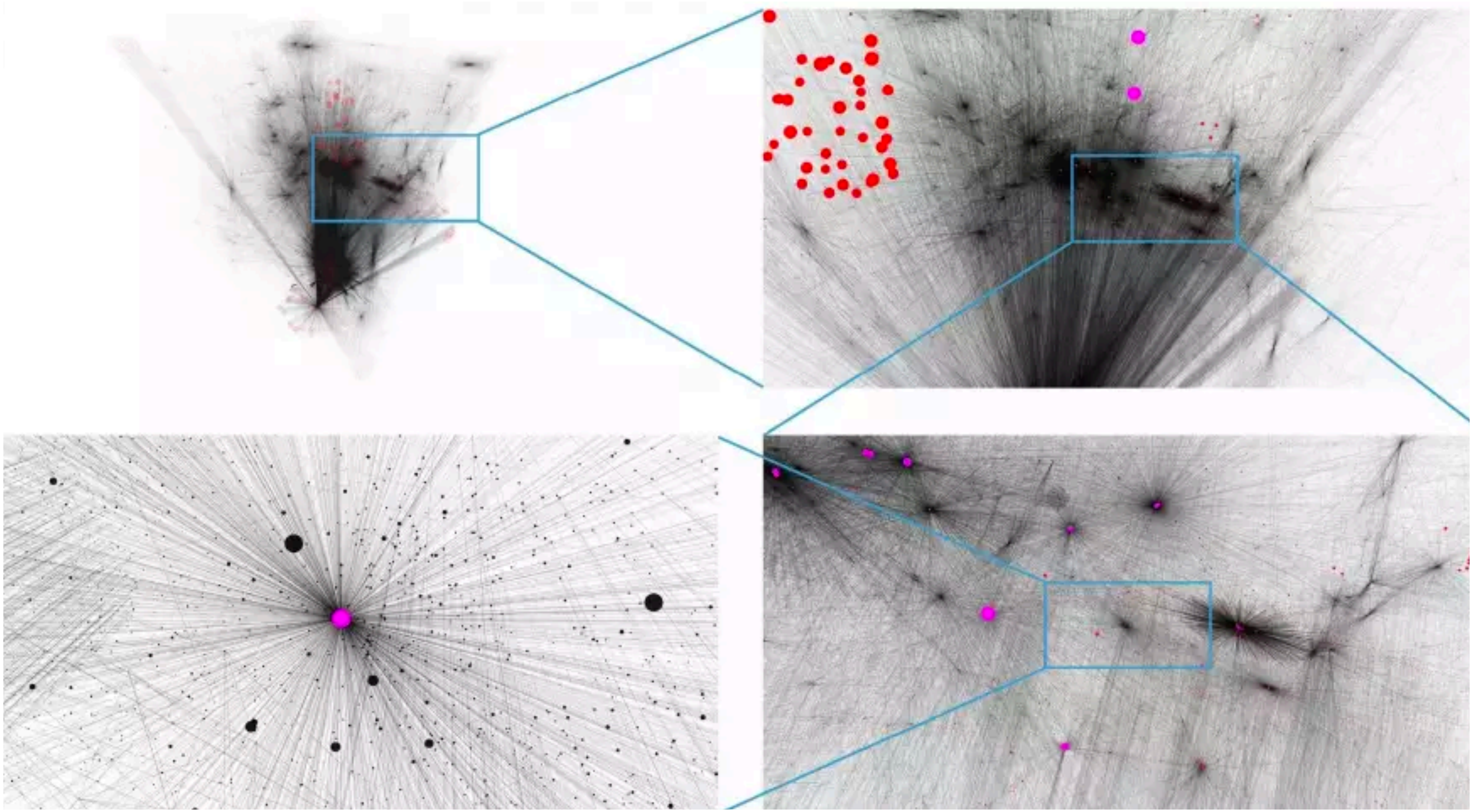
Received 5th May 2009, Accepted 12th August 2009
First published as an Advance Article on the web 2nd October 2009
DOI: 10.1039/b908681a



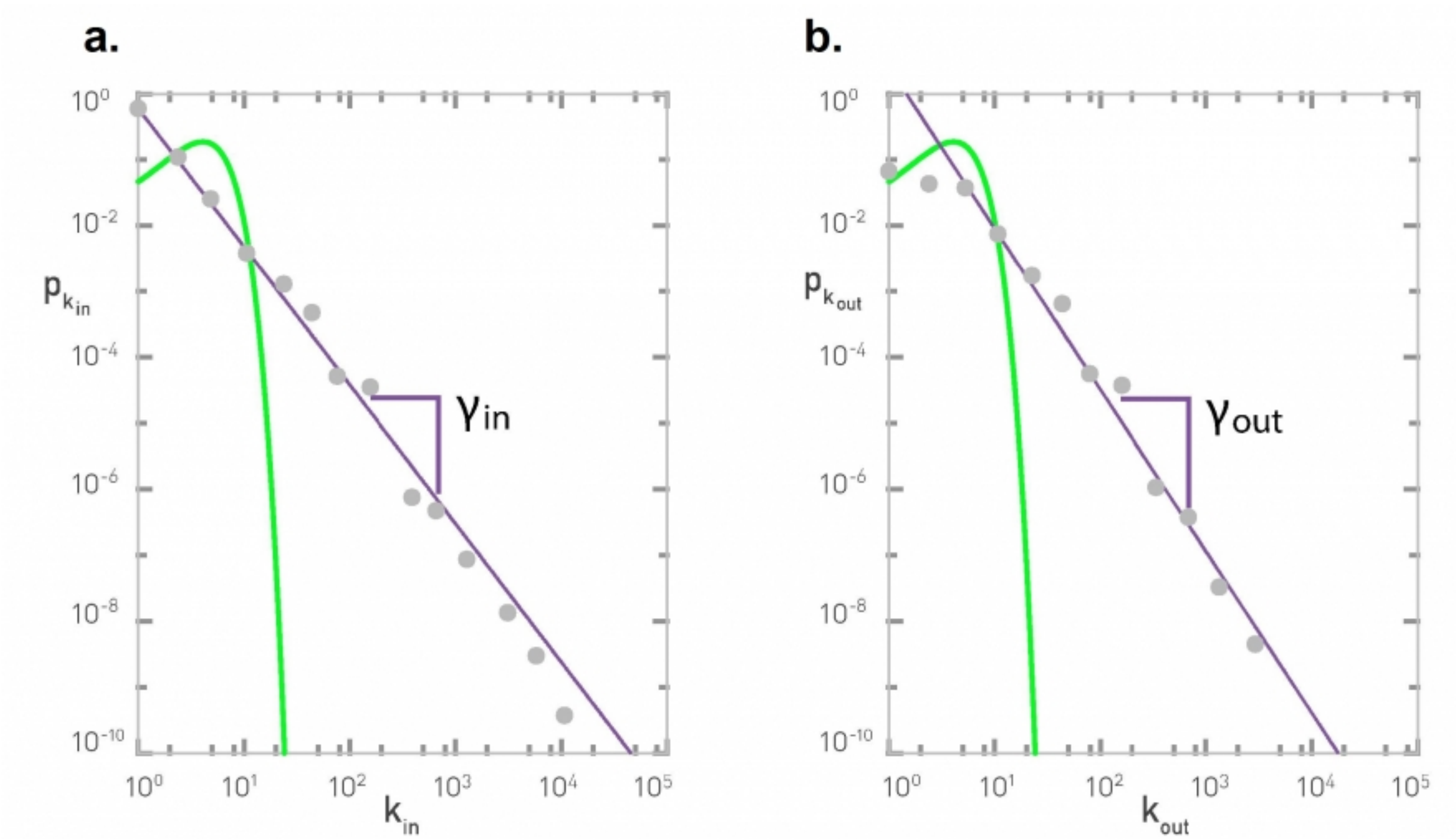
The Degree Distribution of the WWW



The topology of the World Wide Web



The Degree Distribution of the WWW



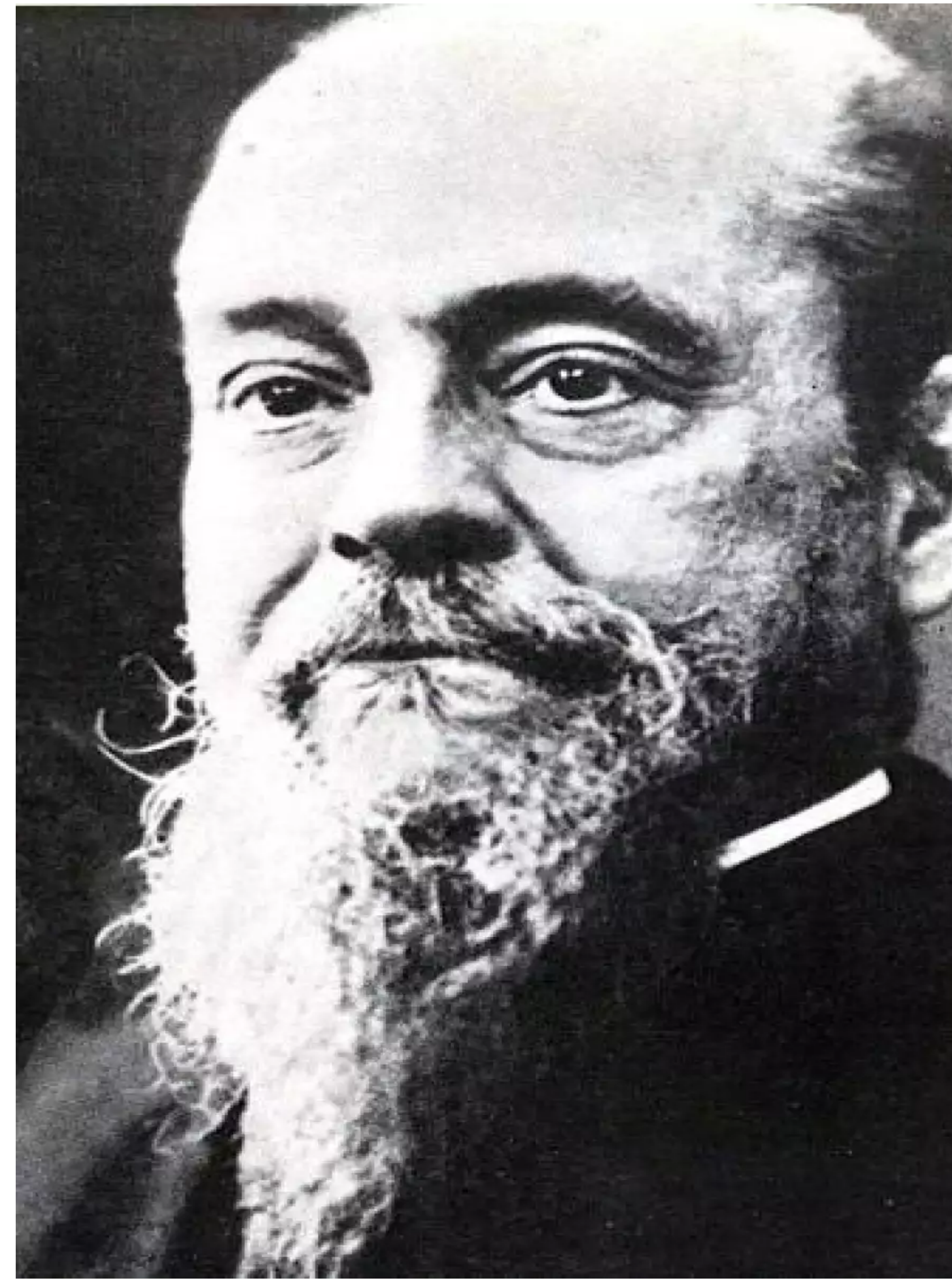
The 80/20 Rule and the Top One Percent

Vilfredo Pareto, a 19th century economist, noticed that in Italy a few wealthy individuals earned most of the money, while the majority of the population earned rather small amounts. He connected this disparity to the observation that incomes follow a power law, representing the first known report of a power-law distribution [3]. His finding entered the popular literature as the *80/20 rule*: Roughly 80 percent of money is earned by only 20 percent of the population.

The 80/20 rule emerges in many areas. For example in management it is often stated that 80 percent of profits are produced by only 20 percent of the employees. Similarly, 80 percent of decisions are made during 20 percent of meeting time.

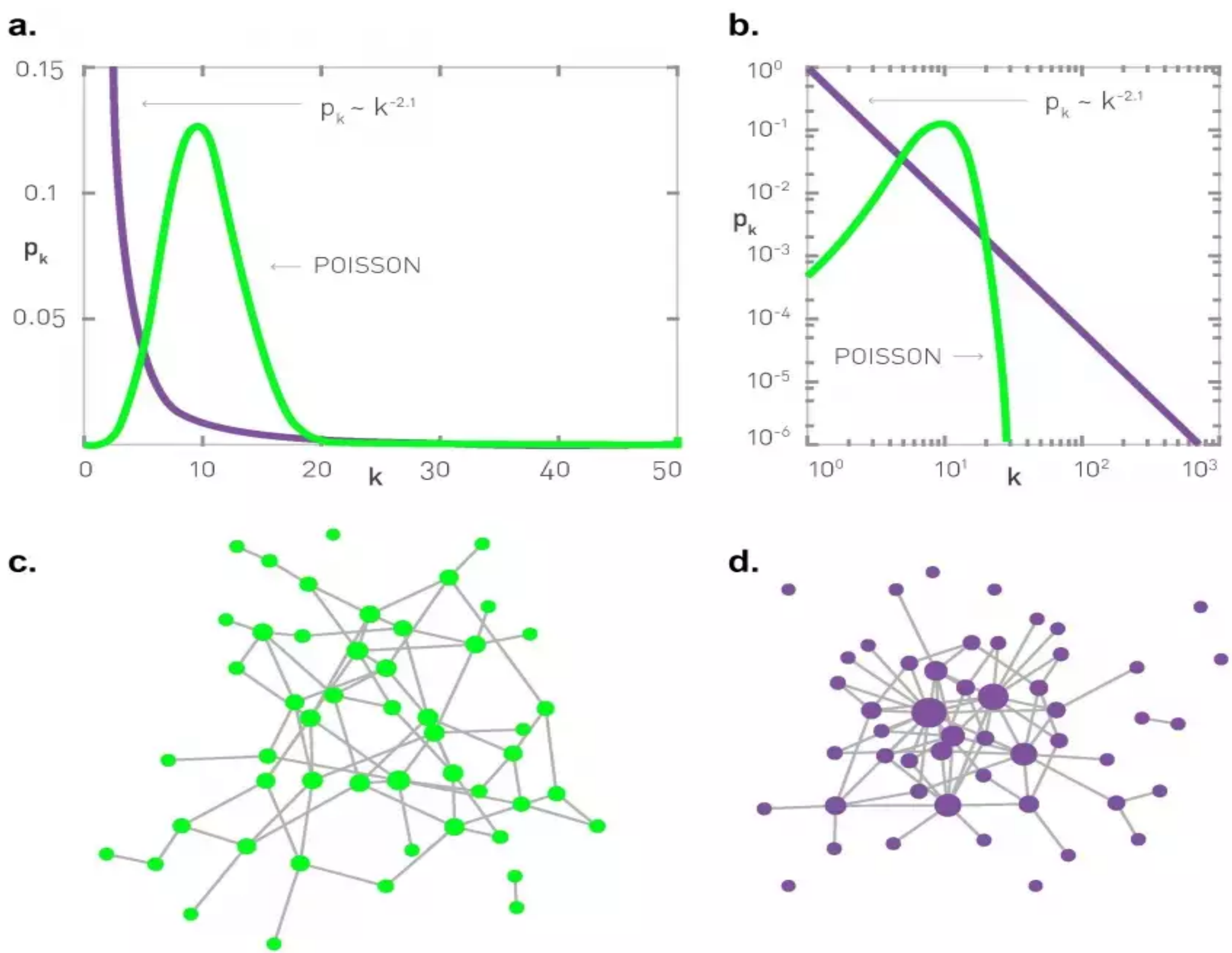
The 80/20 rule is present in networks as well: 80 percent of links on the Web point to only 15 percent of webpages; 80 percent of citations go to only 38 percent of scientists; 80 percent of links in Hollywood are connected to 30 percent of actors [4]. Most quantities following a power law distribution obey the 80/20 rule.

During the 2009 economic crisis power laws gained a new meaning: The Occupy Wall Street Movement draw attention to the fact that in the US 1% of the population earns a disproportionate 15% of the total US income. This 1% phenomena, a signature of a profound income disparity, is again a consequence of the power-law nature of the income distribution.



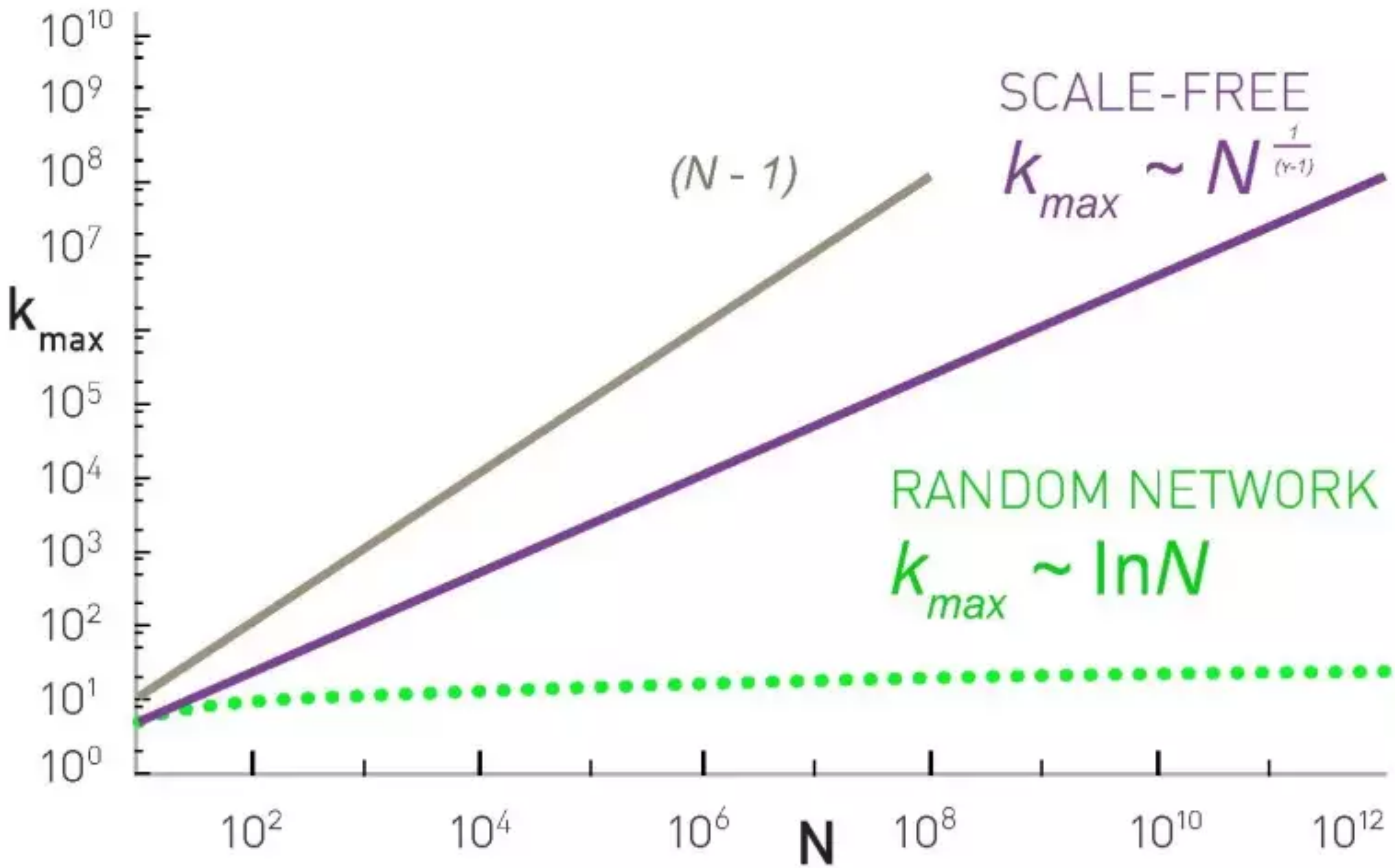
Vilfredo Federico Damaso Pareto (1848 – 1923)
Italian economist, political scientist, and philosopher, who had important contributions to our understanding of income distribution and to the analysis of individual choices. A number of fundamental principles are named after him, like *Pareto efficiency, Pareto distribution* (another name for a power-law distribution), the *Pareto principle* (or 80/20 law).

Hubs

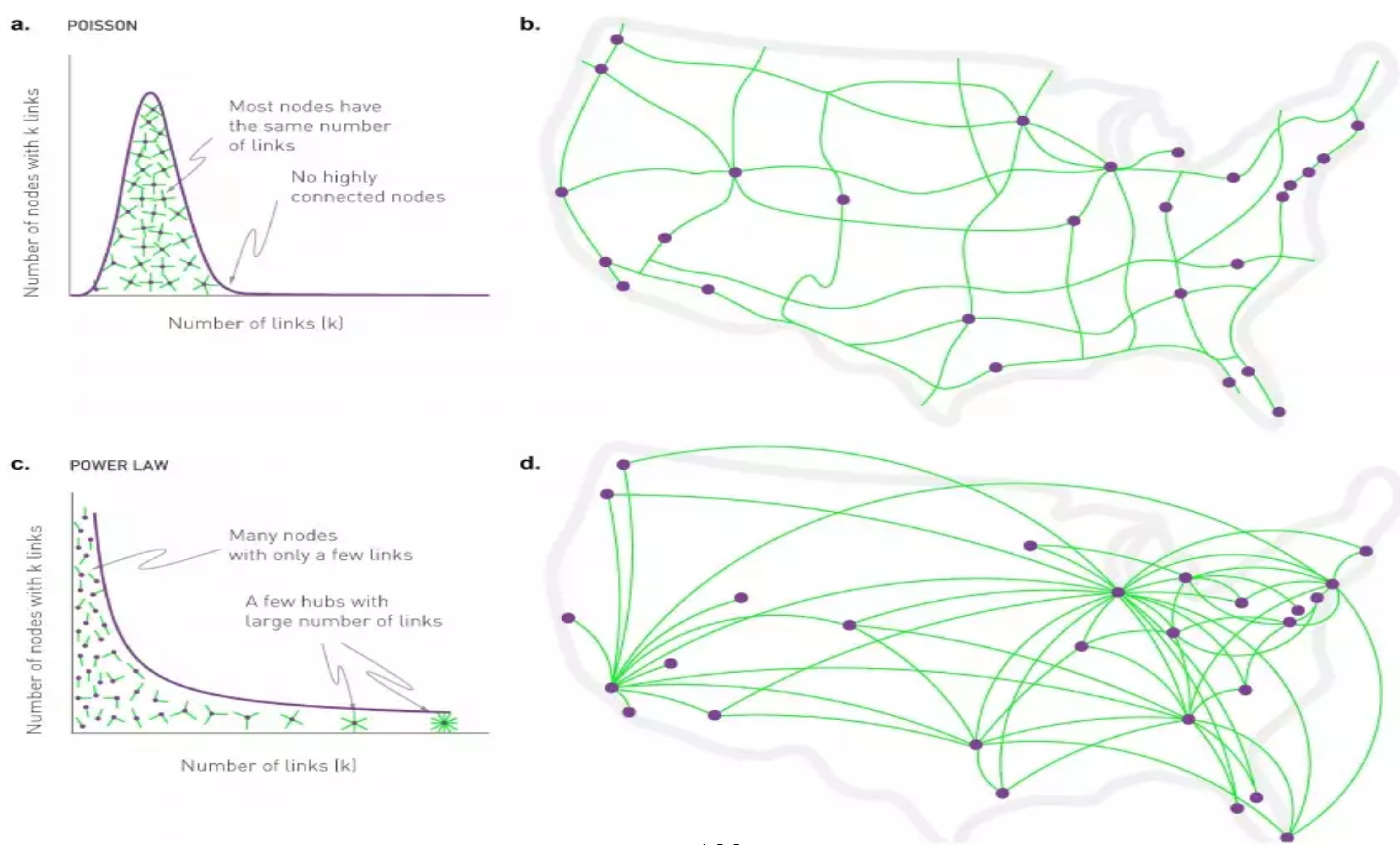


a.
0.15
0.1
0.05
p_k
$p_k \sim k^{-2.1}$
POISSON
0
10
20
30
40
50
k
b.
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
p_k
$p_k \sim k^{-2.1}$
POISSON
10^0
10^1
10^2
10^3
k
c.
d.

Hubs are Large in Scale-free Networks



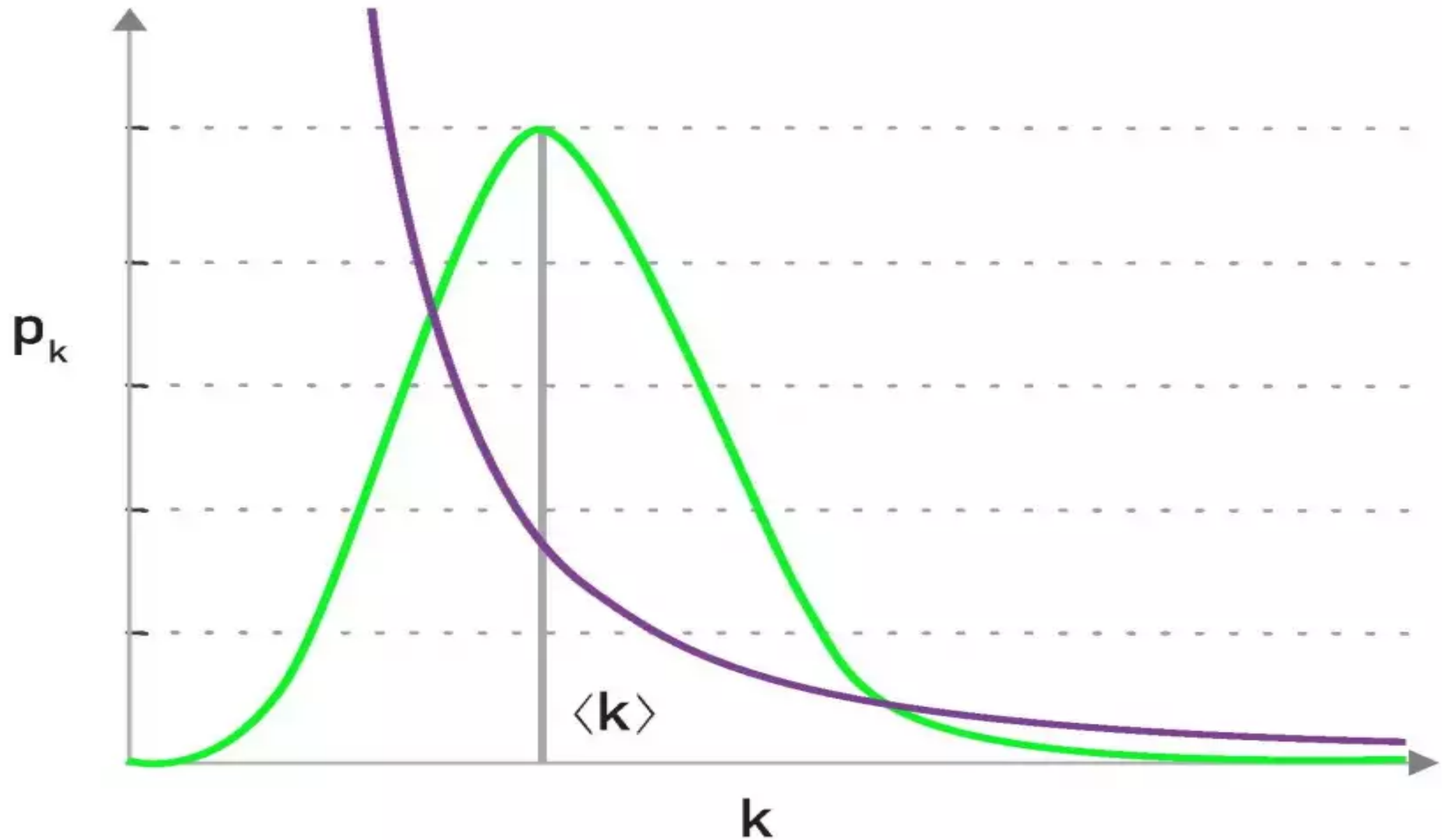
Random vs. Scale-free Networks



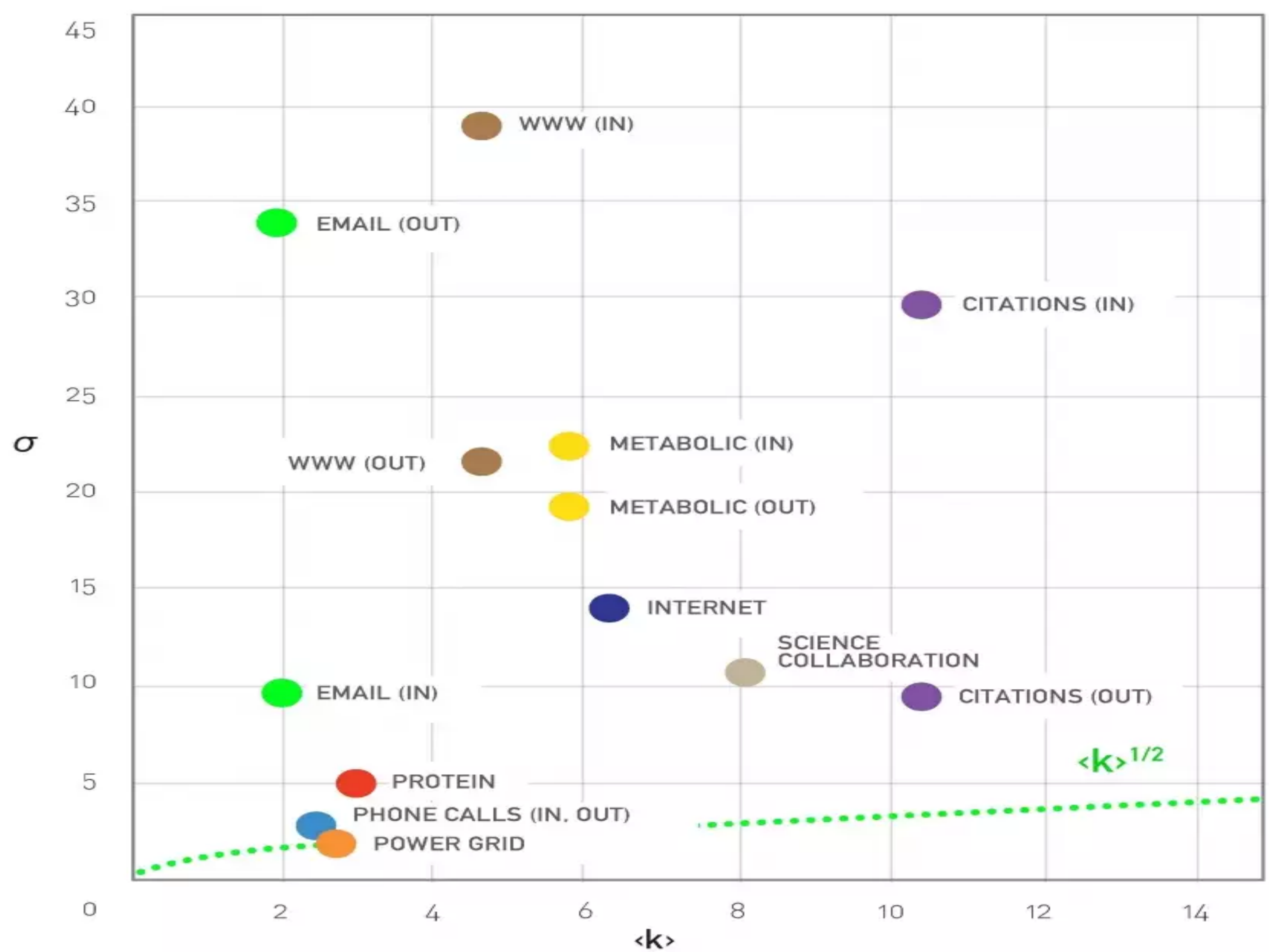
Degree Fluctuations in Real Networks

Network	N	L	⟨k⟩	⟨kin2⟩	⟨kout2⟩	⟨k2⟩	γin	γout	γ
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.6	1546	482.4	-	2	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.70	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.9	-	-	32.3	-	-	2.89*-

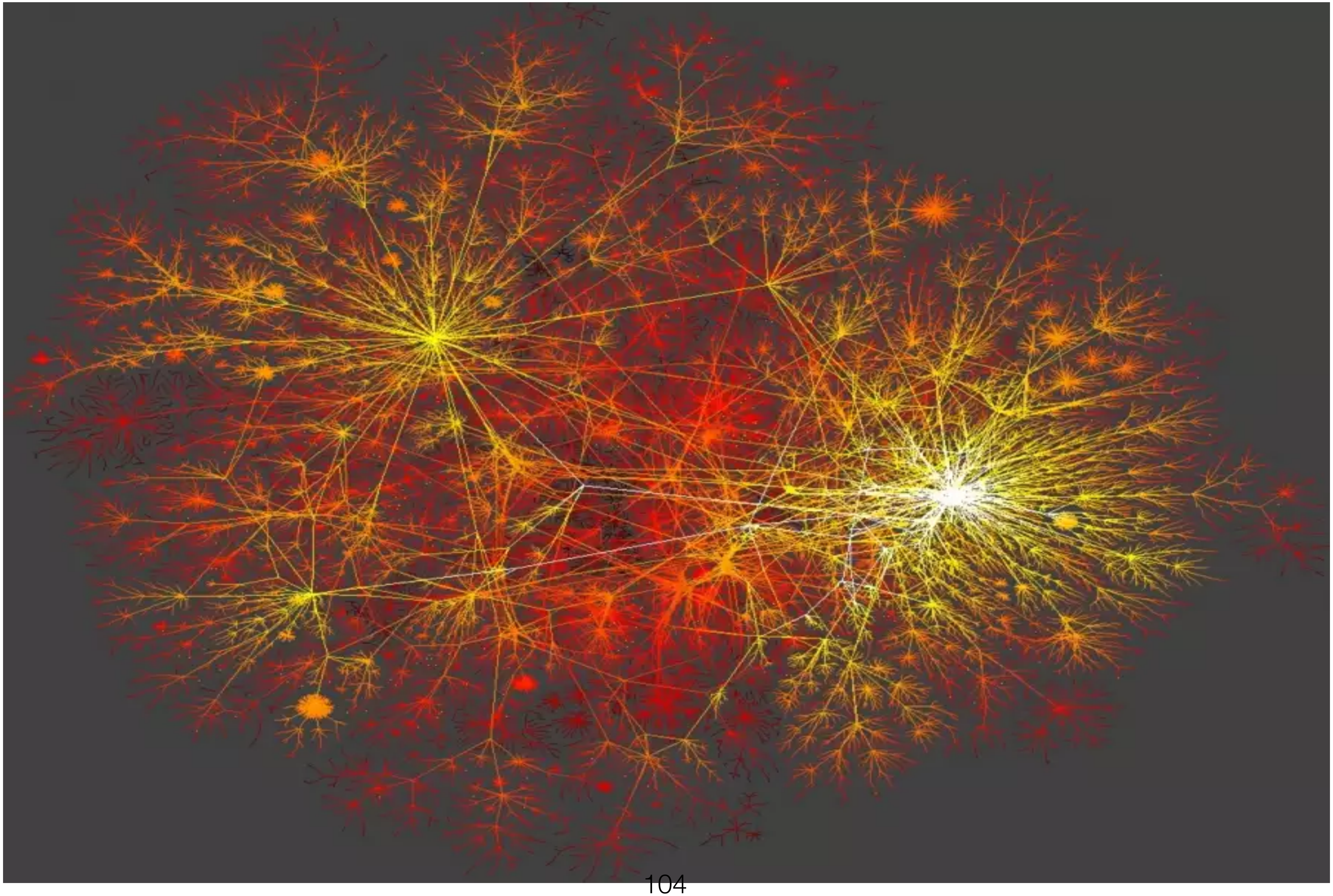
Lack of an Internal Scale



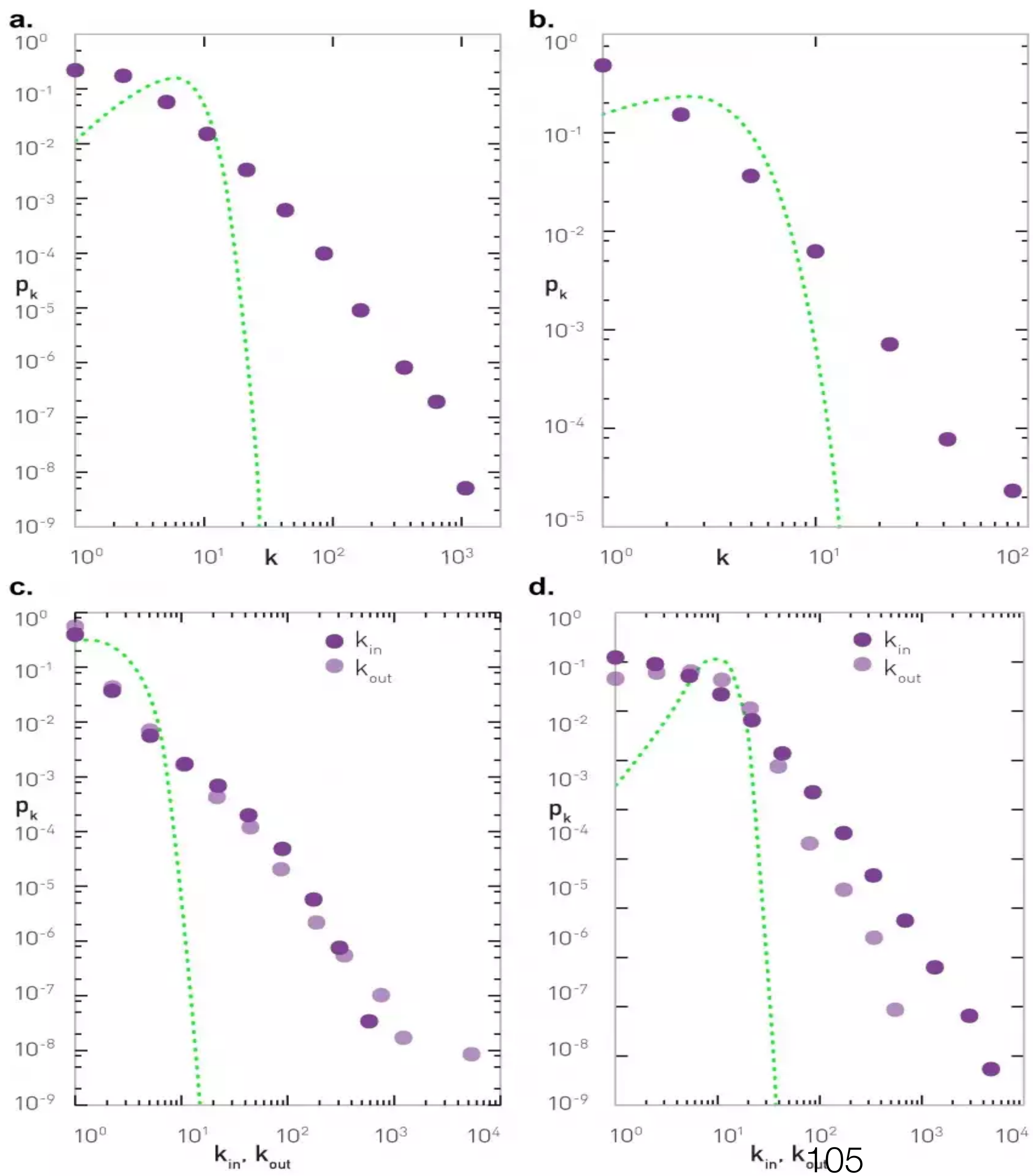
Standard Deviation is Large in Real Networks



Universality



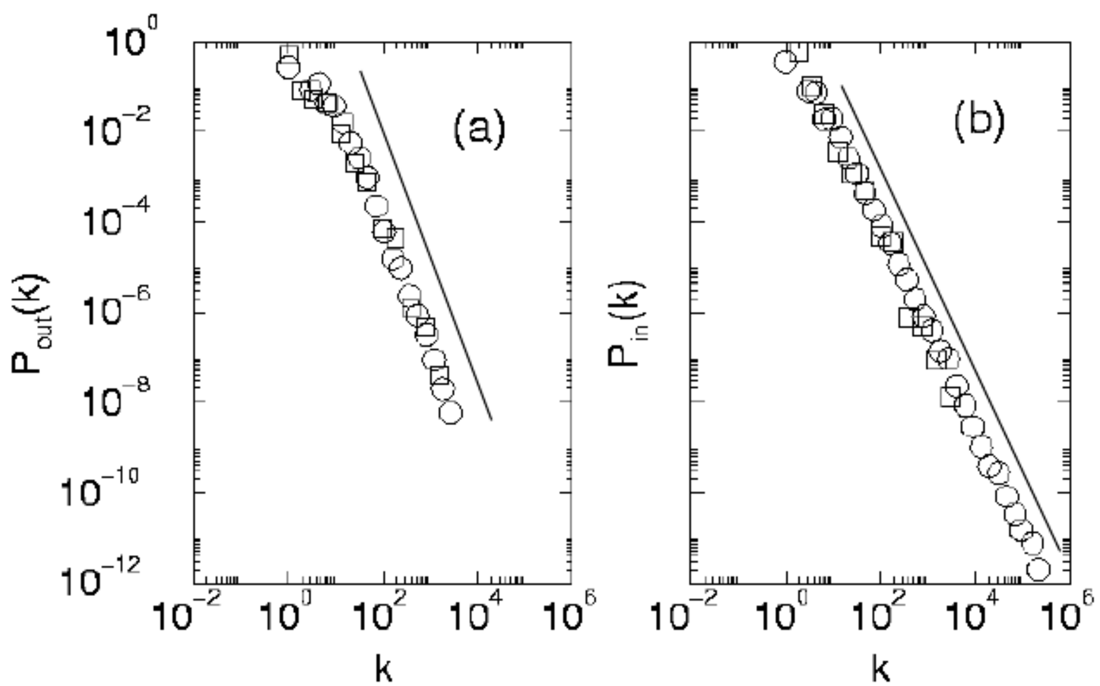
Many real networks are free-scale



The degree distribution of four networks

1. Internet at the router level.
2. Protein-protein interaction network.
3. Email network.
4. Citation network.

In- and out-degree distribution of the WWW



nodes: webpages

edges: hyperlinks

$$P_{out}(k) \approx k^{-2.45}$$

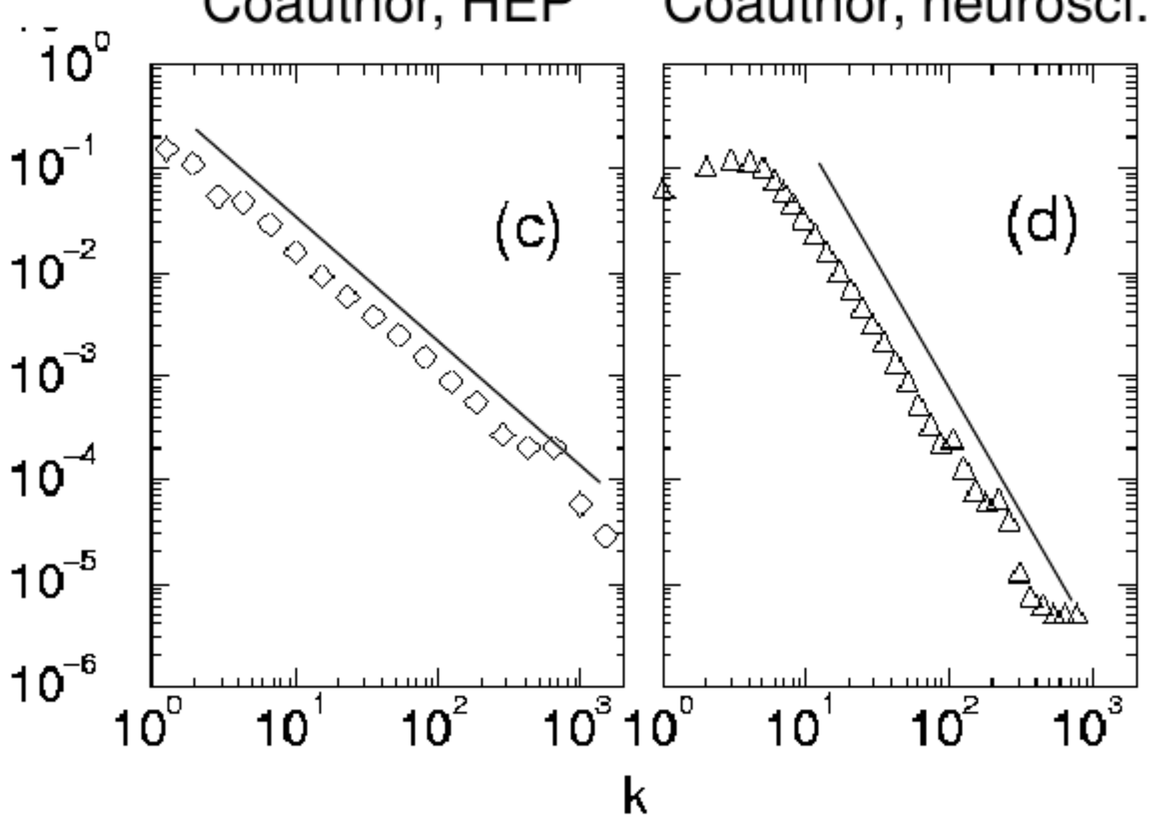
$$P_{in}(k) \approx k^{-2.1}$$

Usage: the degree distribution scales as a power law

R. Albert, H. Jeong, A.-L. Barabási, Nature 401, 130 (1999)

A. Broder *et al.*, Comput. Netw. 33, 309 (1999)

Degree distributions in networks of science collaborations



$P(k)$

$P(k) \approx k^{-1.2}$

$P(k) \approx k^{-2.1}$

M. E. J. Newman, Phys. Rev. E 64, 016131 (2001)

A.-L. Barabási et al., cond-mat/0104162 (2001)

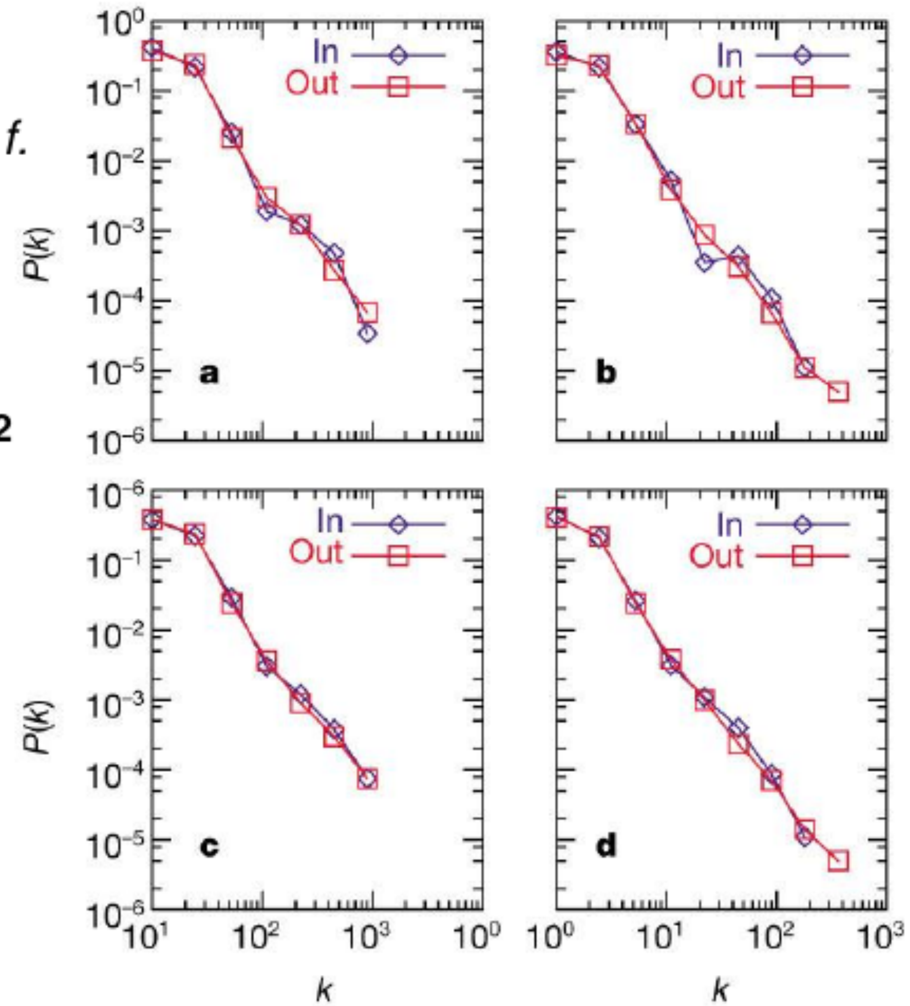
Metabolic networks have a power-law degree distribution

Archaeoglobus f.

$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

C. elegans



E. coli

bipartite

nodes: metabolites,

reactions

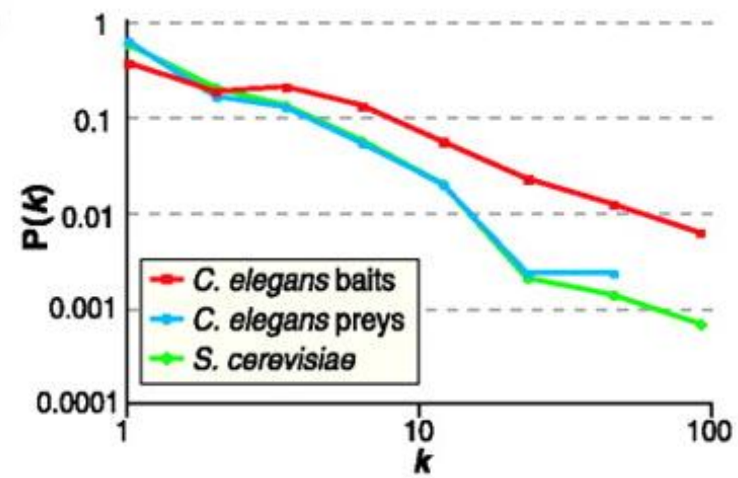
directed edges,

out: reactant (substrate)

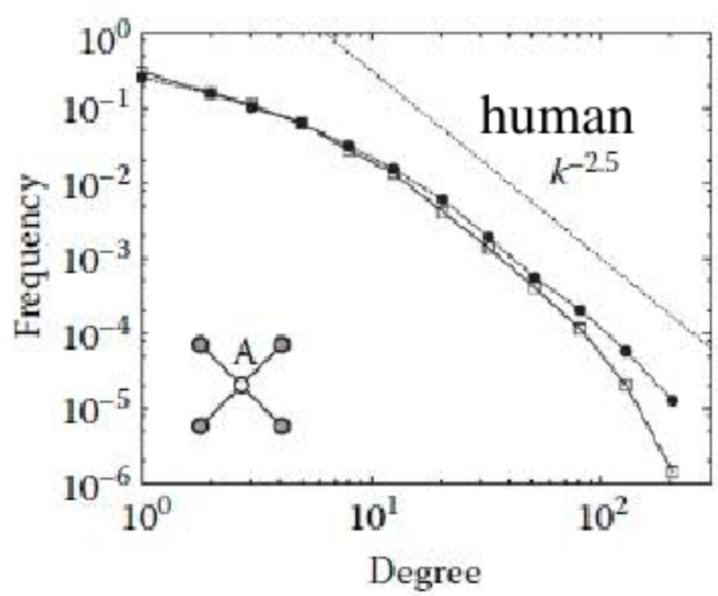
in: product of reaction

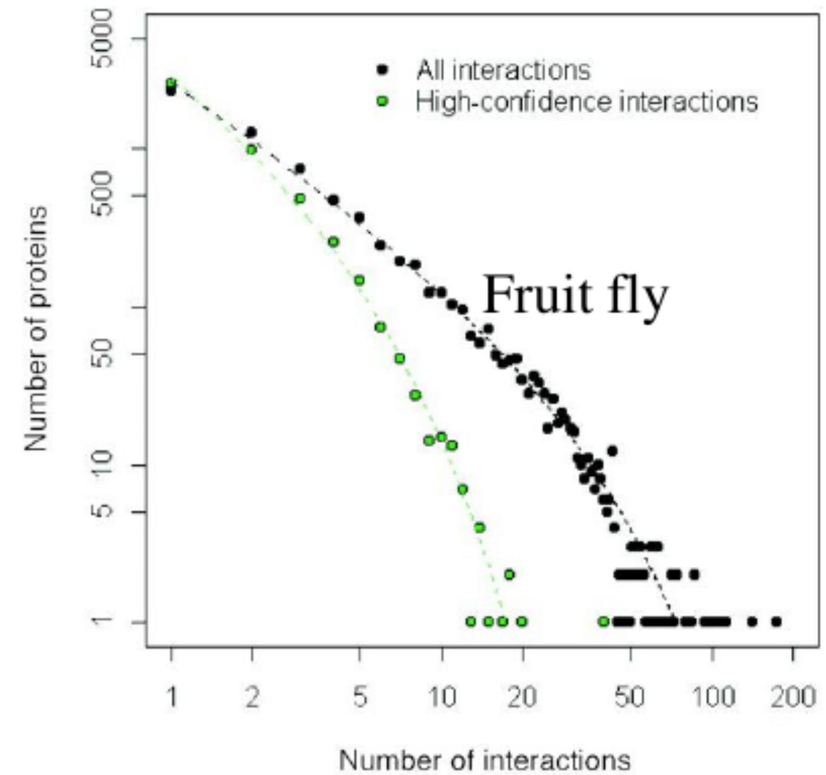
H. Jeong et al., Nature 407, 651 (2000)

Degree distribution of protein networks



$$P(k) \approx Ak^{-\gamma}$$

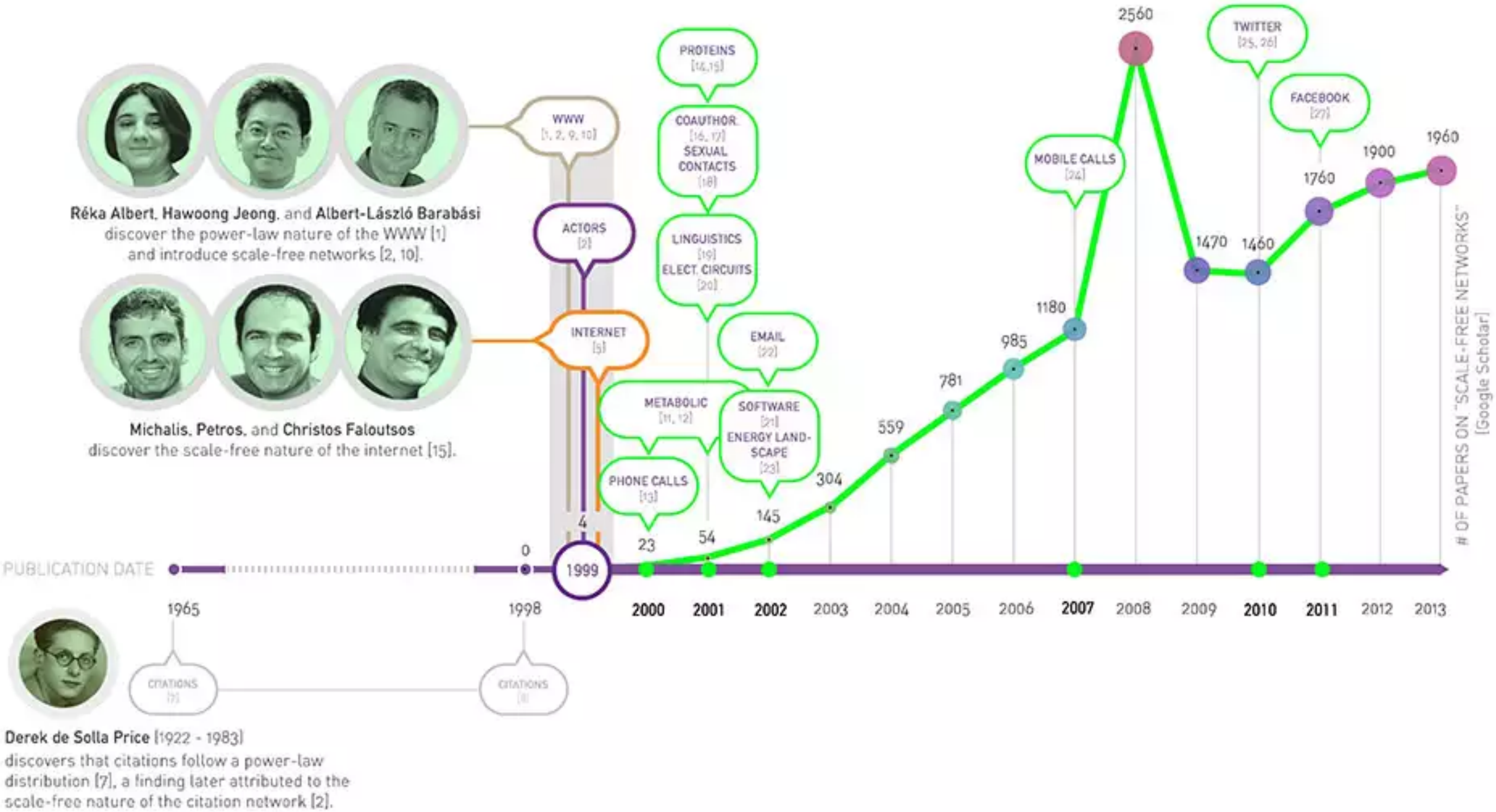




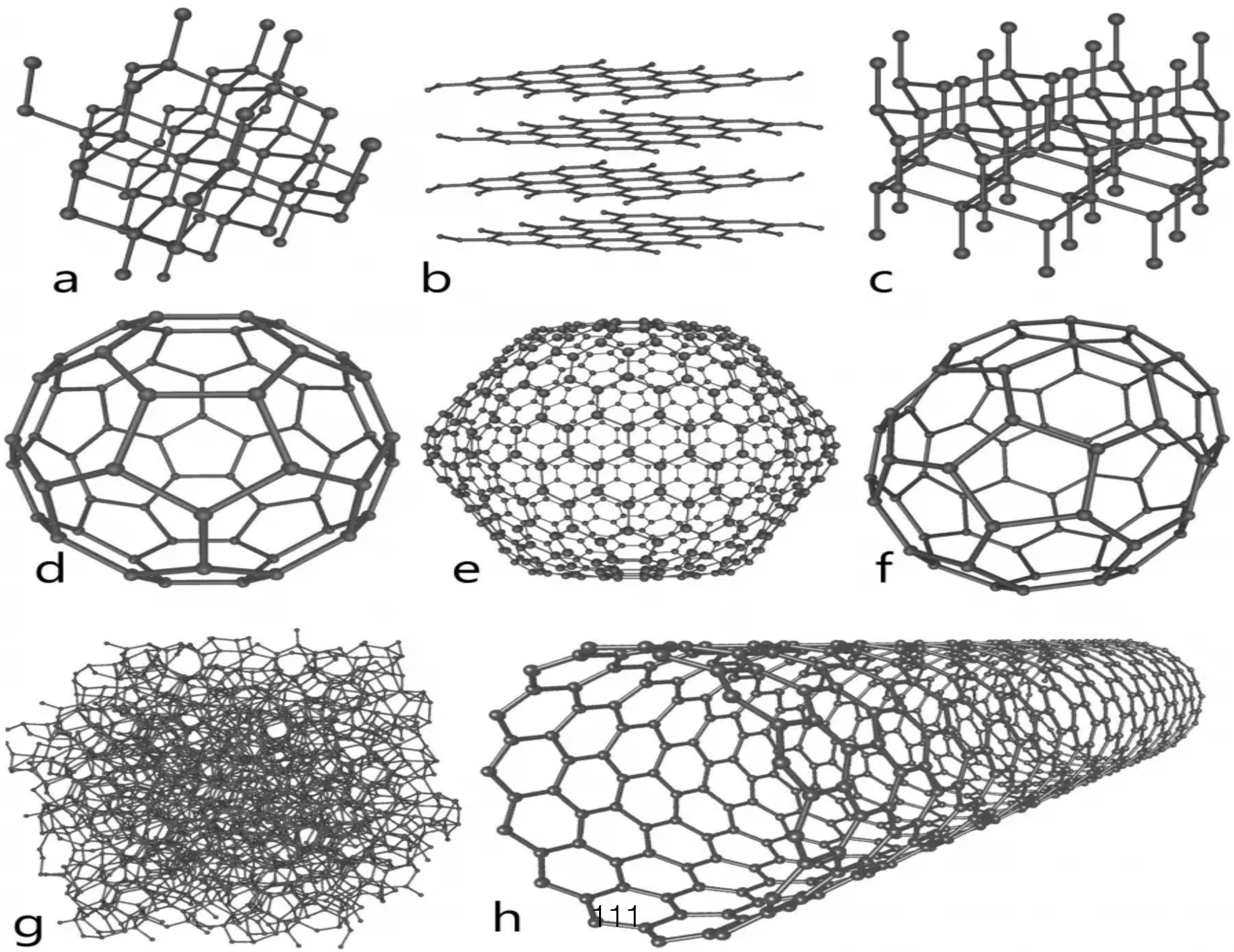
$$P(k) \approx Ak^{-\gamma} \exp(-\beta k)$$

Giot et al. Science 2003 – *Drosophila m.*
Li et al. Science 2004 – *C. elegans*
Rual et al. Nature 2005 – human
Stelzl et al. Cell 2005 - human

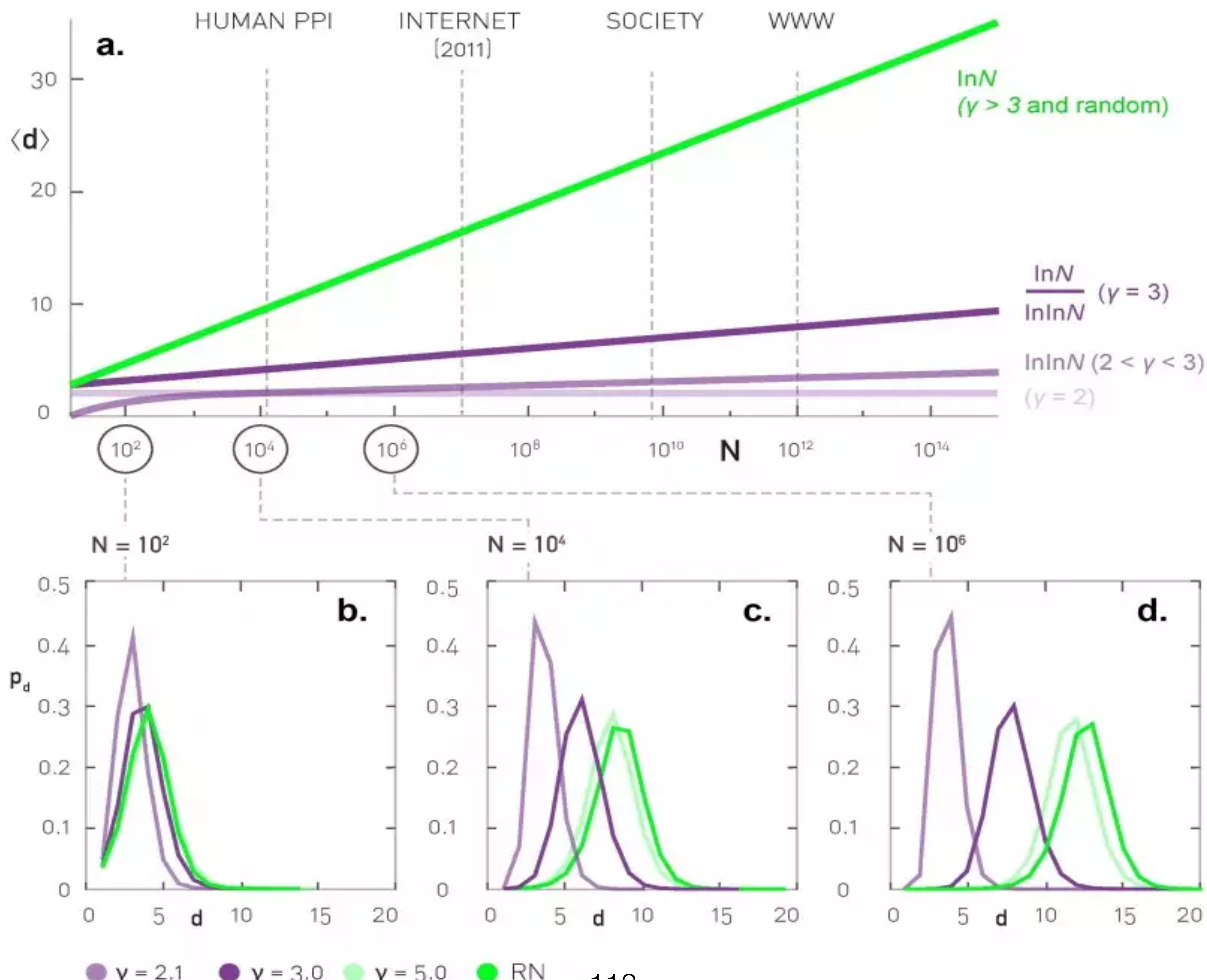
Timeline:Scale-Free Networks



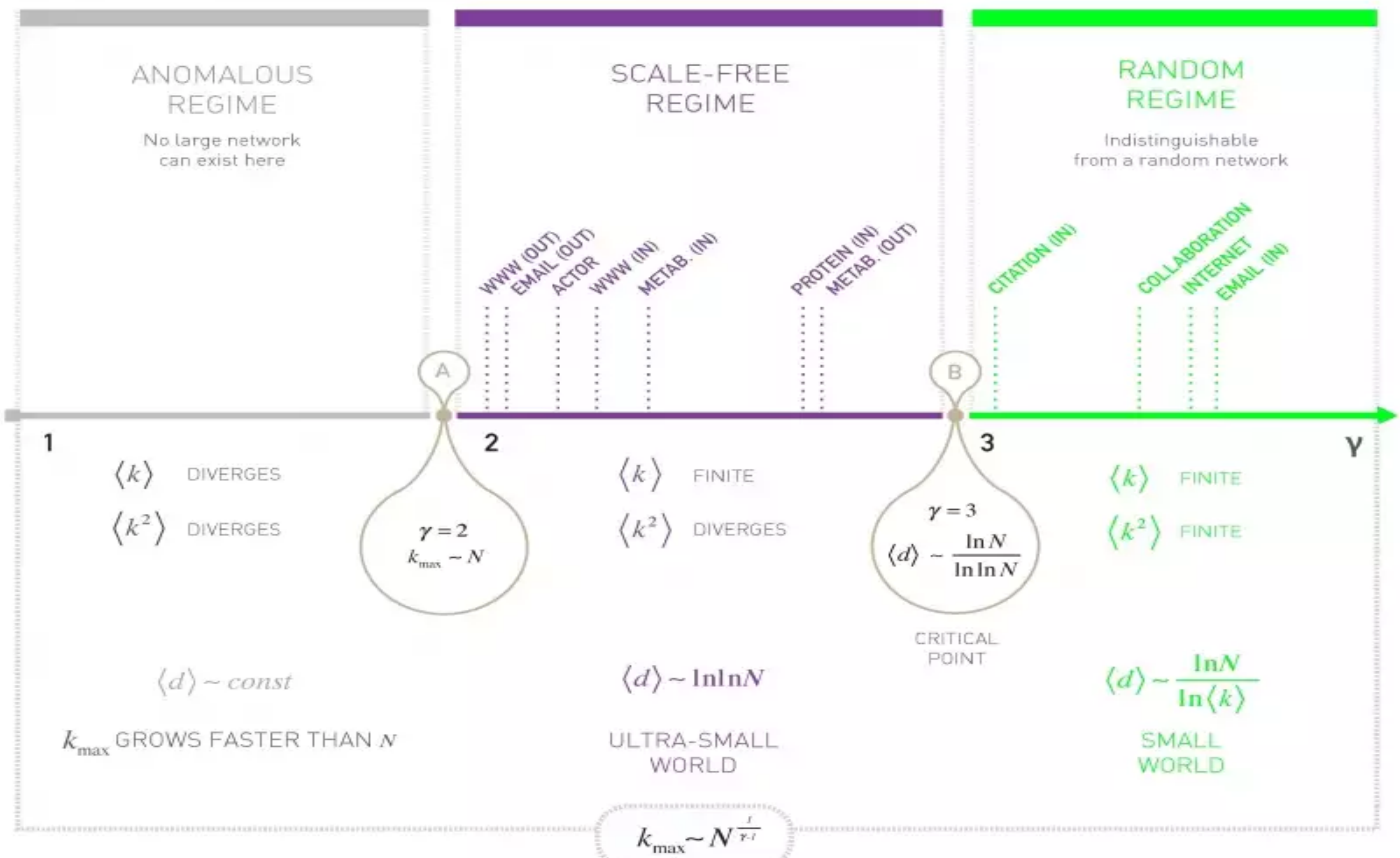
Not All Network Are Scale-Free



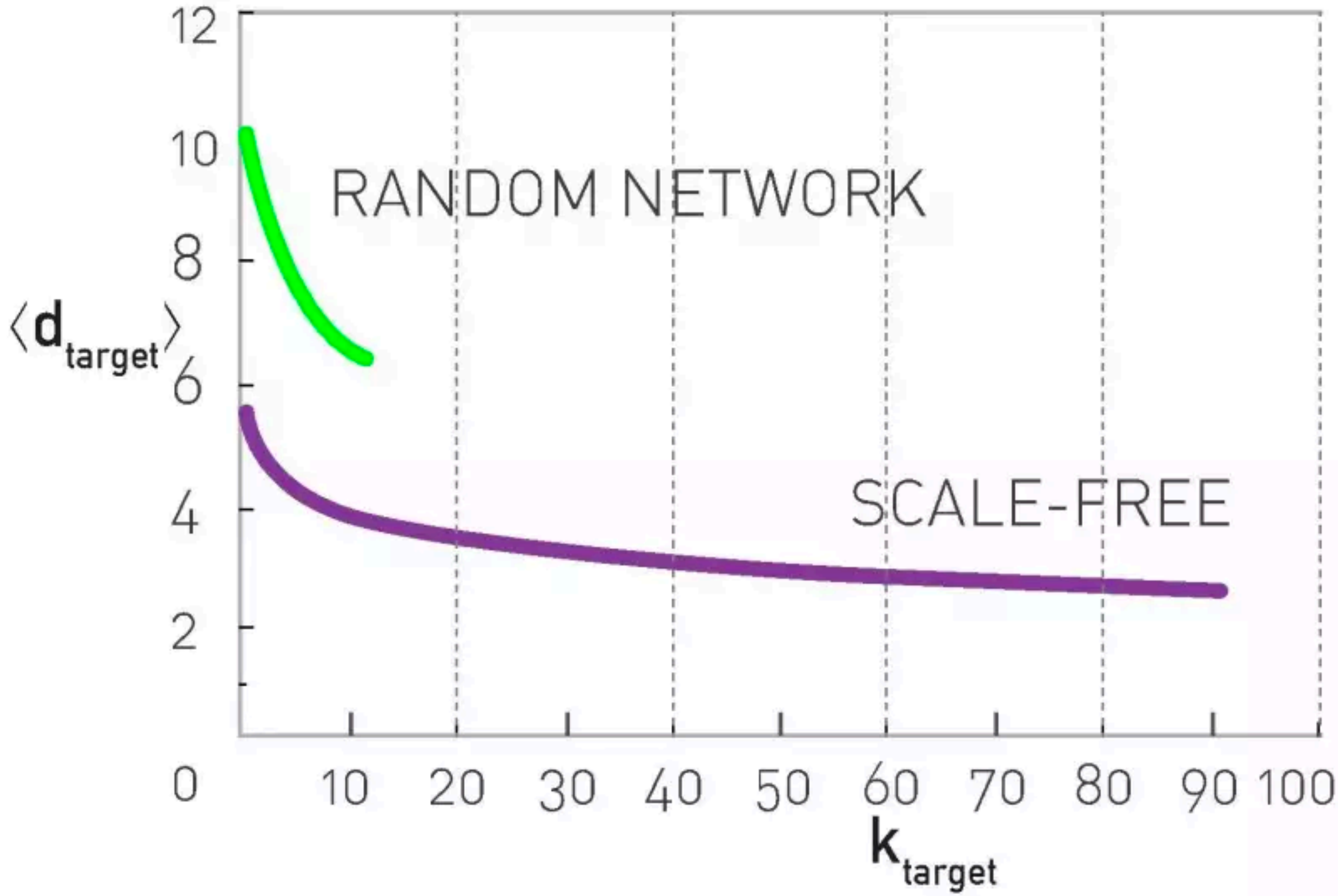
Ultra-Small Property



The γ Dependent Properties of Scale-Free Networks



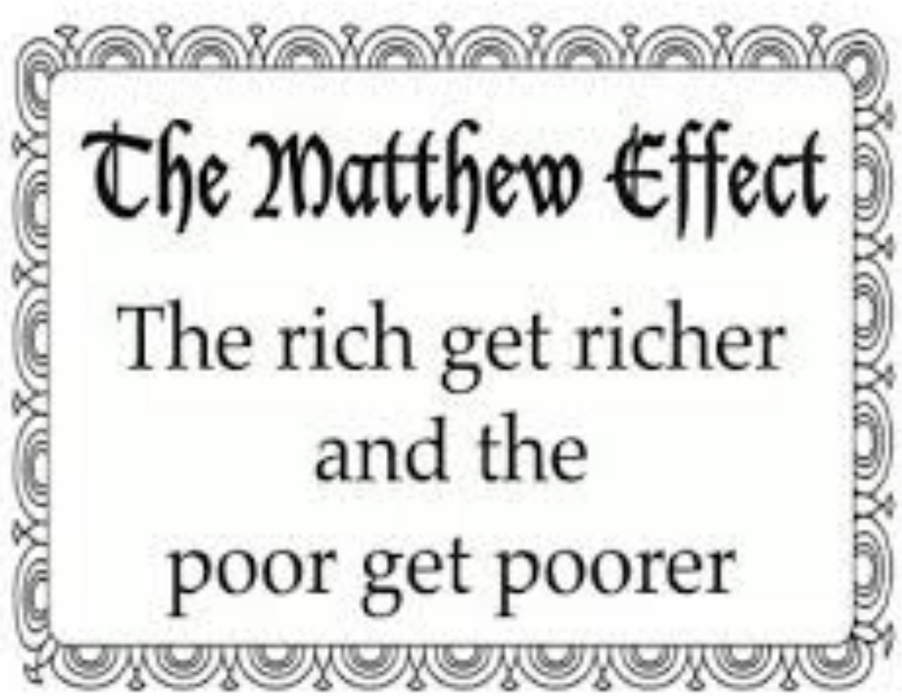
Closing on the hubs



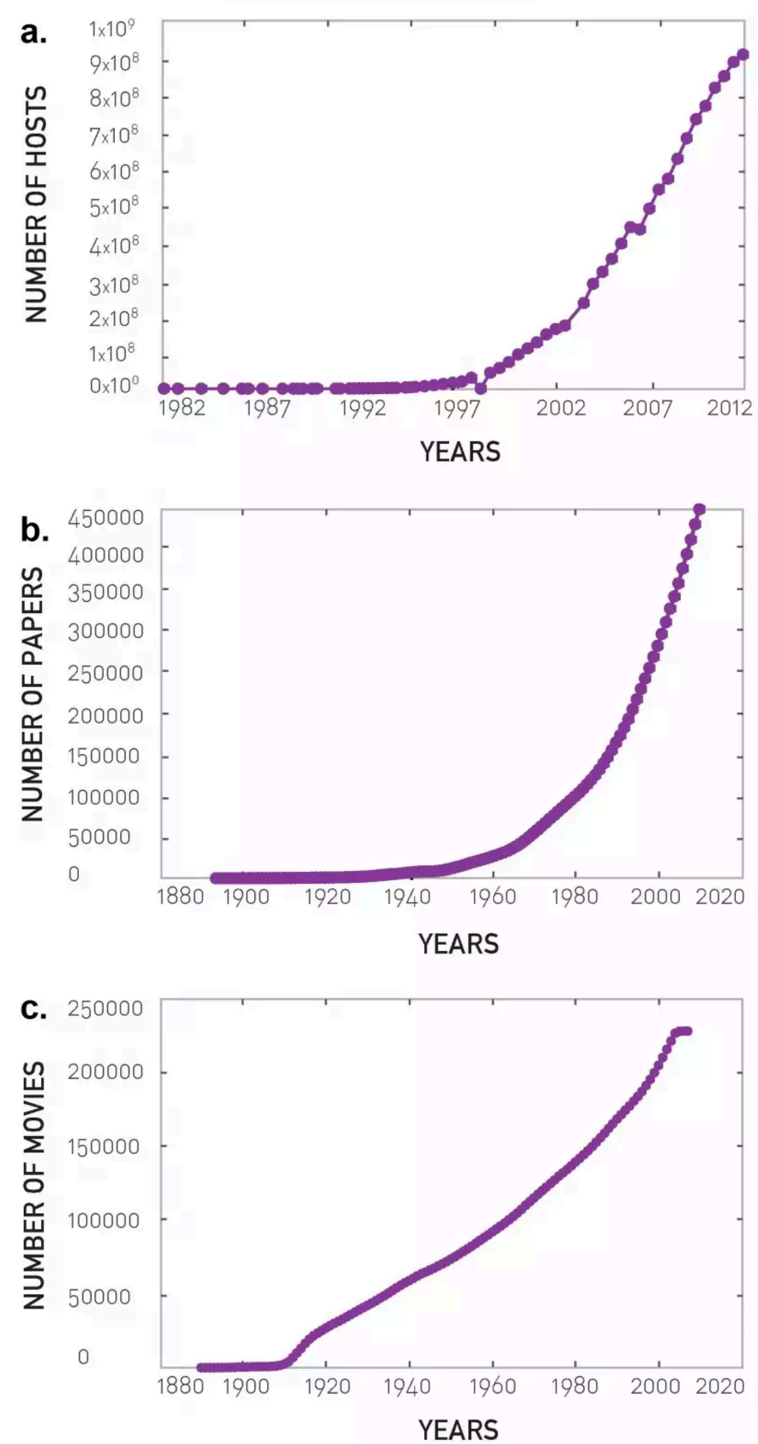
Exercises 3

The Free-Scale Property

Preferential Attachment

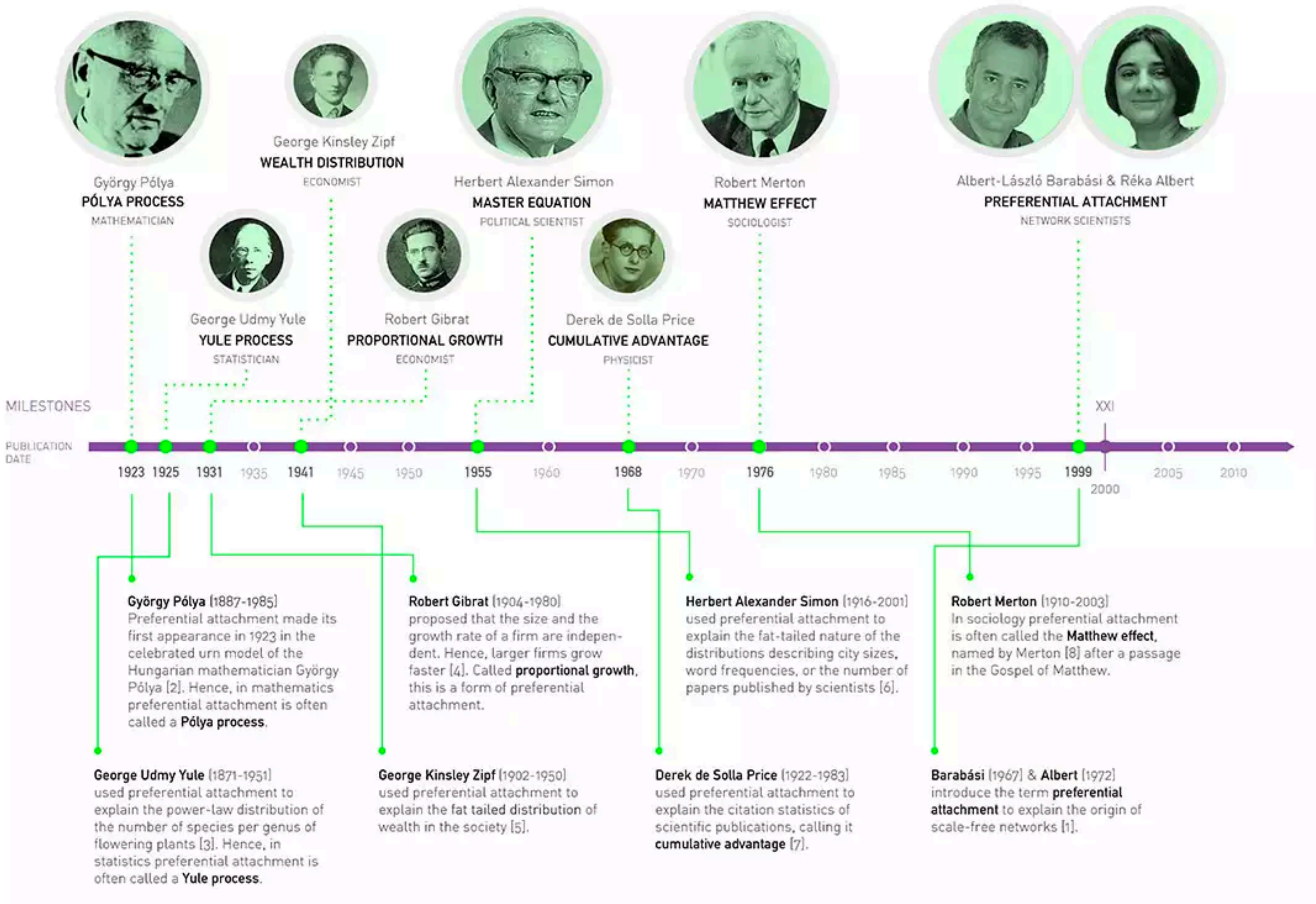


The Growth of Networks

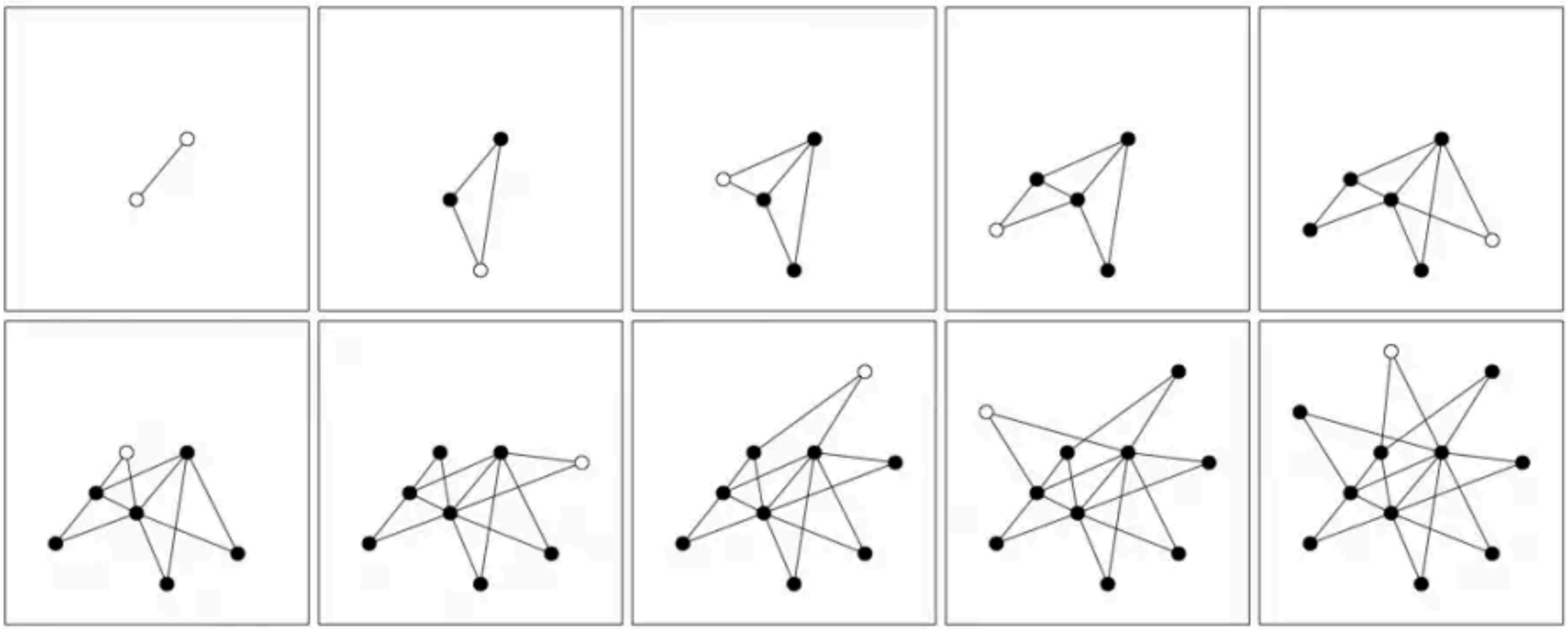


"For whosoever hath, to him shall be given, and he shall have more abundance: but whosoever hath not, from him shall be taken away even that he hath."

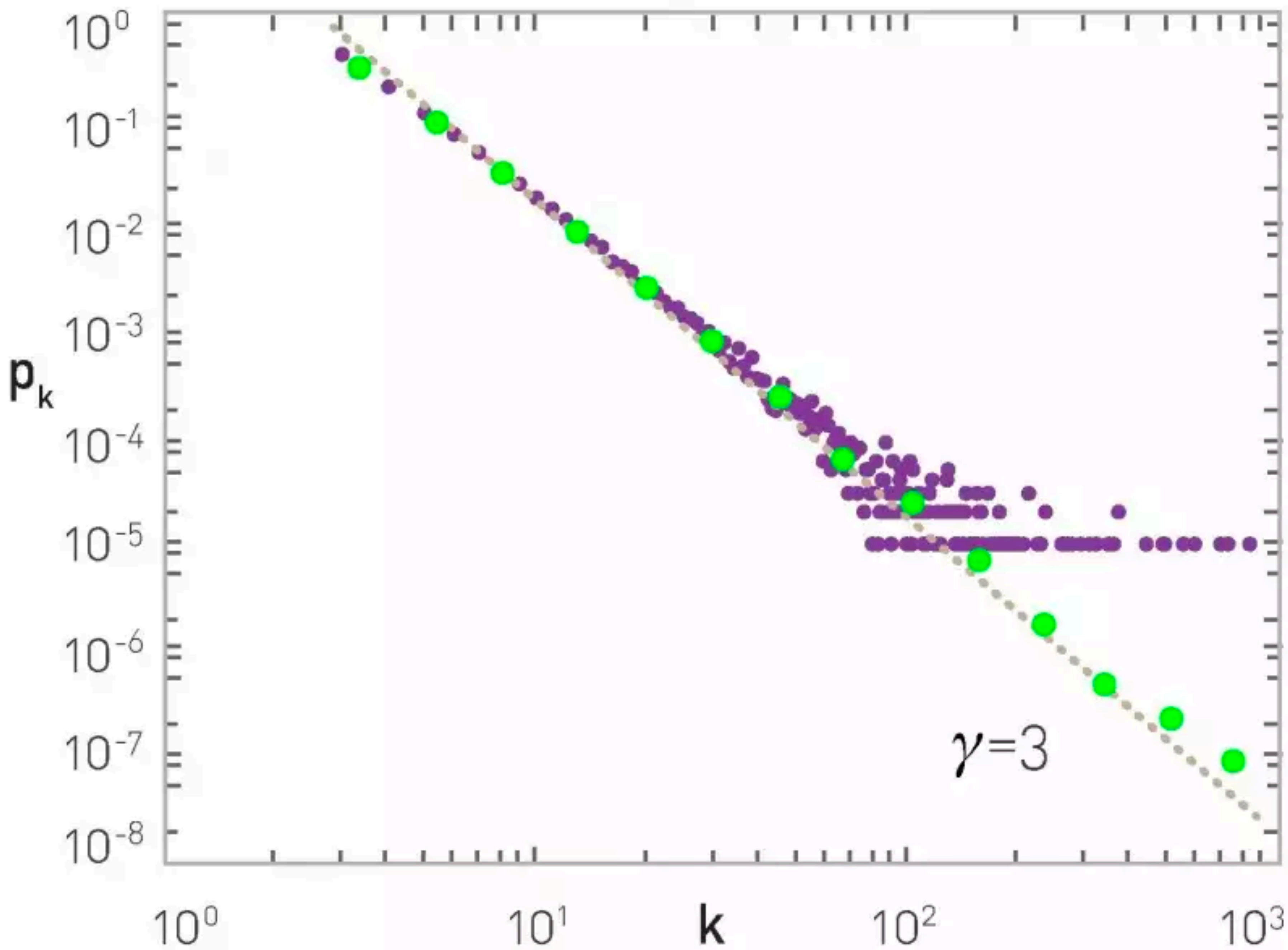
Preferential Attachment: A brief history



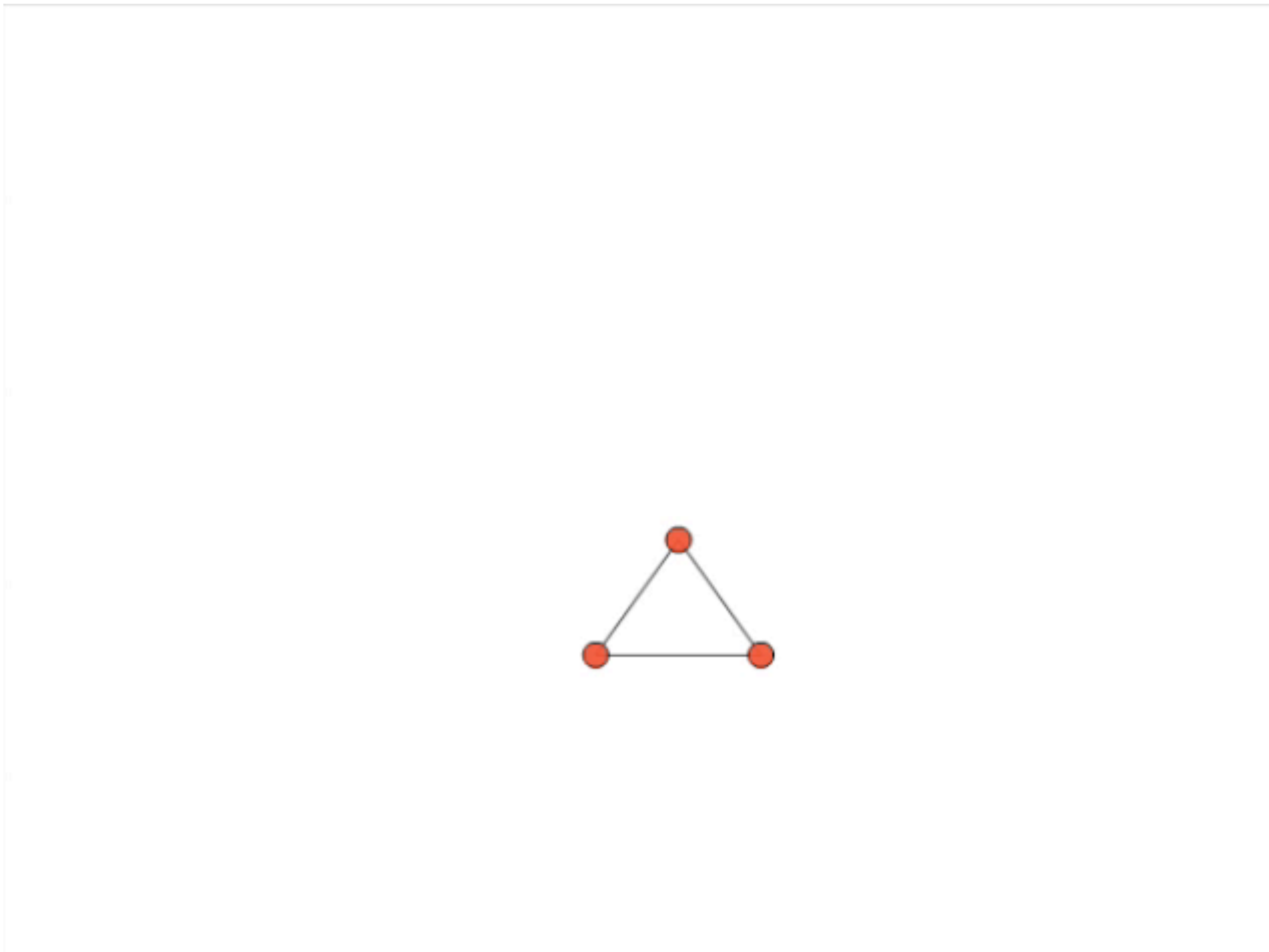
Evolution of the Barabási-Albert Model



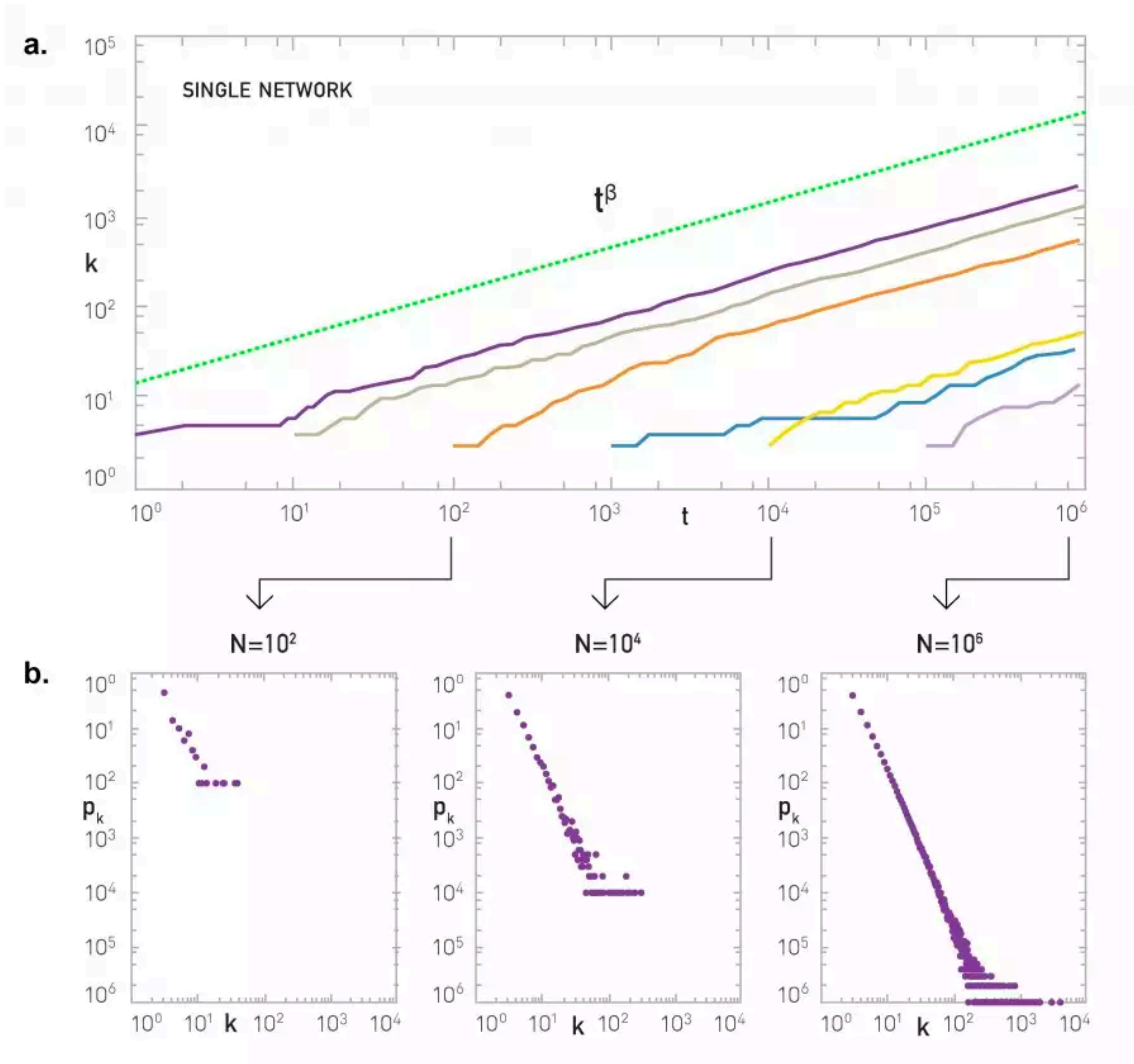
Degree Distribution



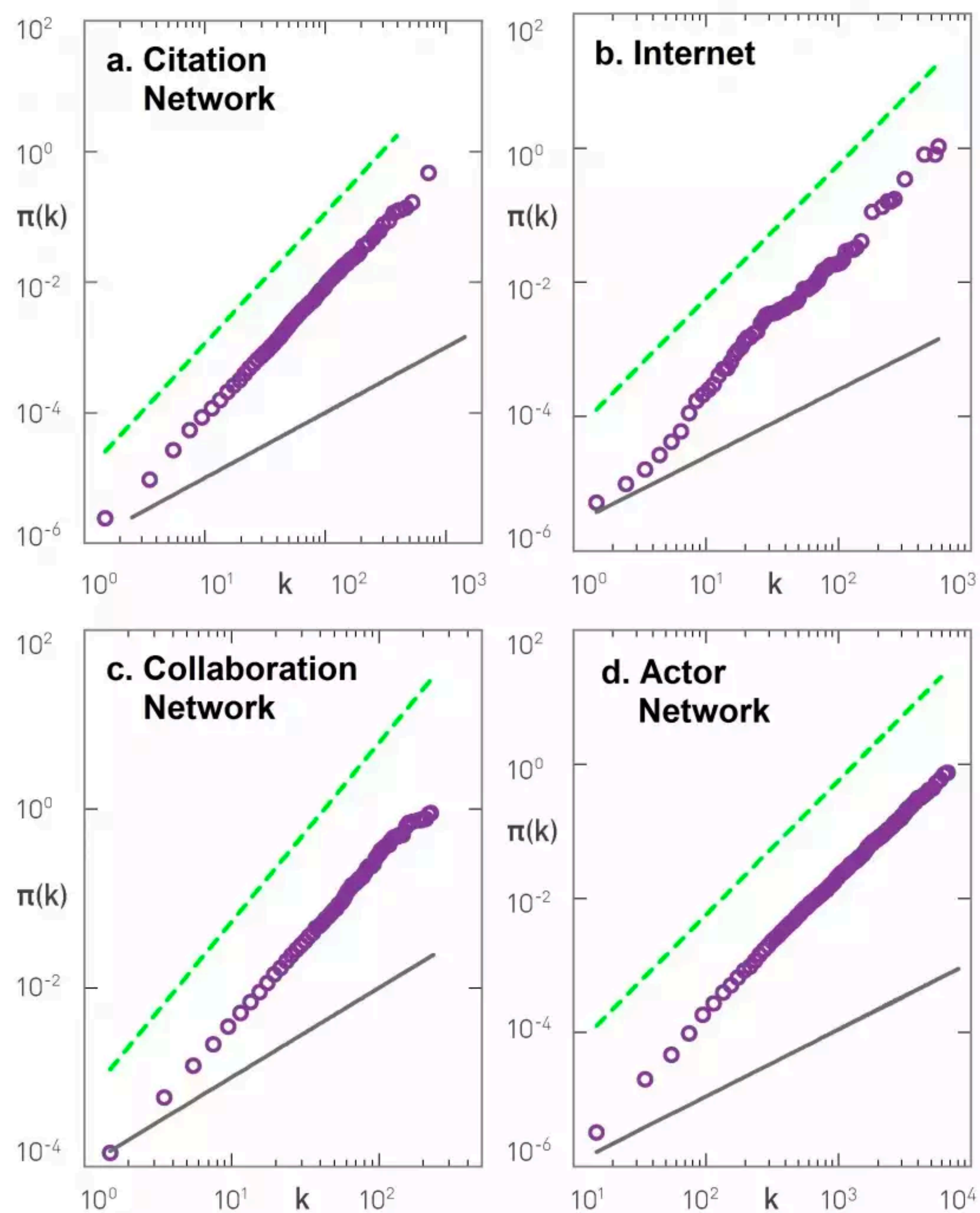
Emergence of a Scale-Free Network



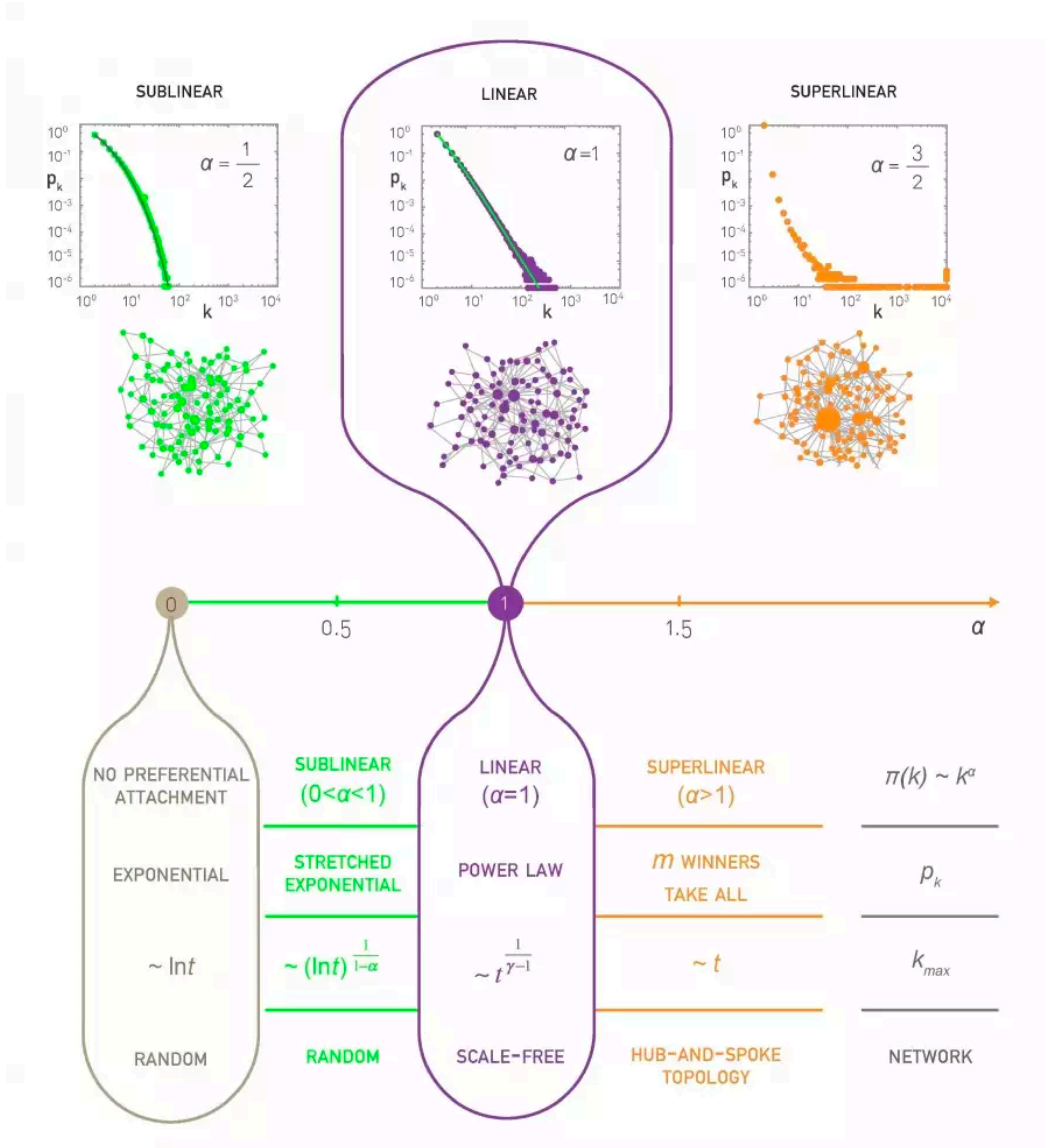
Degree Dynamics



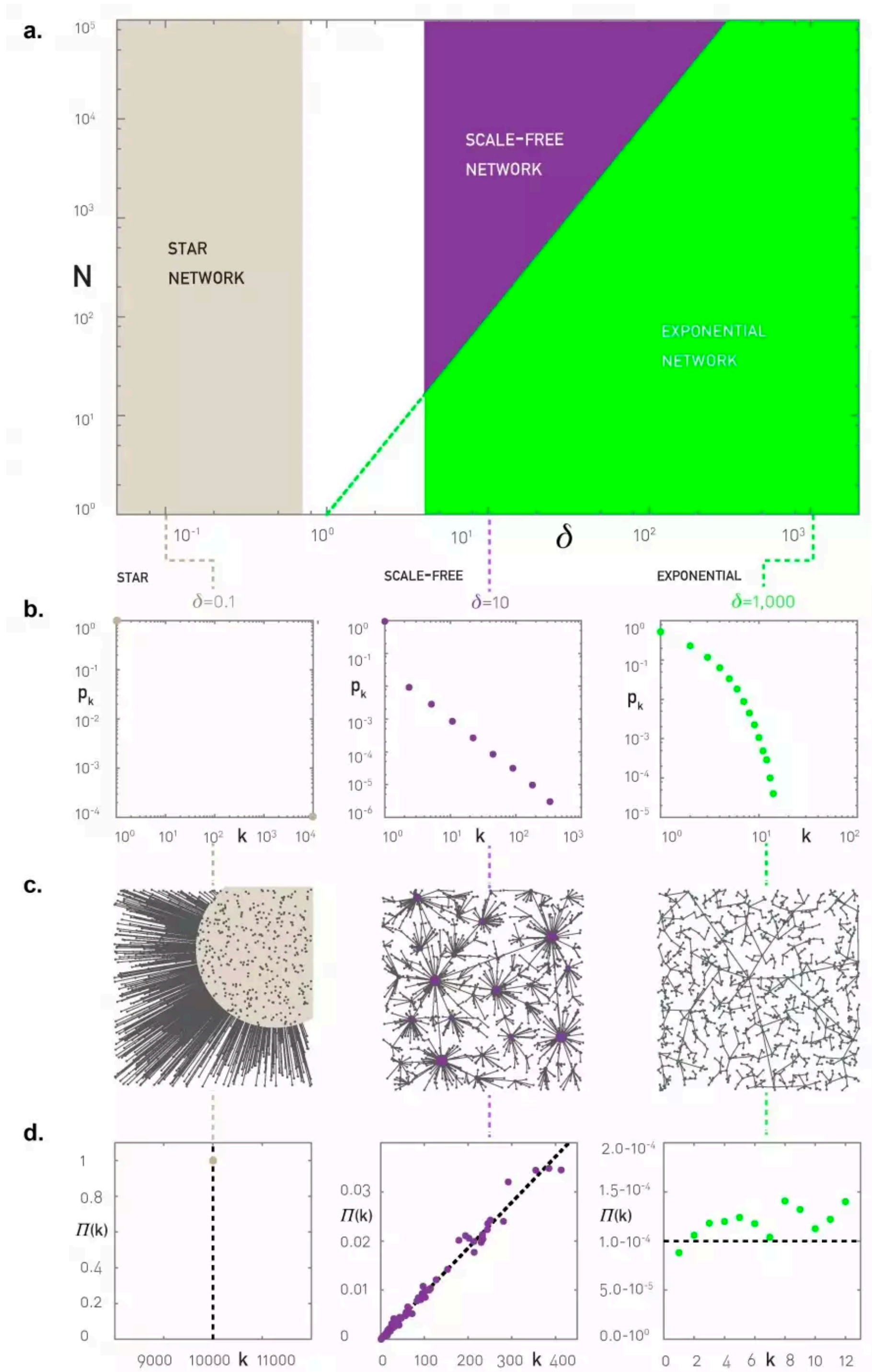
Evidence of Preferential Attachment



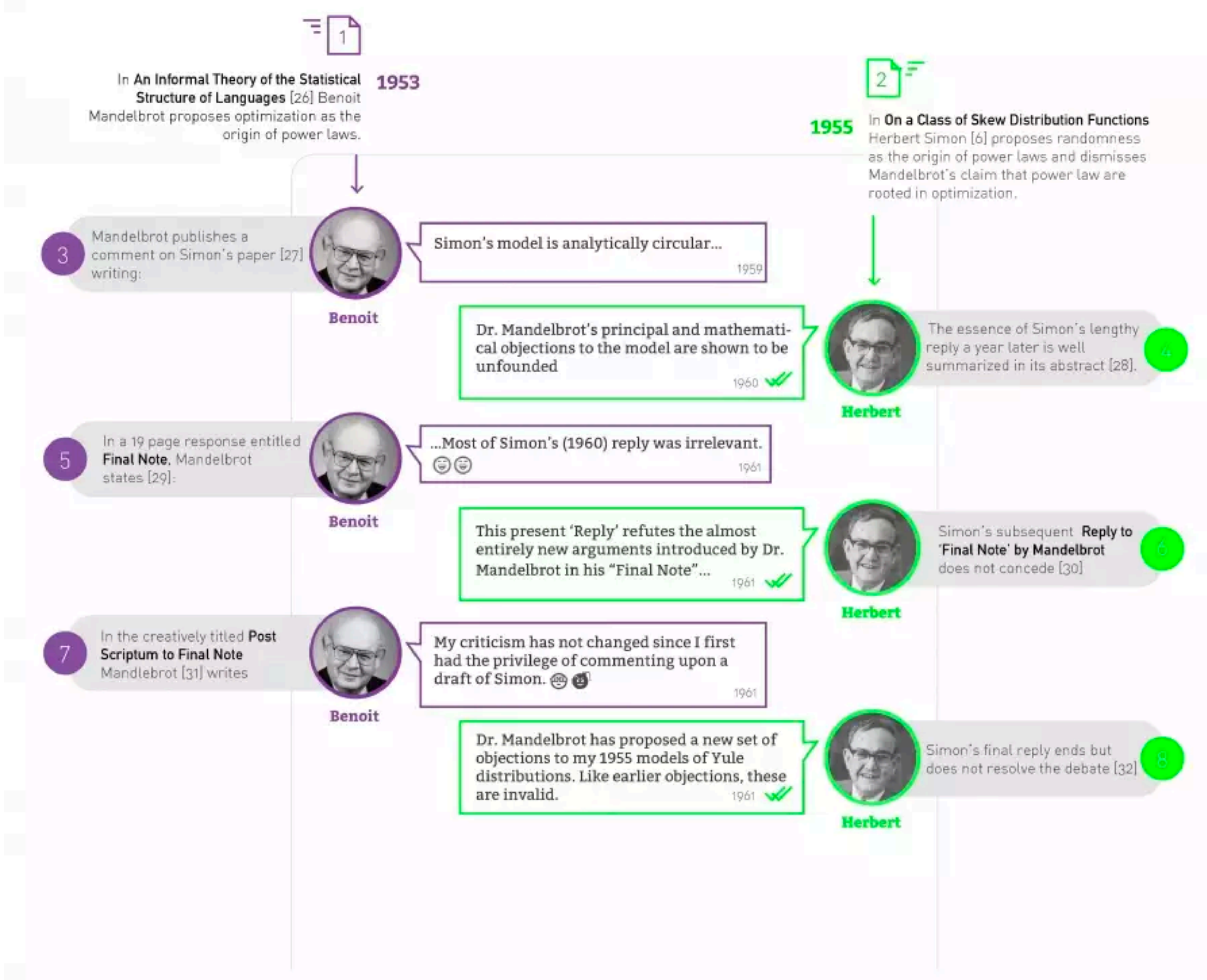
Nonlinear Preferential Attachment



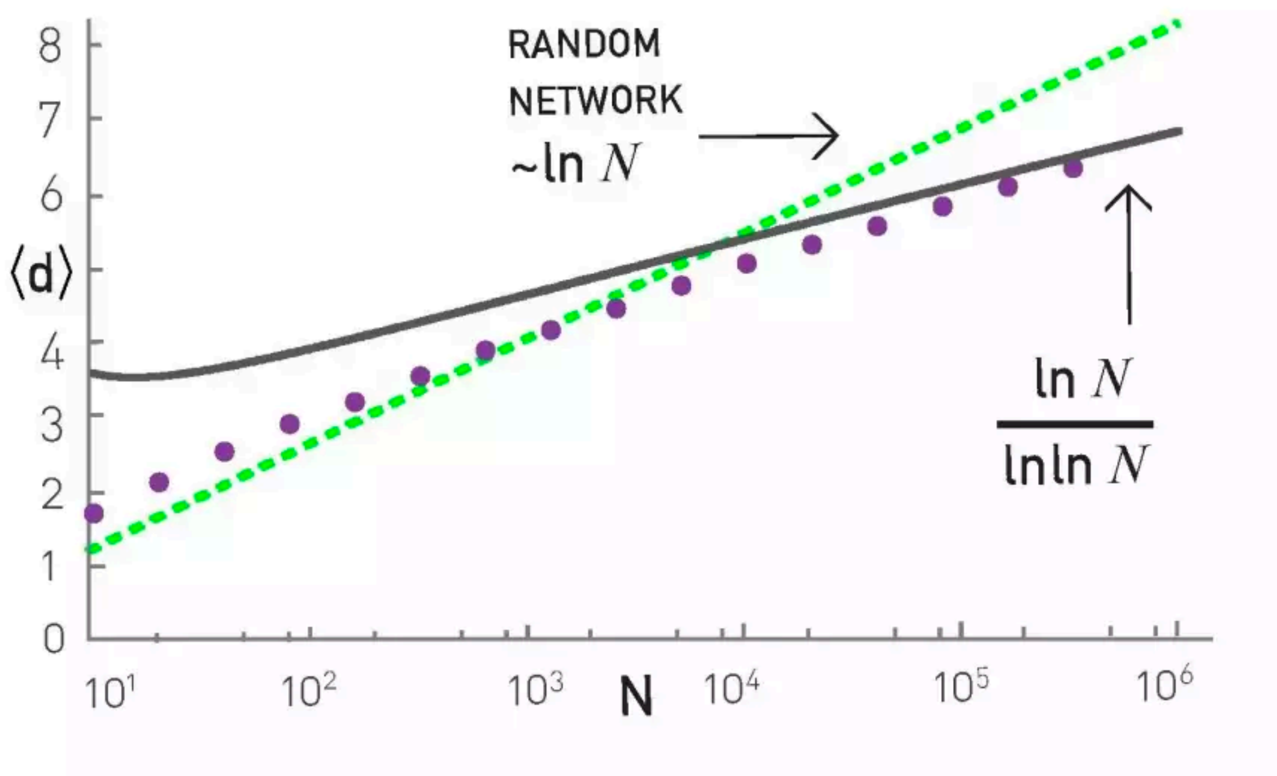
Scaling in the Optimization Model

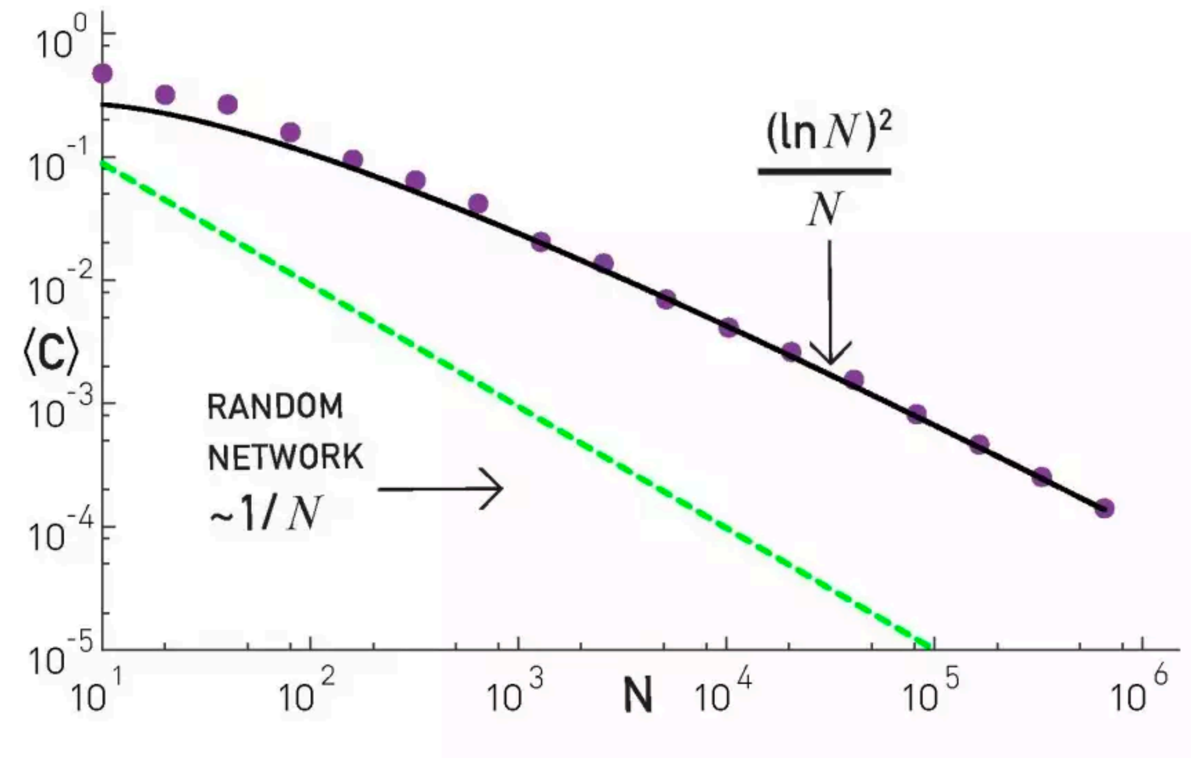


Luck or Reason: An Ancient Fight



Clustering Coefficient





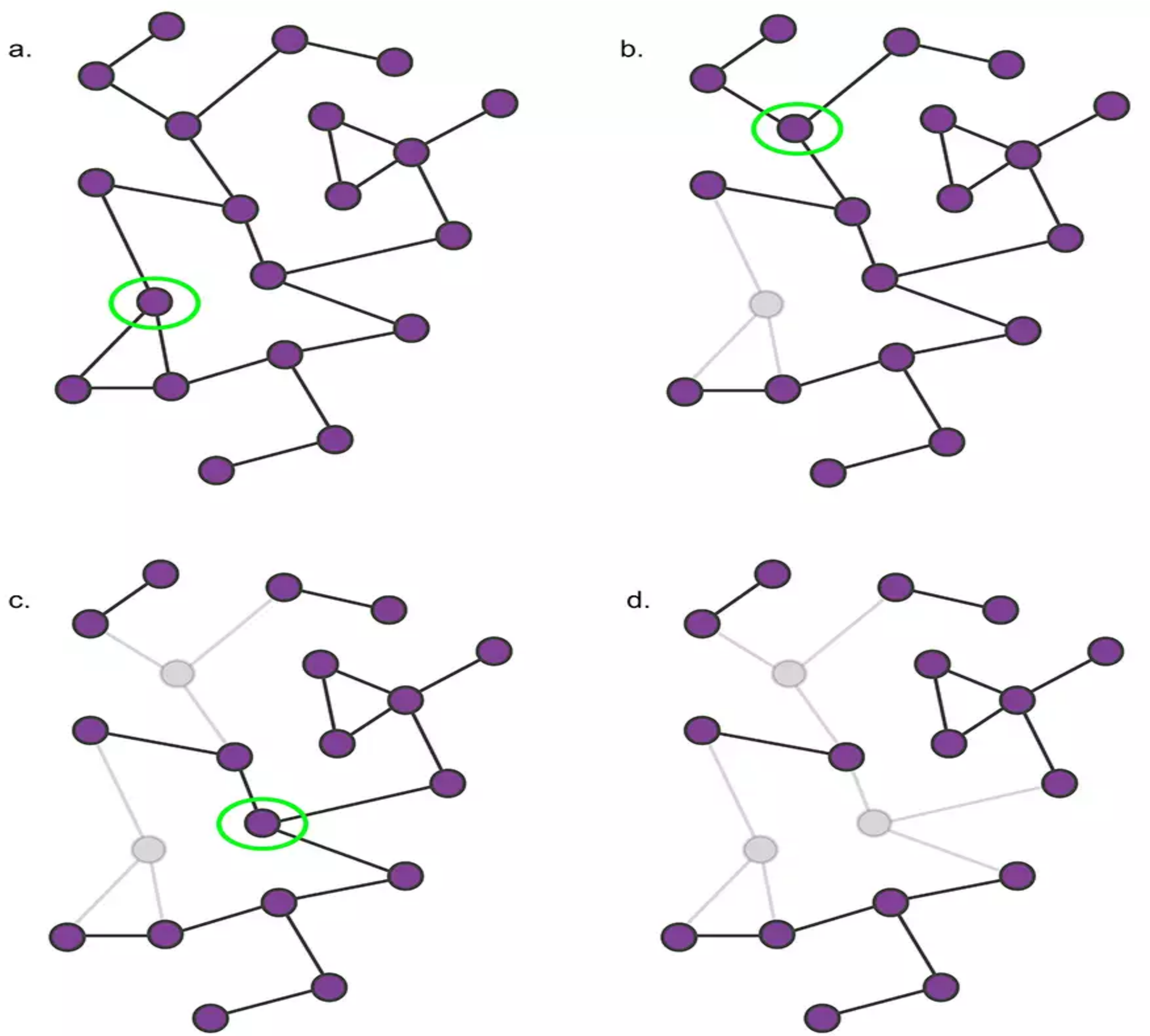
"Robust" comes from the latin Quercus Robur, meaning oak, the symbol of strength and longevity in the ancient world.

4. Network robustness

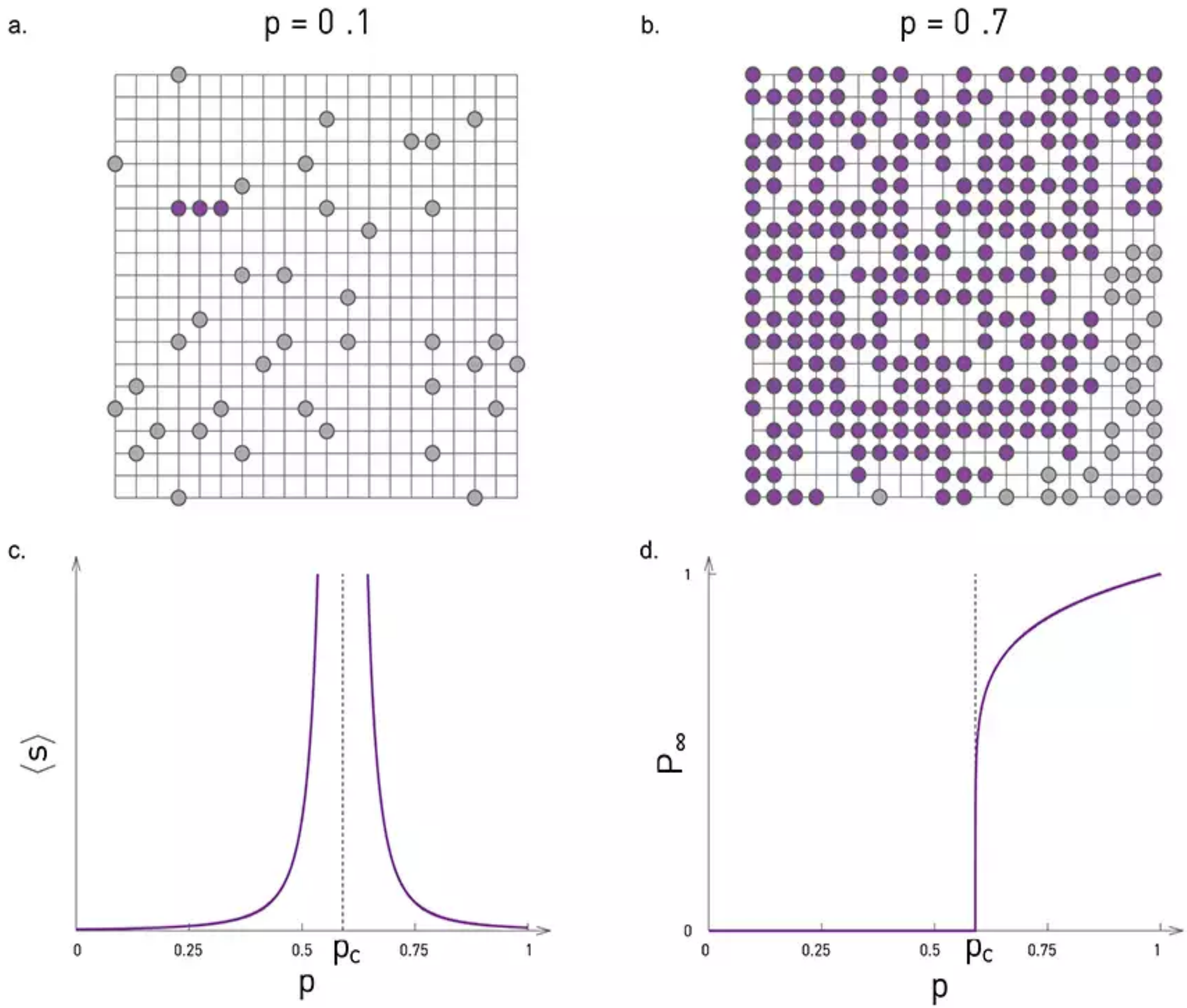
Or Why (Real) Networks are the way they are

The Achille's Heel of Complex Networks

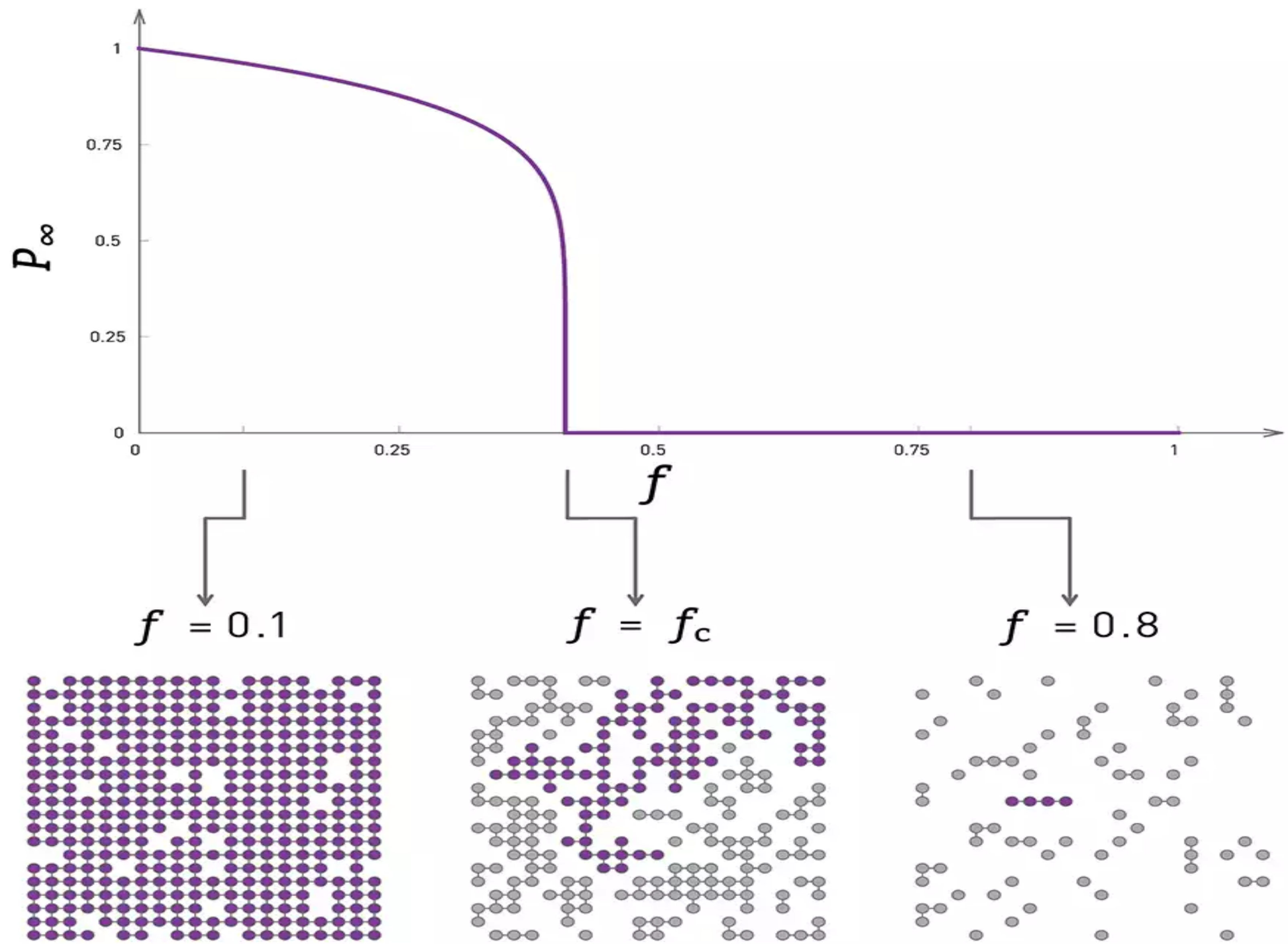
The Impact of Node Removal



Percolation



Network Breakdown as Inverse Percolation



$0 < f < f_c$:

There is a giant component.

$P_\infty \sim |f - f_c|^\beta$

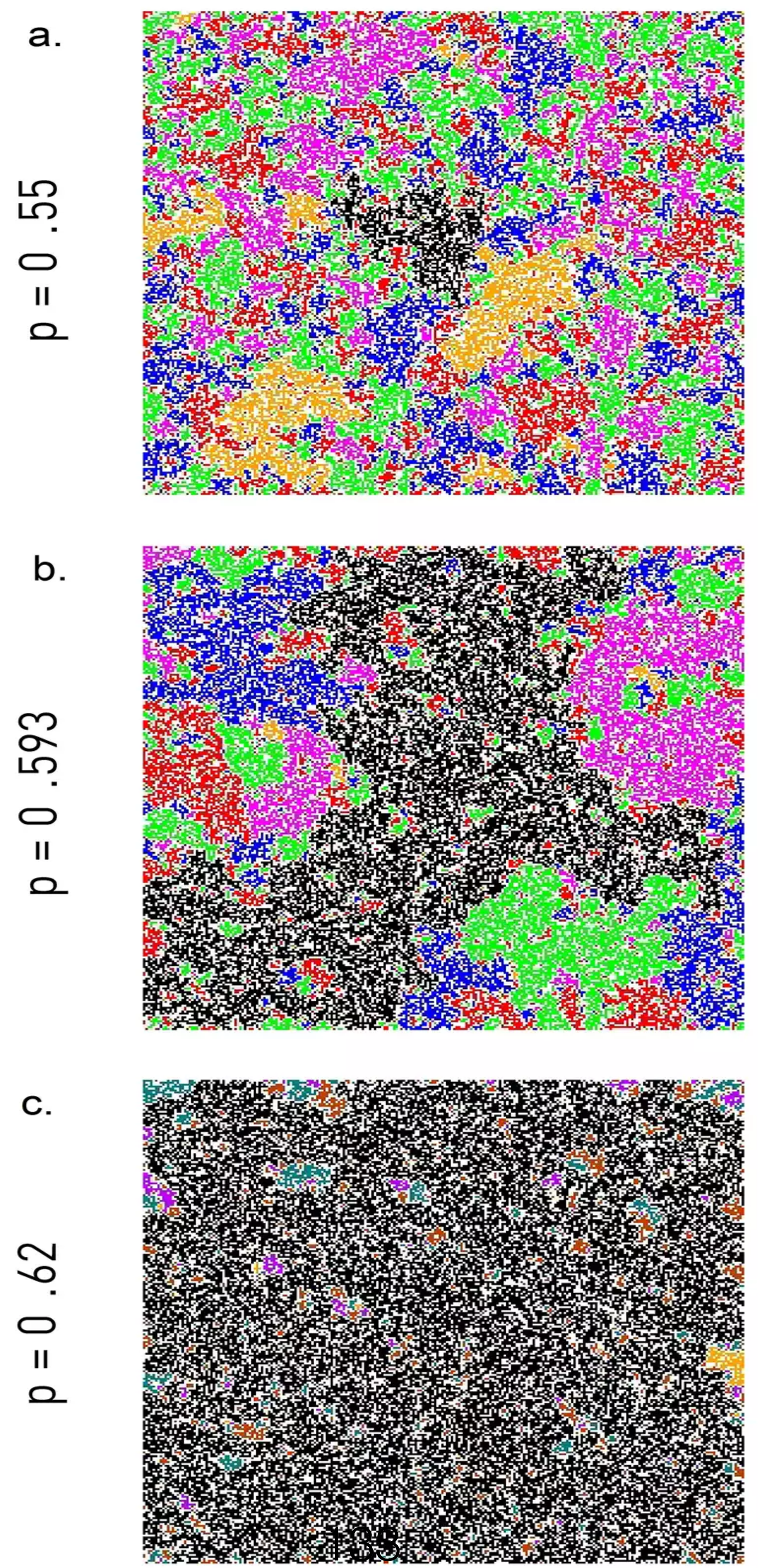
$f = f_c$:

The giant component vanishes.

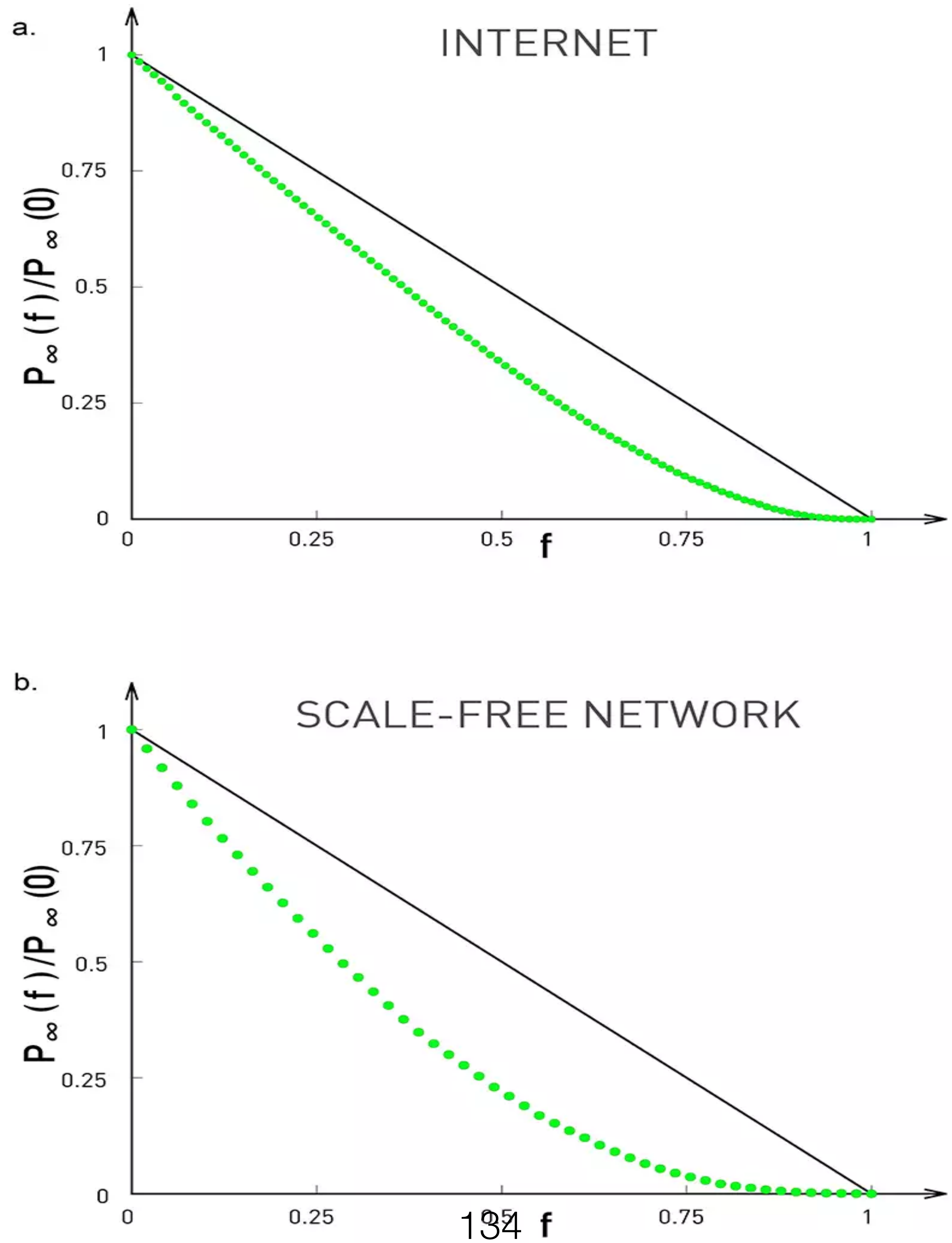
$f > f_c$:

The lattice breaks into many tiny components.

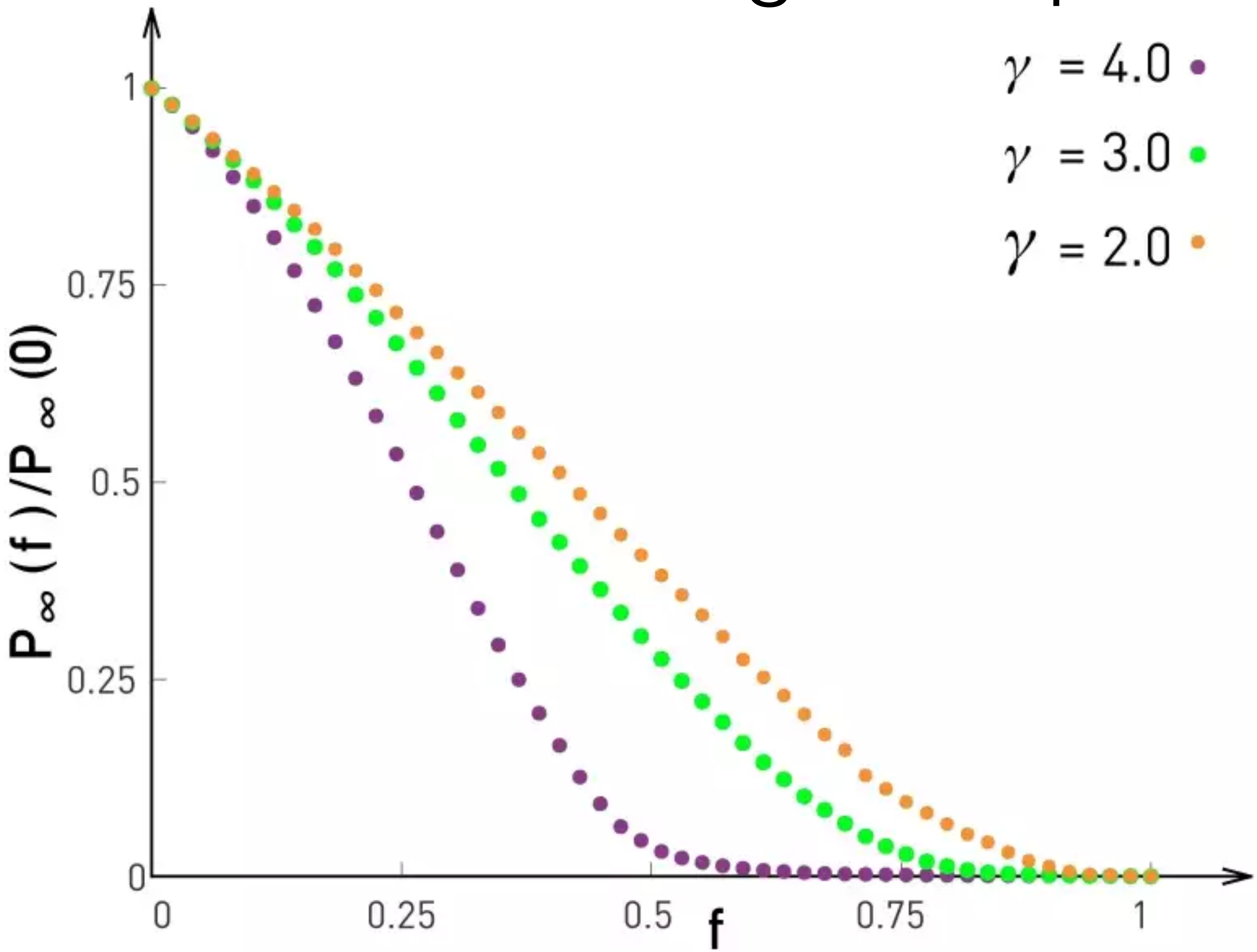
From Forest Fires to Percolation Theory



Robustness of Scale-Free Networks



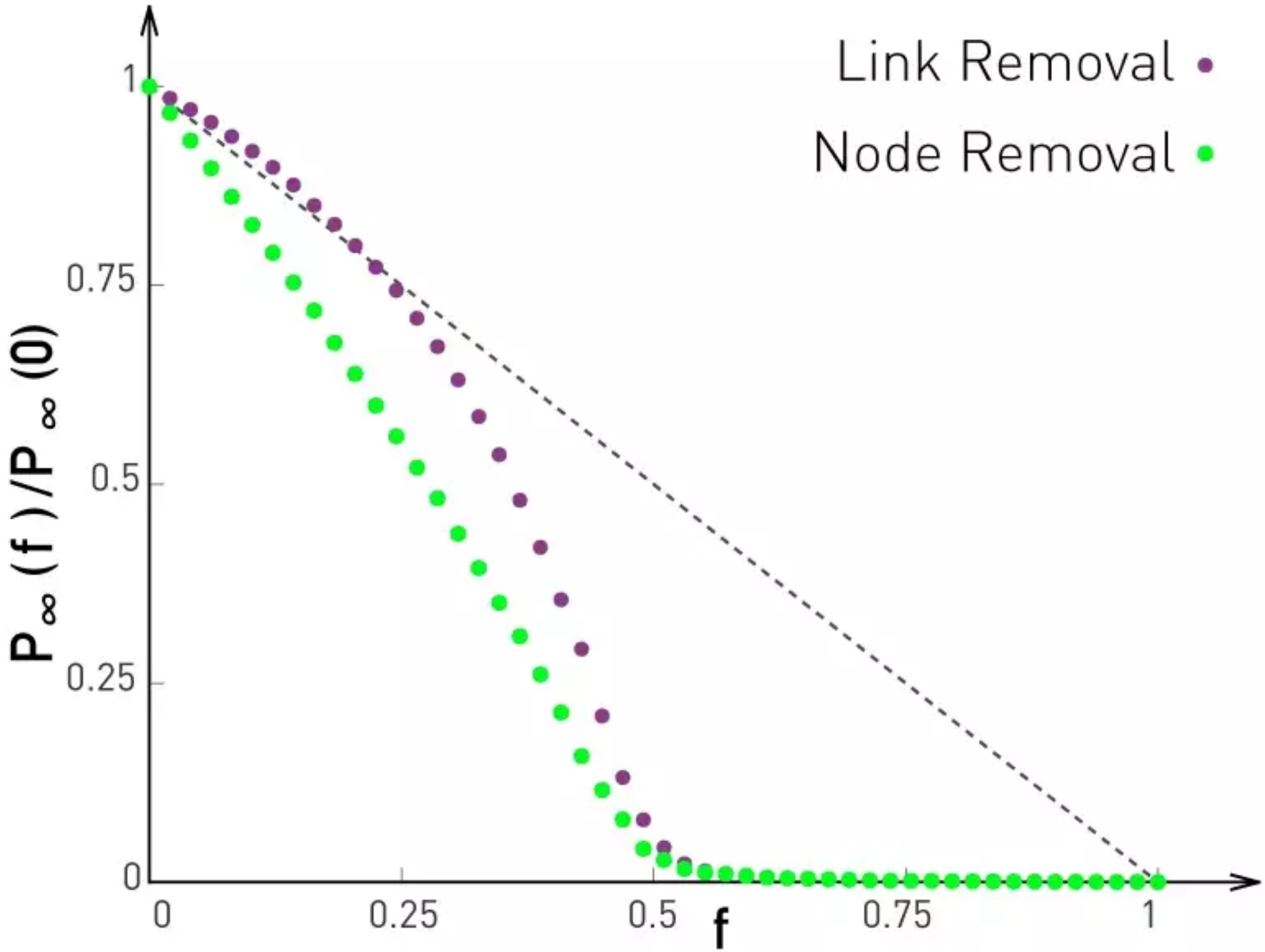
Robustness and Degree Exponent



Scale-Free Network under Nodes Failures

http://barabasi.com/networksciencebook/images/ch-08/video-8-1.webm

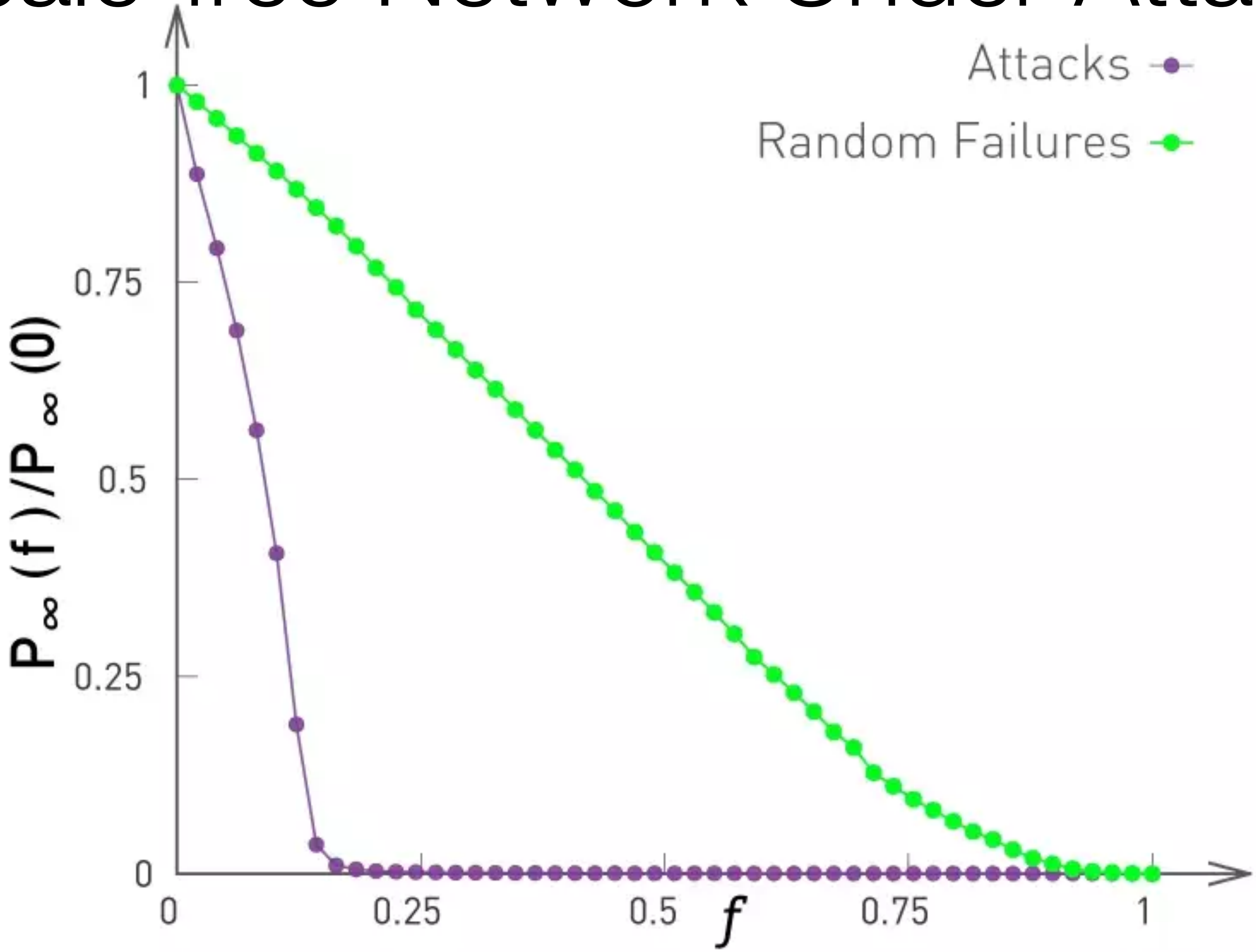
Robustness and Link Removal

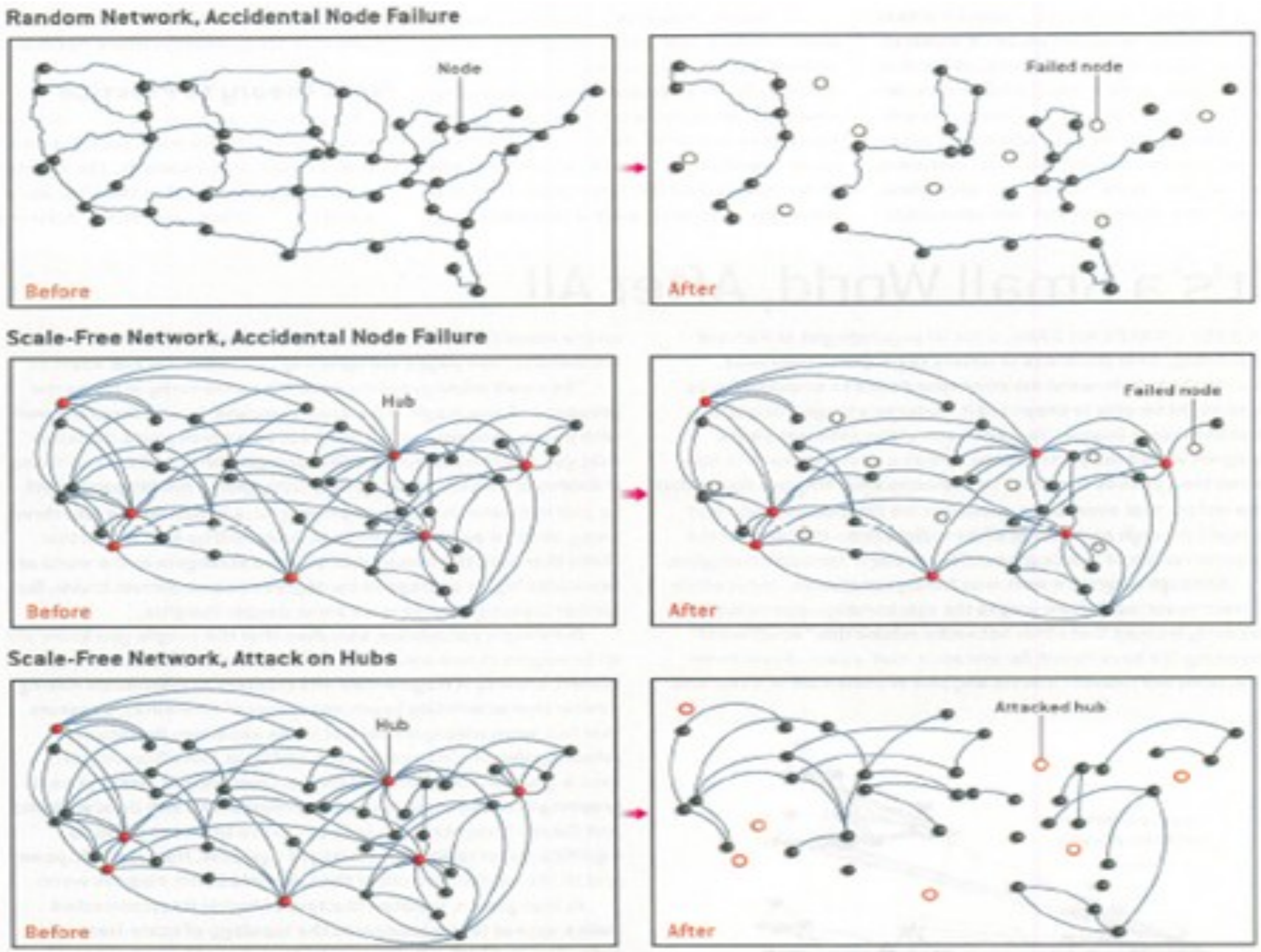


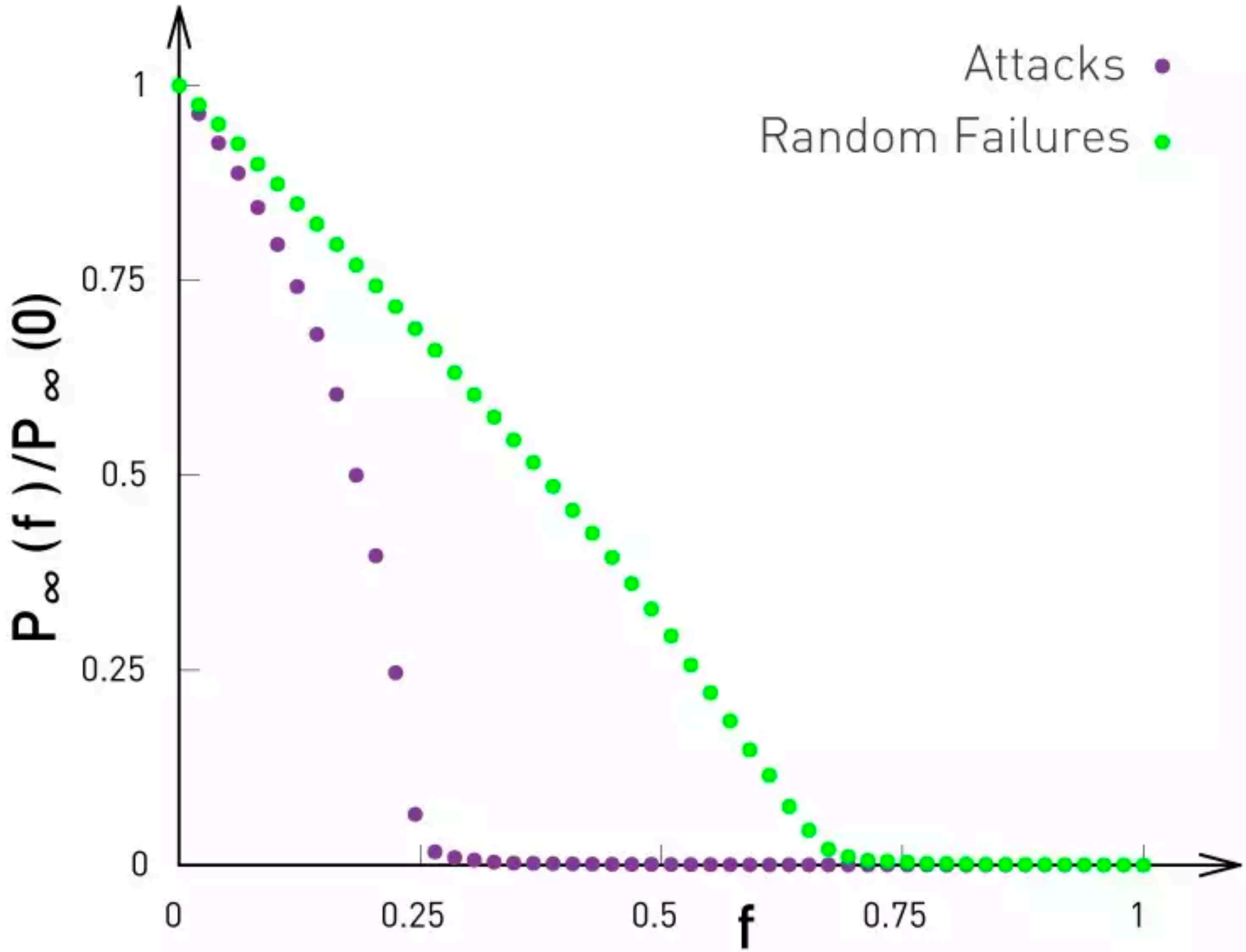
Breakdown Thresholds Under Random Failures and Attacks

Network	Random Failures (Real Network)	Random Failures (Randomized Network)	Attack (Real Network)
Internet	0.92	0.84	0.16
WWW	0.88	0.85	0.12
Power Grid	0.61	0.63	0.2
Mobile Phone Calls	0.78	0.68	0.2
Email	0.92	0.69	0.04
Science Collaboration	0.92	0.88	0.27
Actor Network	0.98	0.99	0.55
Citation Network	0.96	0.95	0.76
E. Coli Metabolism	0.96	0.9	0.49
Protein Interactions	0.88	0.66	0.06

Scale-free Network Under Attack




Random Network, Accidental Node Failure
Node
Before
Failed node
After
Scale-Free Network, Accidental Node Failure
Hub
Before
Failed node
After
Scale-Free Network, Attack on Hubs
Hub
Before
Attacked hub
After



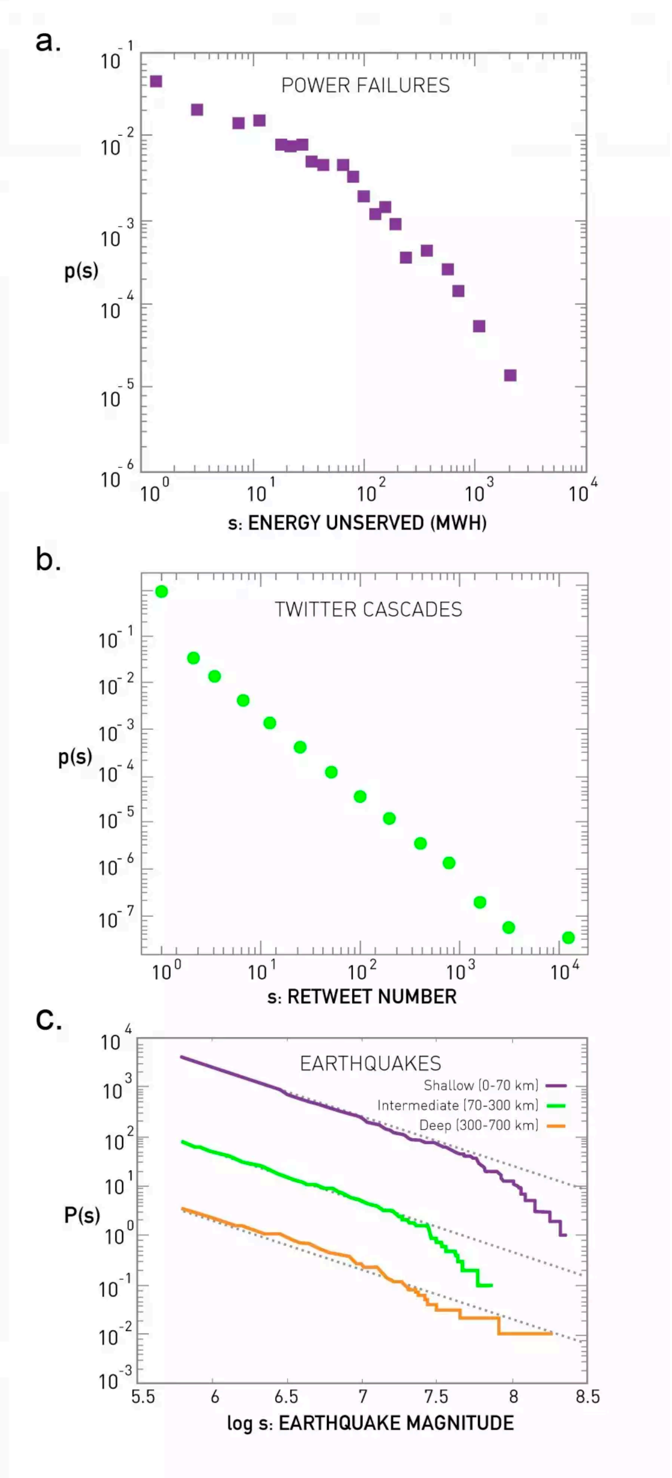
Attacks
Random Failures
$P_\infty(f)/P_\infty(0)$
1
0.75
0.5
0.25
0
0
0.25
0.5
f
0.75
1

Avalanche Exponents in Real Systems

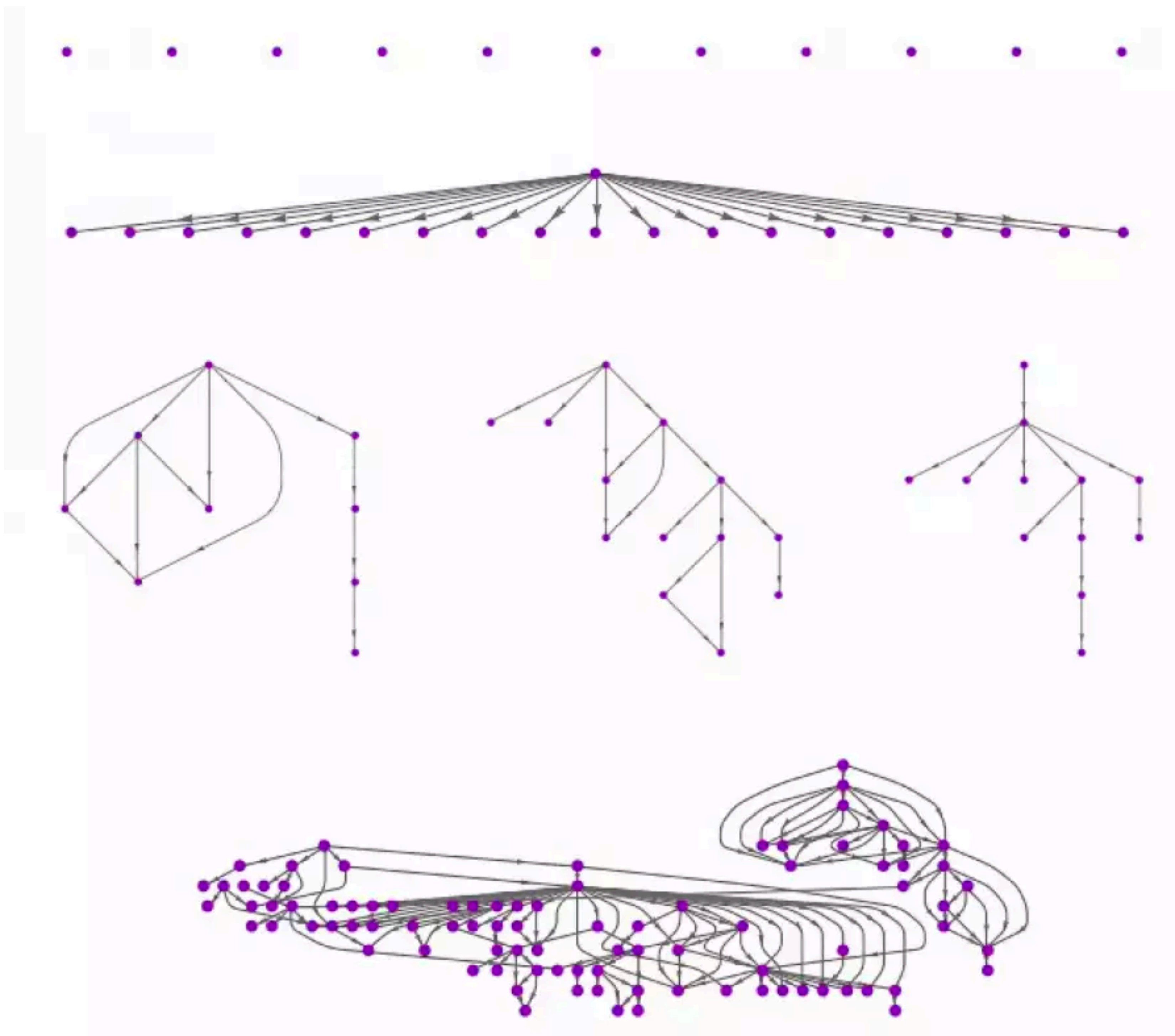
Source	Exponent	Cascade
Power grid (North America)	2.0	Power
Power grid (Sweden)	1.6	Energy
Power grid (Norway)	1.7	Power
Power grid (New Zealand)	1.6	Energy
Power grid (China)	1.8	Energy
Twitter Cascades	1.75	Retweets
Earthquakes	1.67	Seismic Wave

Northeast Blackout of 2003

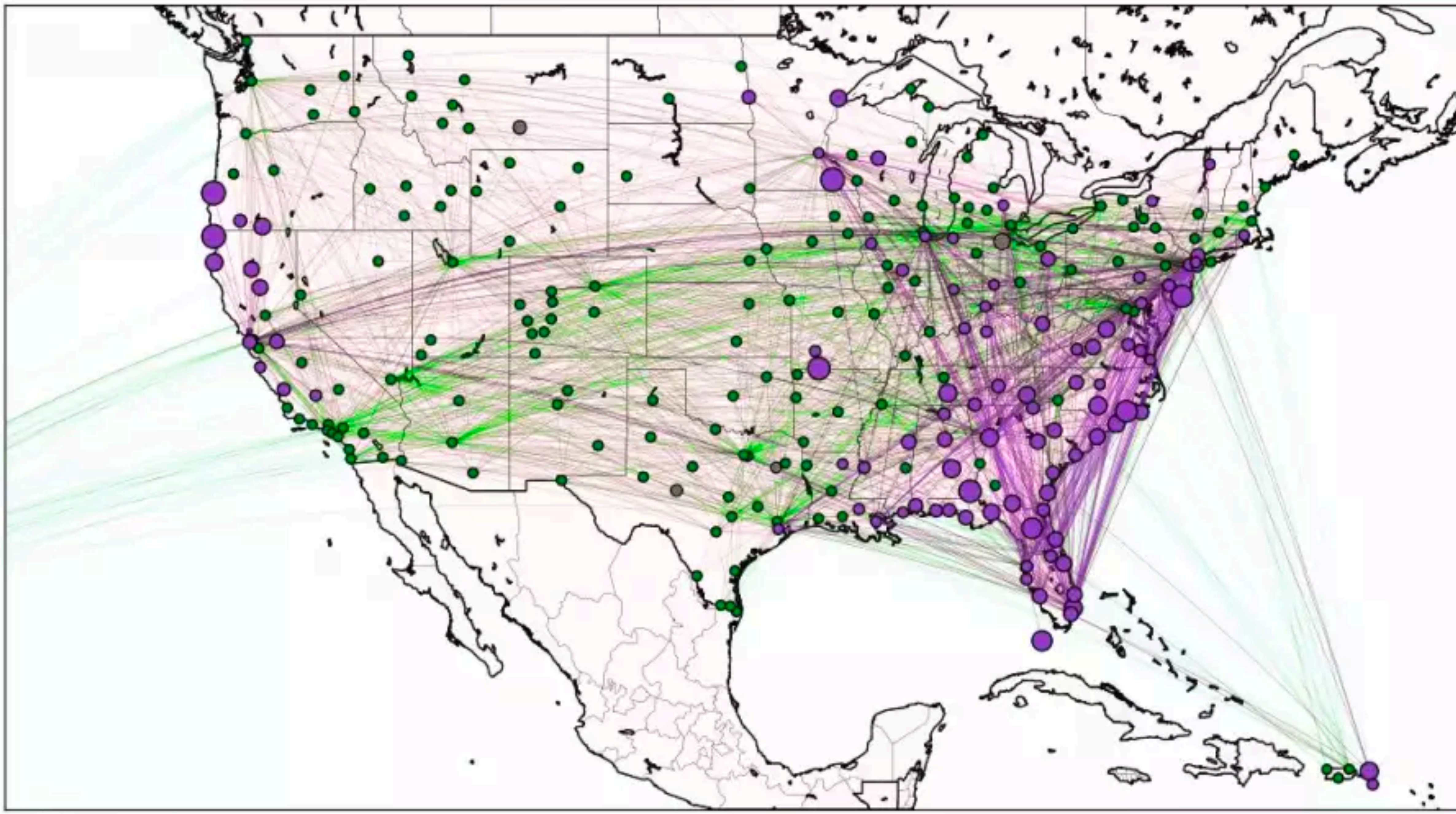
Cascade Size Distribution



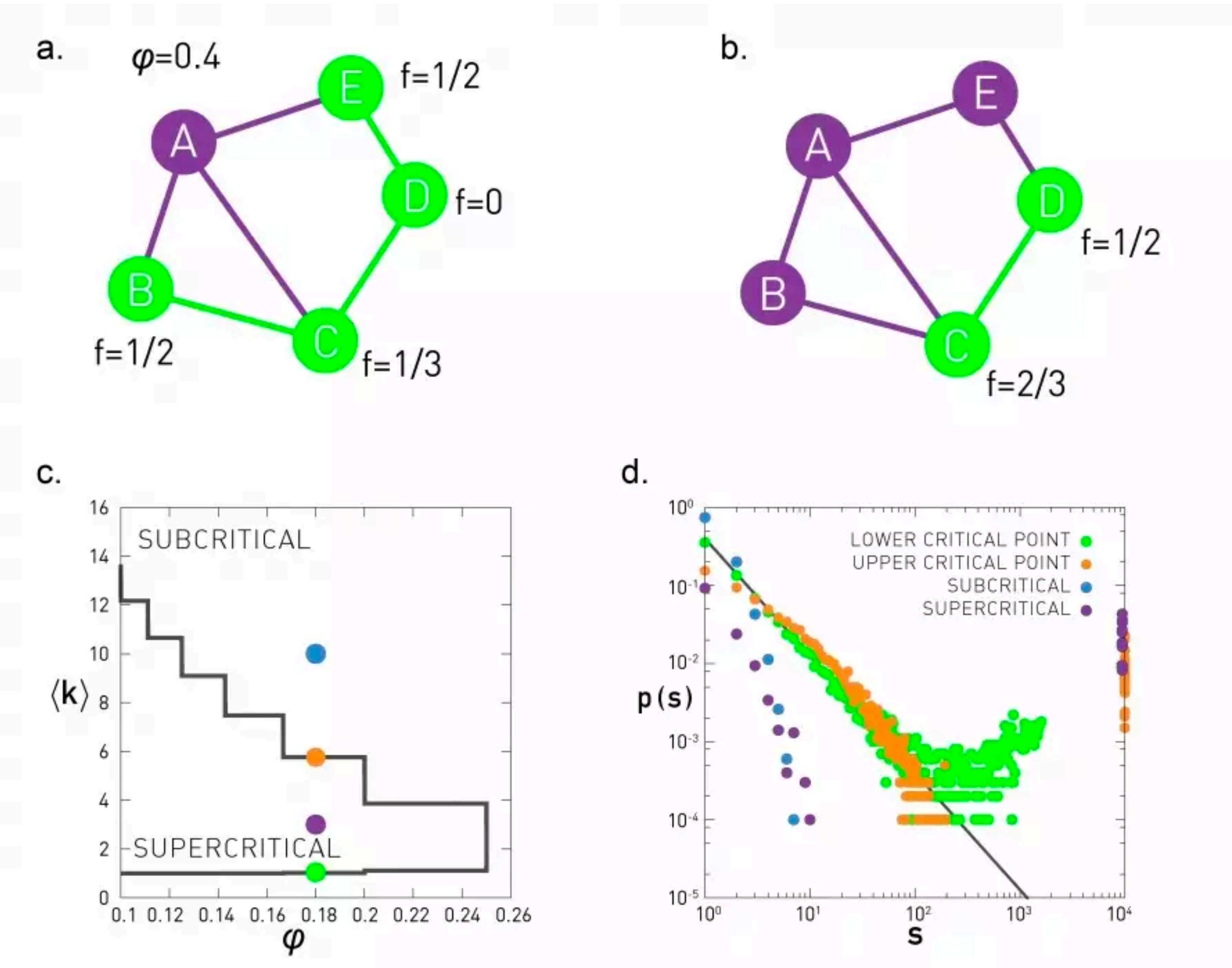
Information Cascades



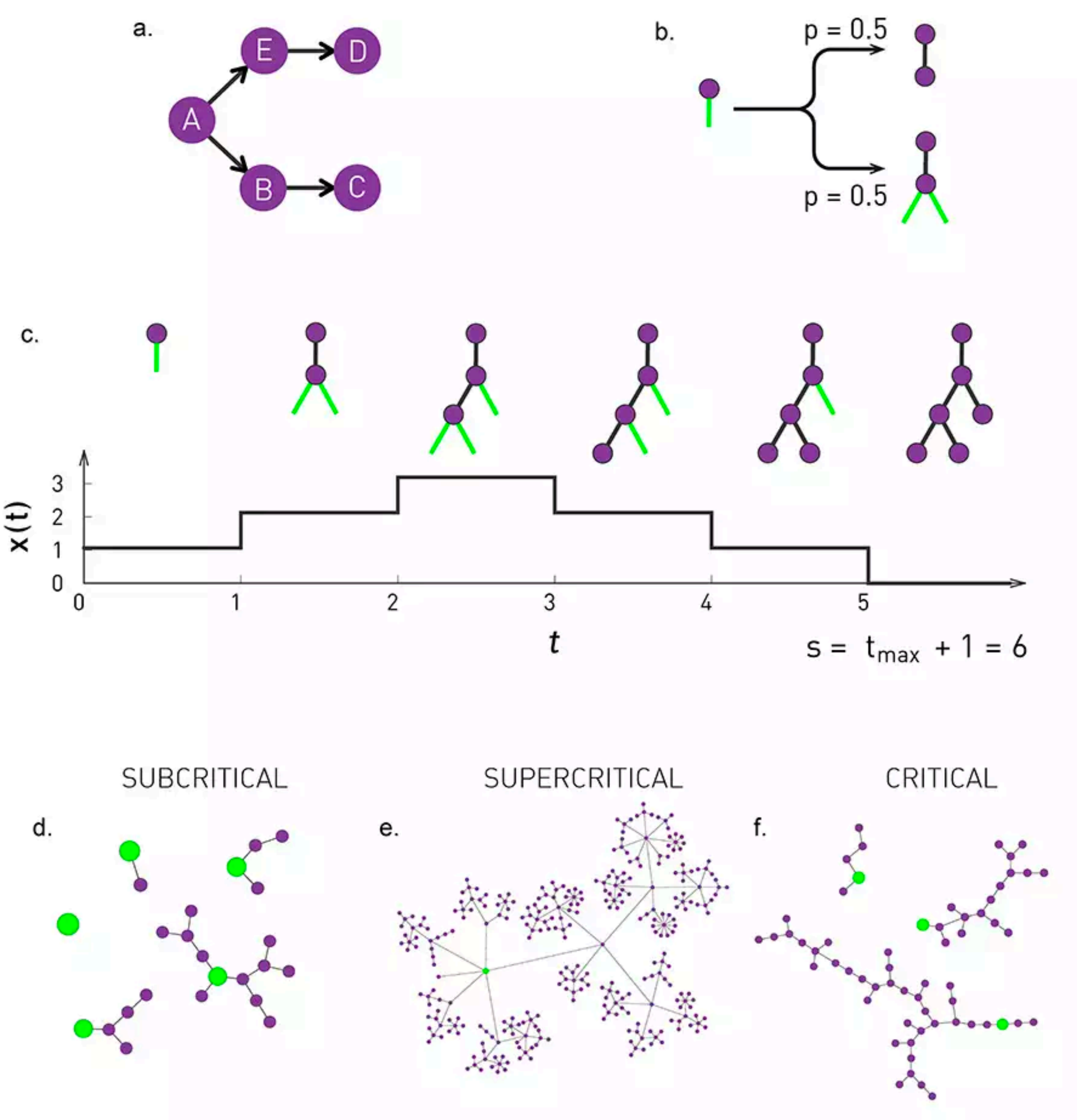
Cluster of Congested Airports



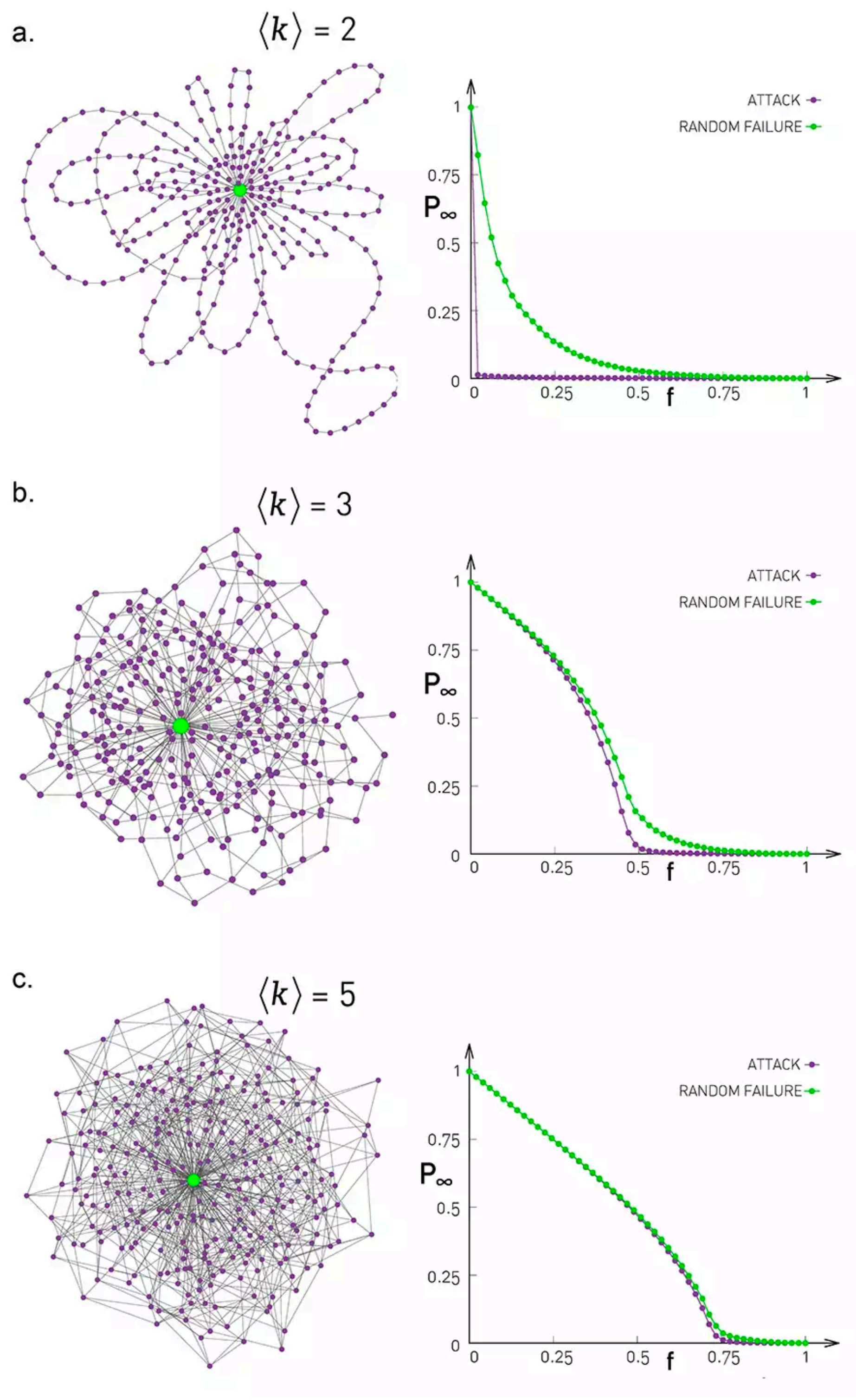
Failure Propagation Model



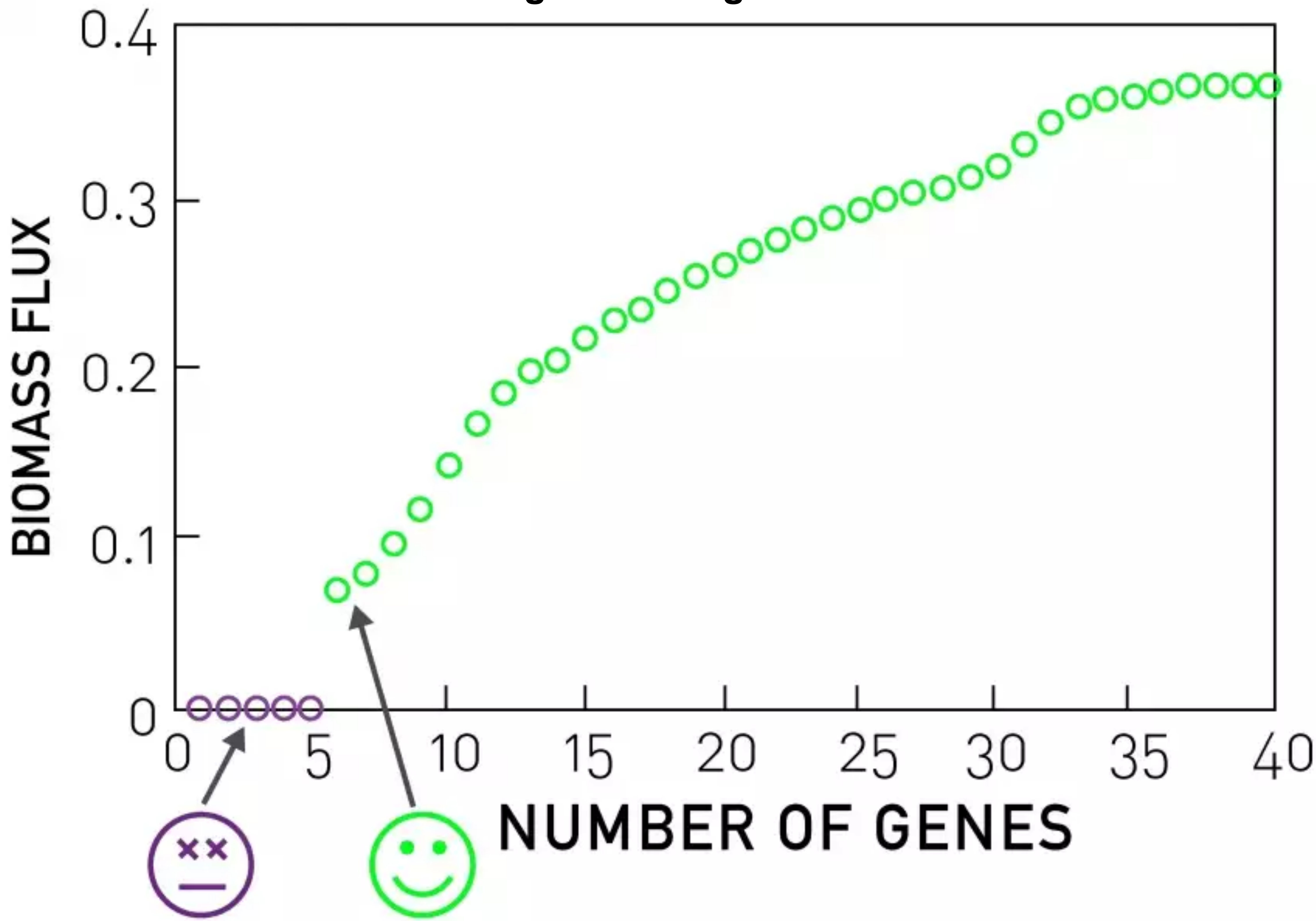
Branching Model



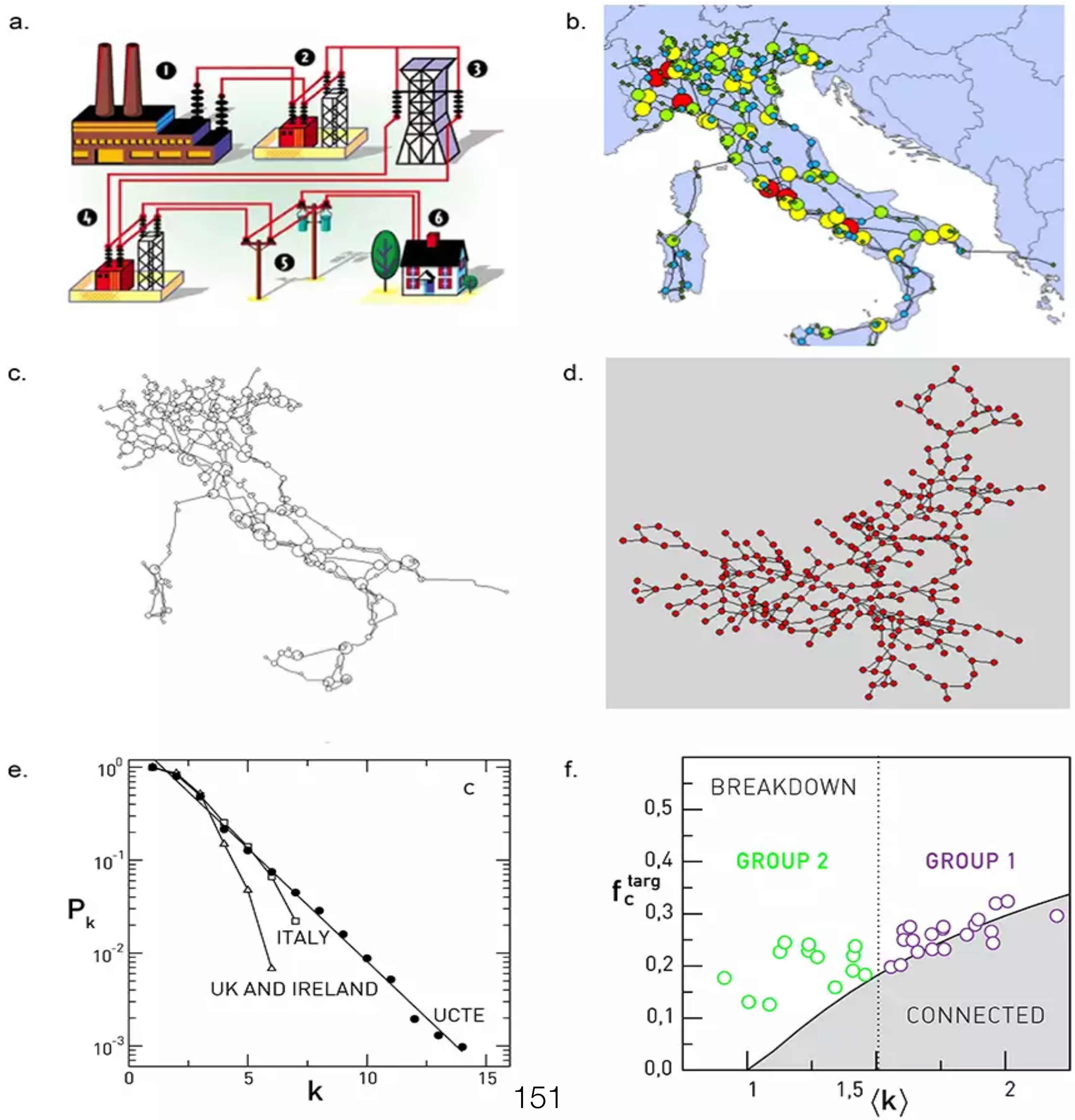
Optimizing Attack and Failure Tolerance



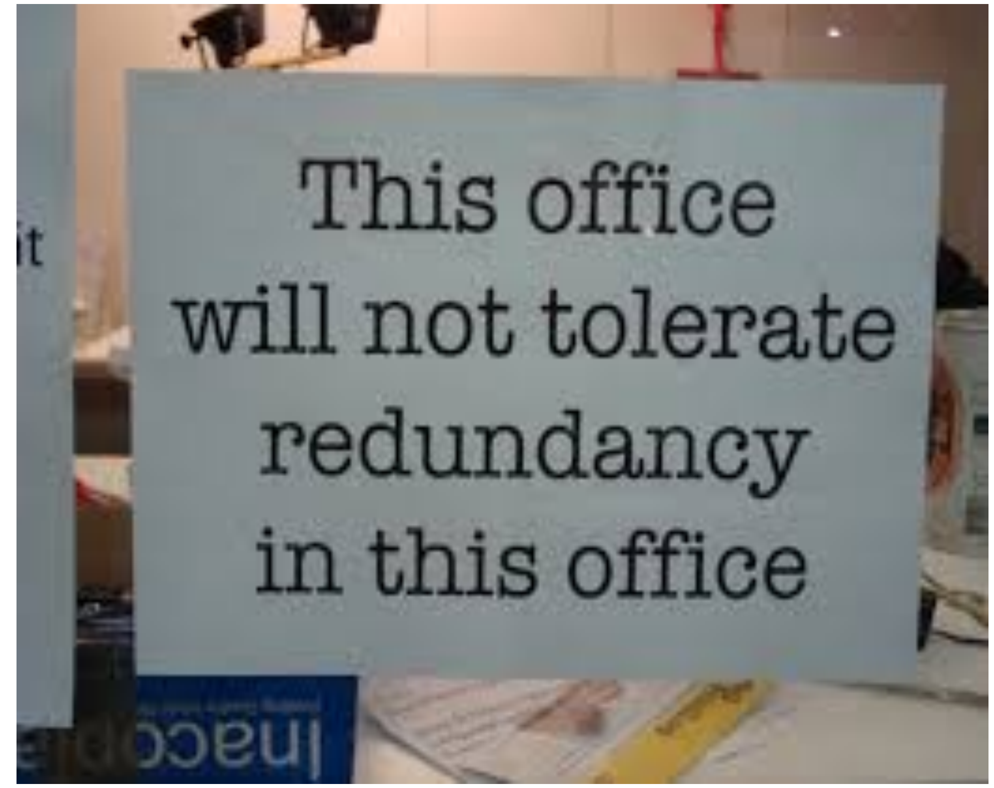
Halting Cascading Failures



The Power Grid

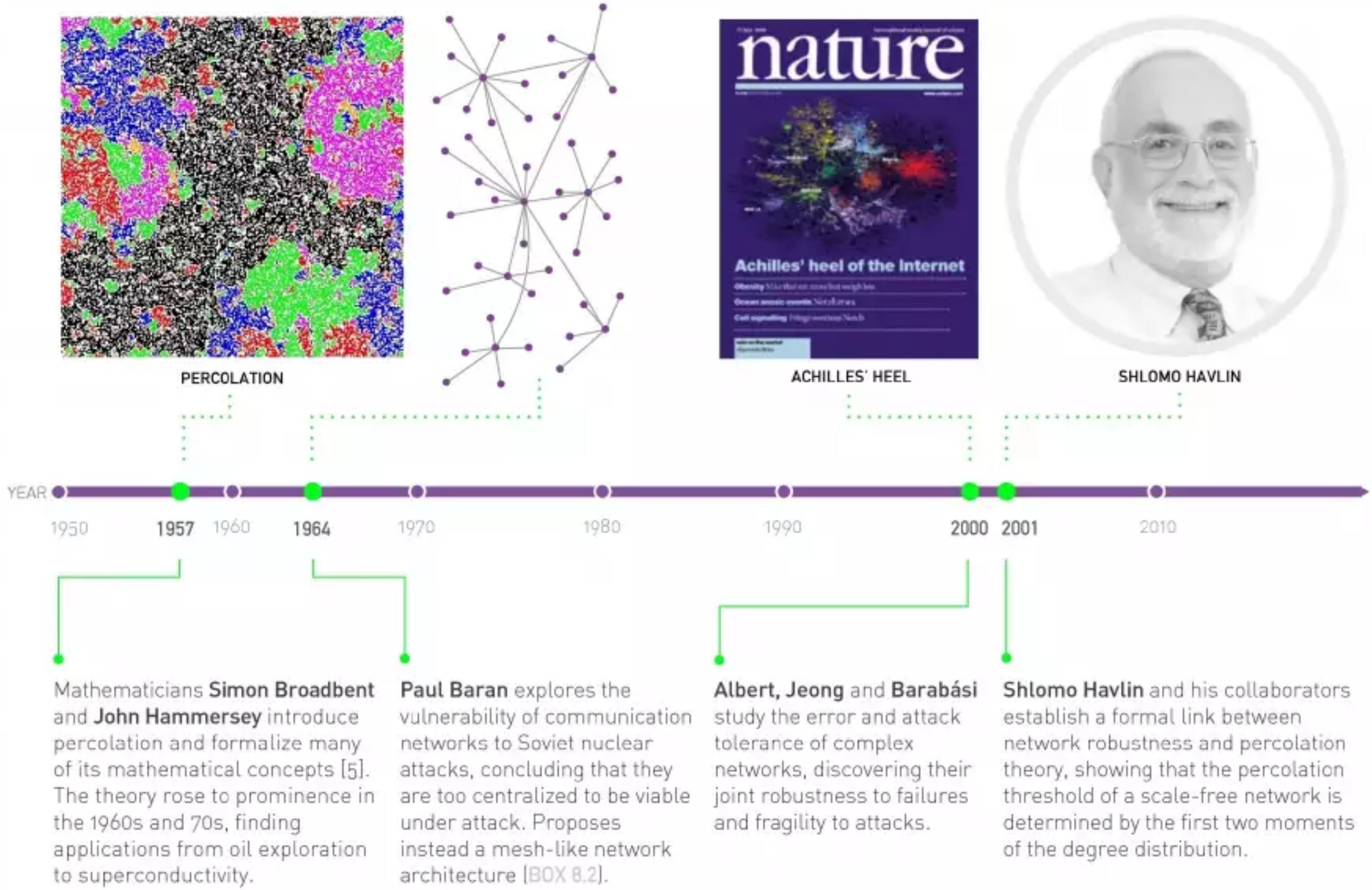


Redundancy


This office
will not tolerate
redundancy
in this office


I LOVE
REDUNDANCY

From Percolation to Robustness: A Brief History



Network Robustness

Malloy-Reed criteria:

A giant component exists if

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > 2$$

Random failures:

$$f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$$

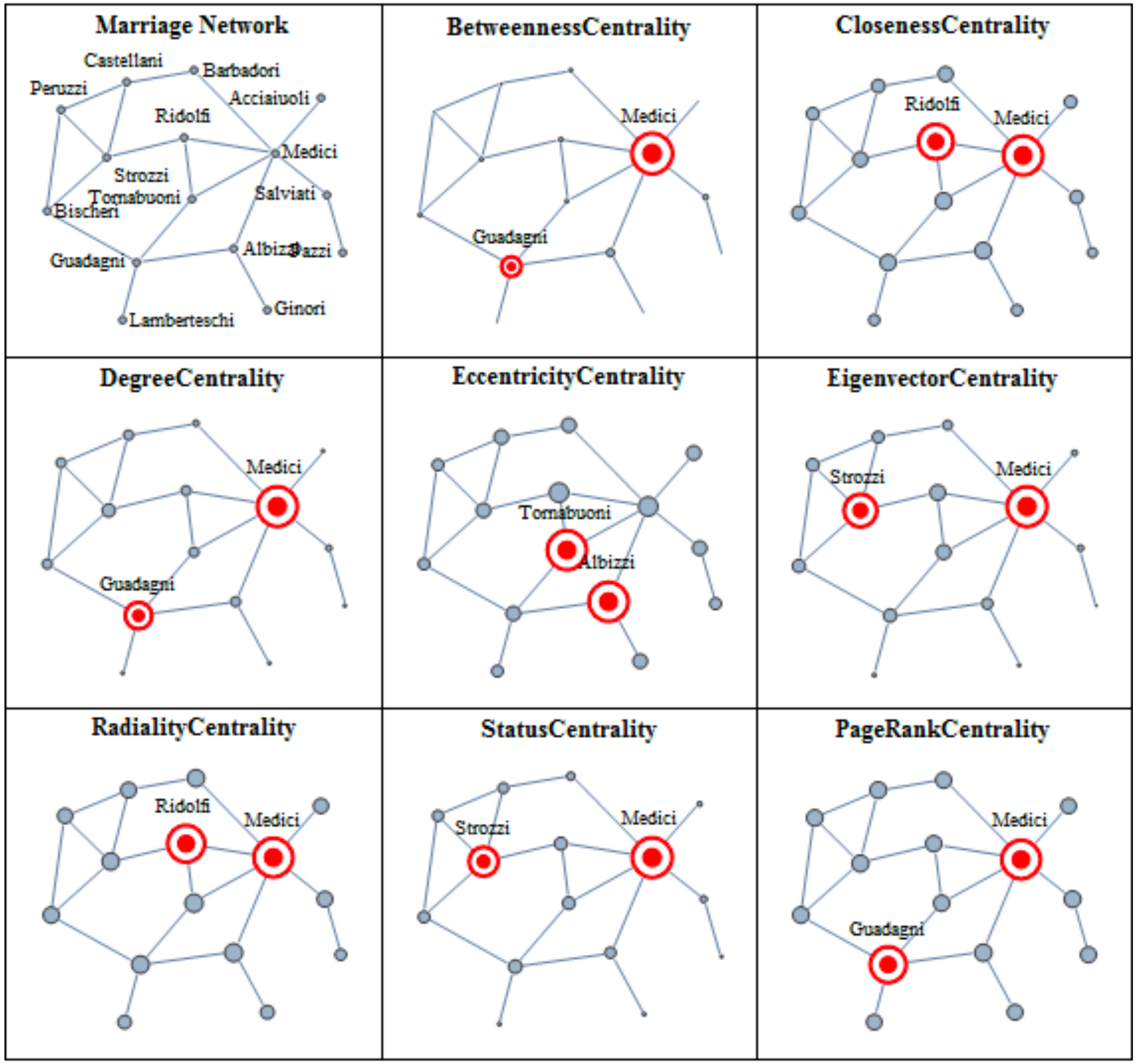
Cascading failures:

$$p(s) \approx s^{-\alpha}$$

$$\alpha = \begin{cases} \frac{3}{2} & \alpha > 3 \\ \frac{\gamma}{\gamma - 1} & 2 < \gamma < 3 \end{cases}$$

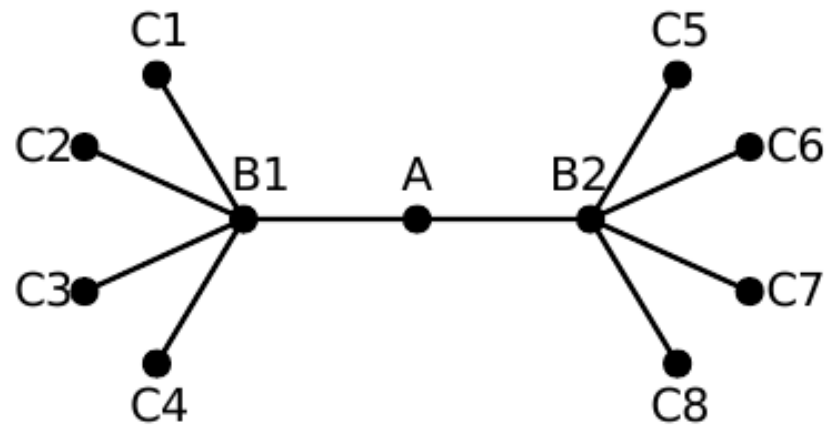
Exercises 4

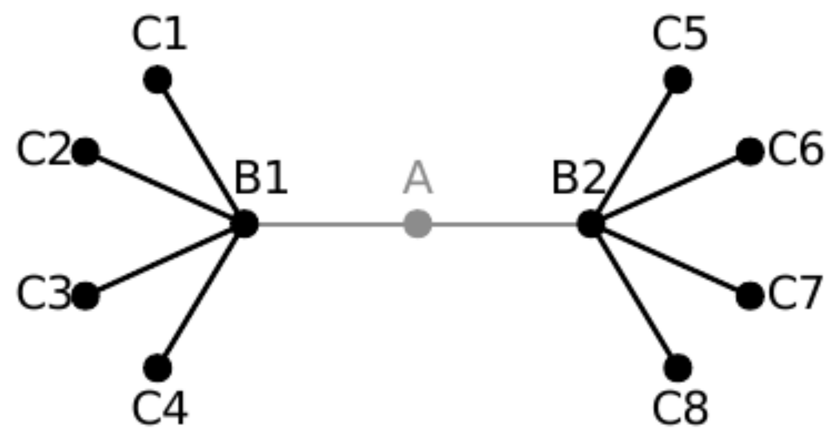
Network Robustness

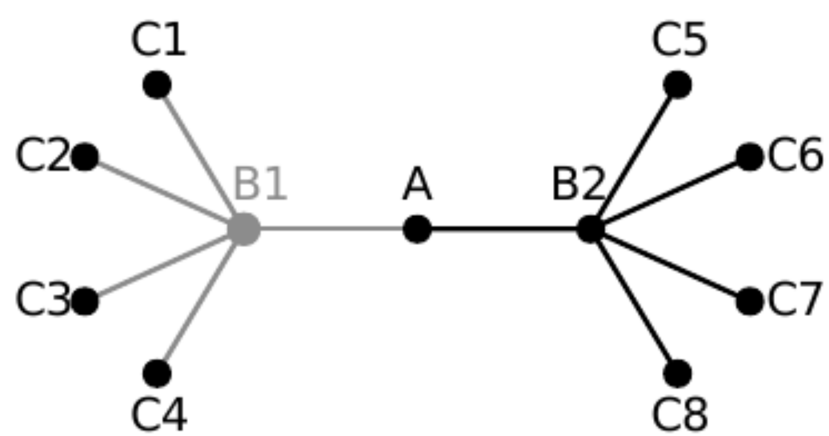


Centralities

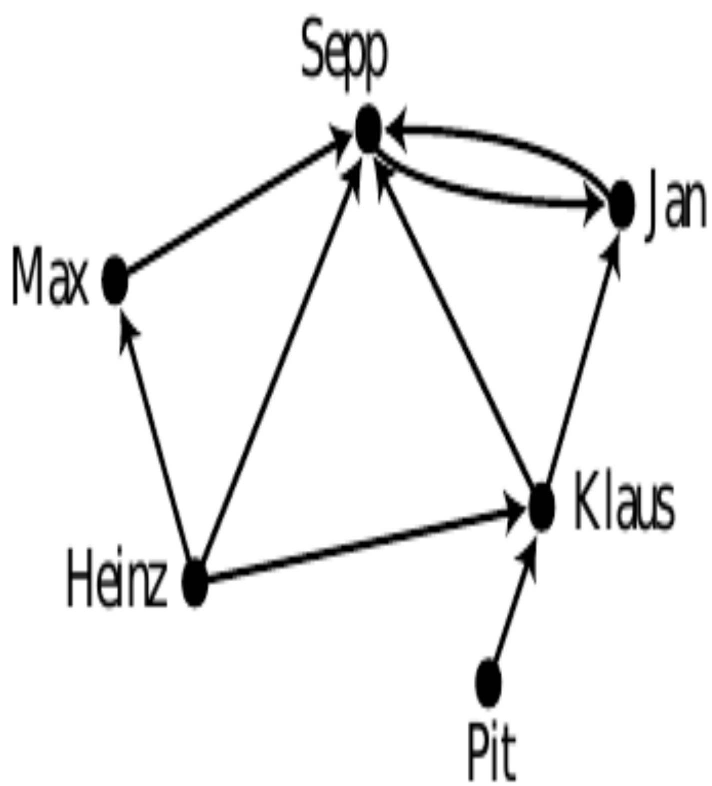
Central?


C1
C2
C3
C4
B1
A
B2
C5
C6
C7
C8
(a)


C1
C2
C3
C4
B1
A
B2
C5
C6
C7
C8
(b)


C1
C2
C3
C4
B1
A
B2
C5
C6
C7
C8
(c)

Degree centrality



Person	Number of votes received
Sepp	4
Jan	2
Klaus	2
Max	1
Heinz	0
Pit	0

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast Saccharomyces cerevisiae is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The S. cerevisiae protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data[3,4], obtained mostly by systematic two-hybrid analyses[3]. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little

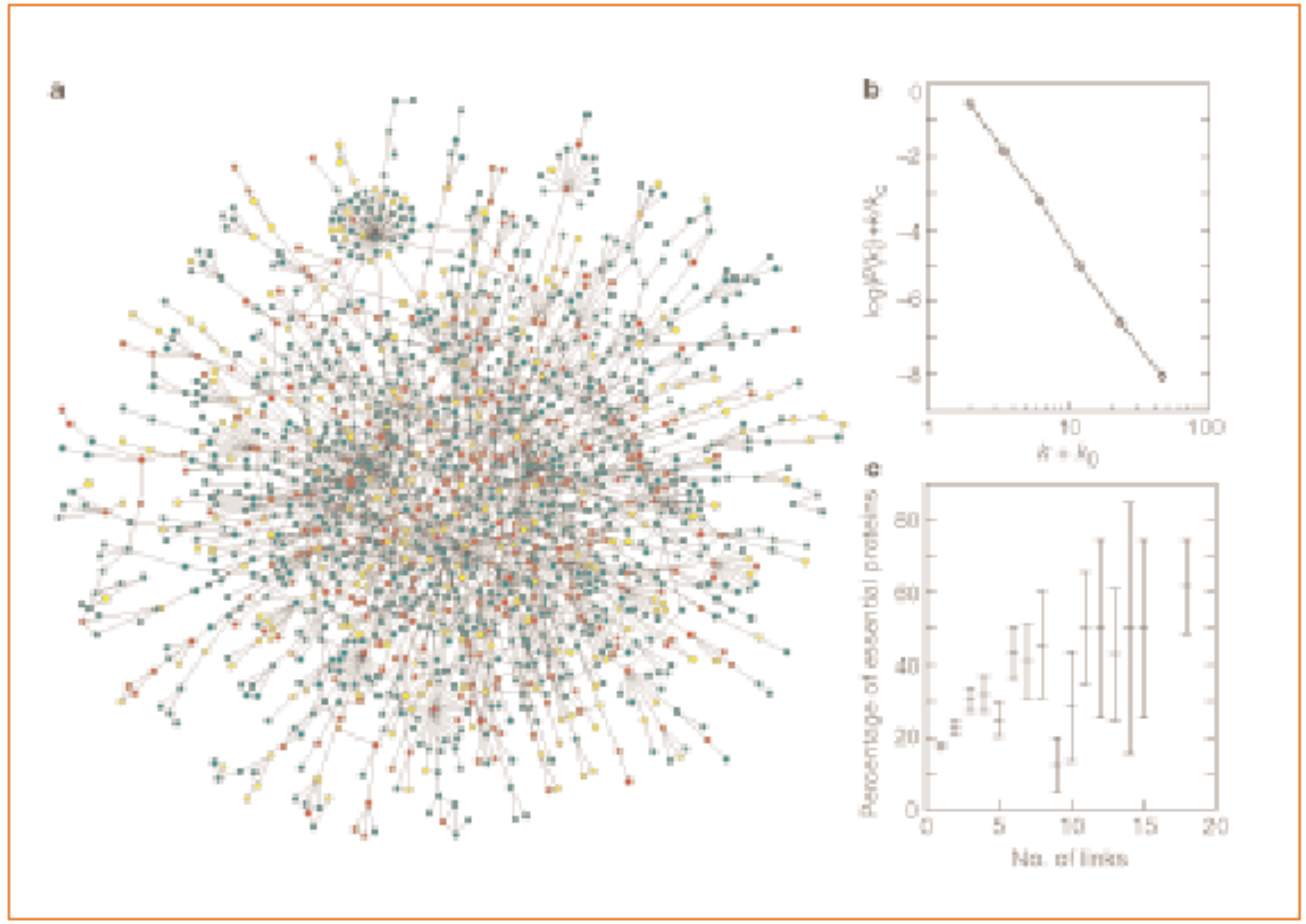
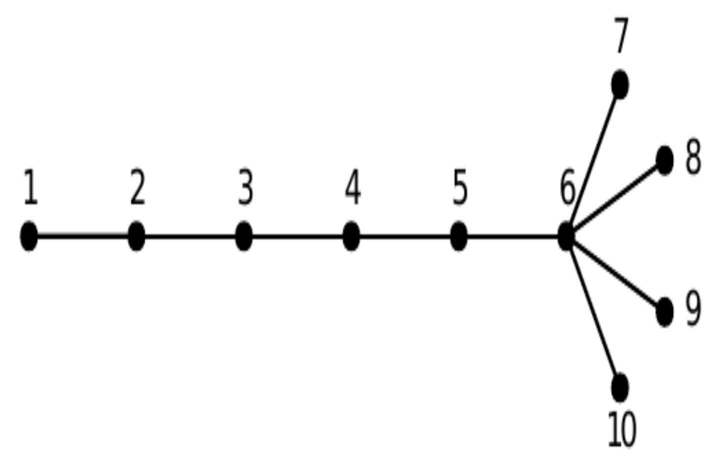


Figure 1 Characteristics of the yeast proteome. **a,** Map of protein–protein interactions. The largest cluster, which contains ~78% of all

Eccentricity

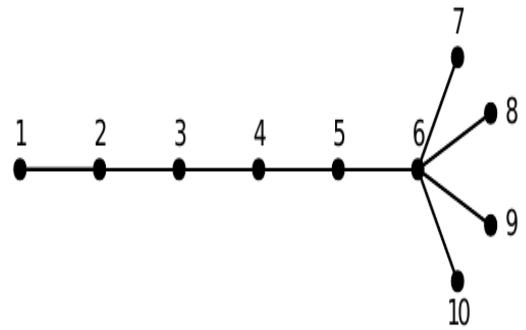
$$C_{exc}(s) = \max\{dist(s, t) : t \in V\}$$



$e(1) = 6$, $e(2) = 5 \ldots$

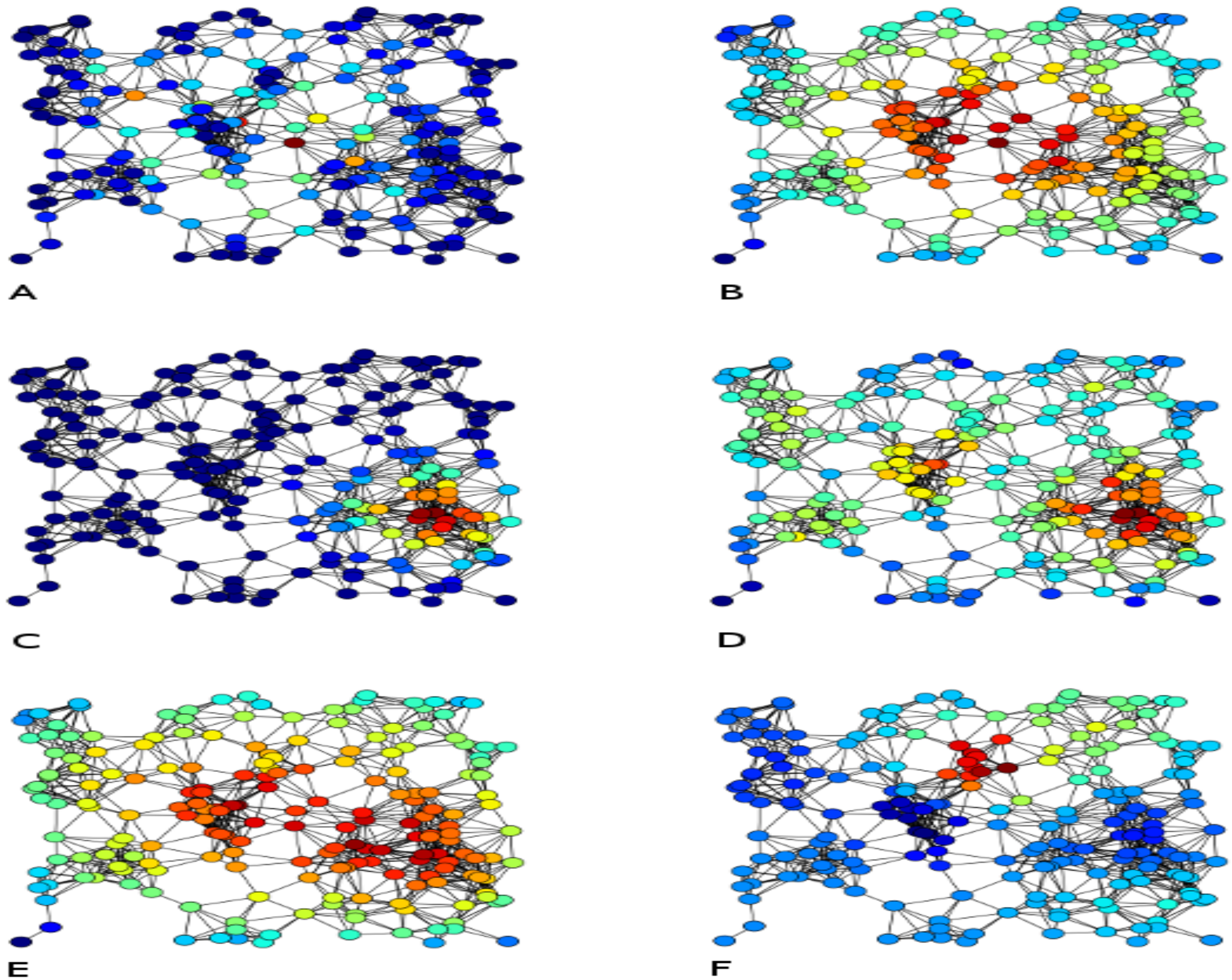
Closeness Centrality

$$C_{close}(s) = \frac{1}{\sum_t dist(s, t)}$$



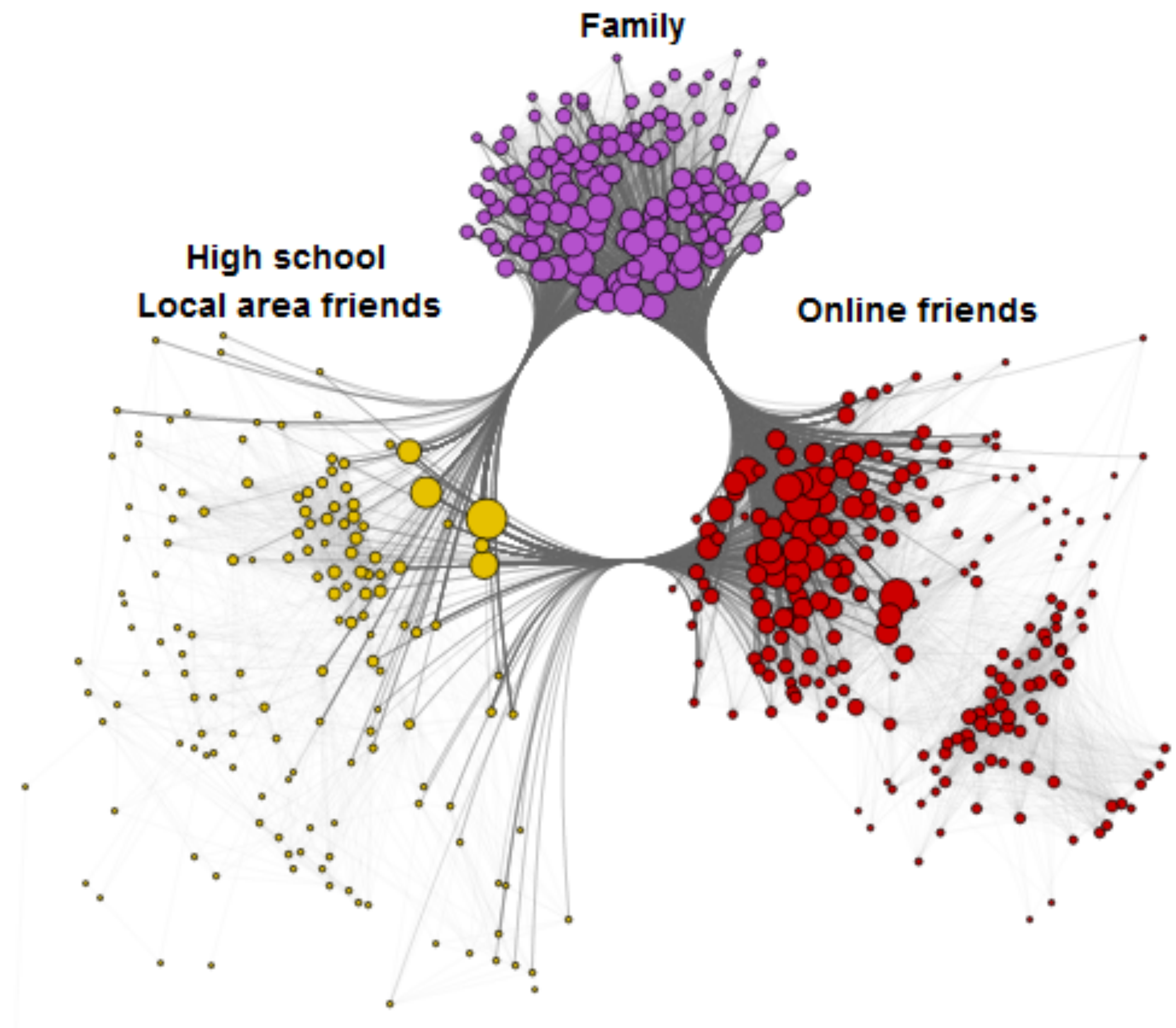
$d(1,2) = 1$, $d(1,3) = 2$, $d(1,4) = 3$, $d(1,5) = 4$, $d(1,6) = 5$, $d(1,7) = d(1,8) = d(1,9) = d(1,10) = 6$

$1 + 2 + 3 + 4 + 5 + 6 * 4 = 39$, $\frac{1}{39} = 0.025$


A
B
C
D
E
F

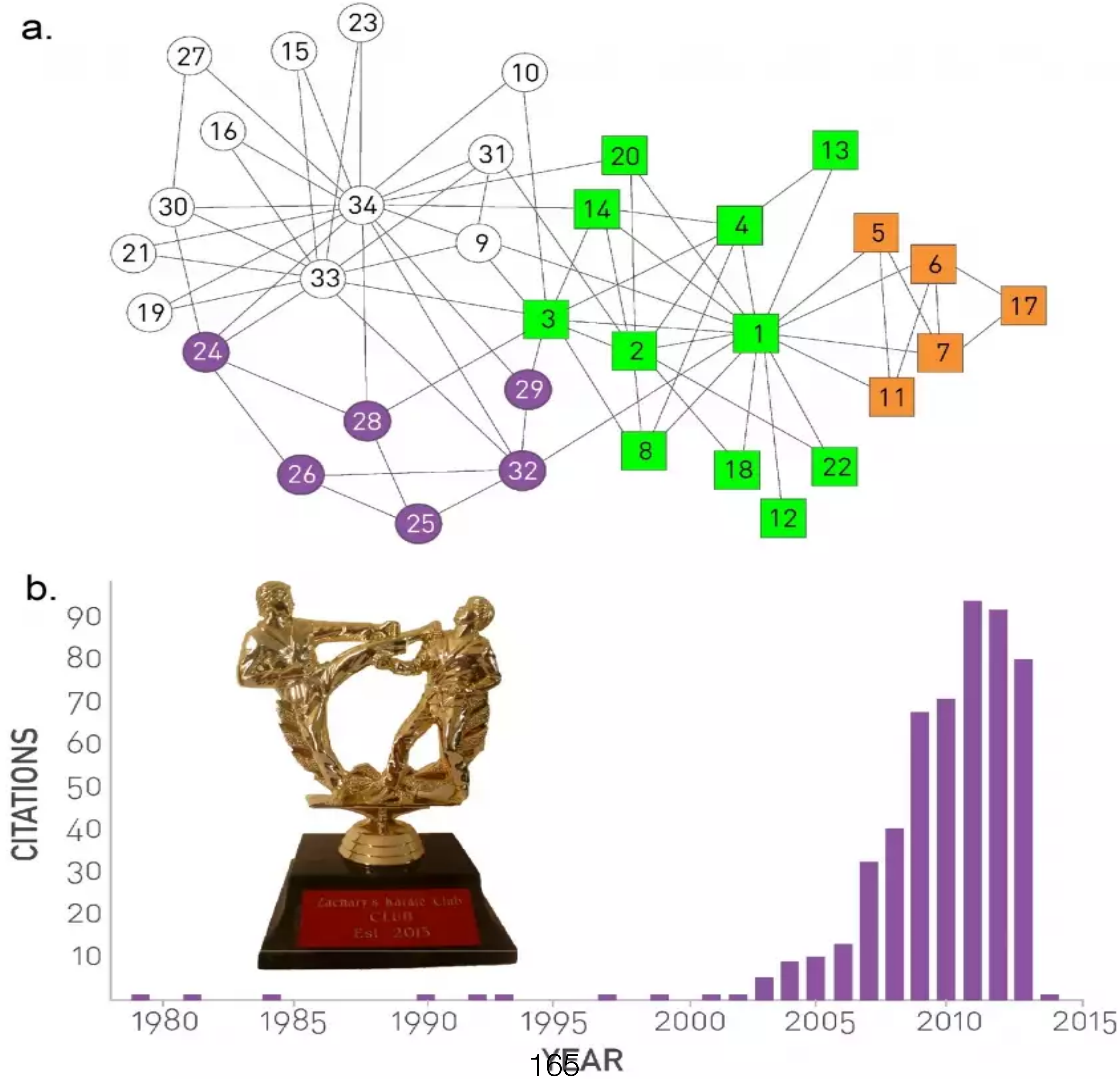
Exercises 5

Centralities

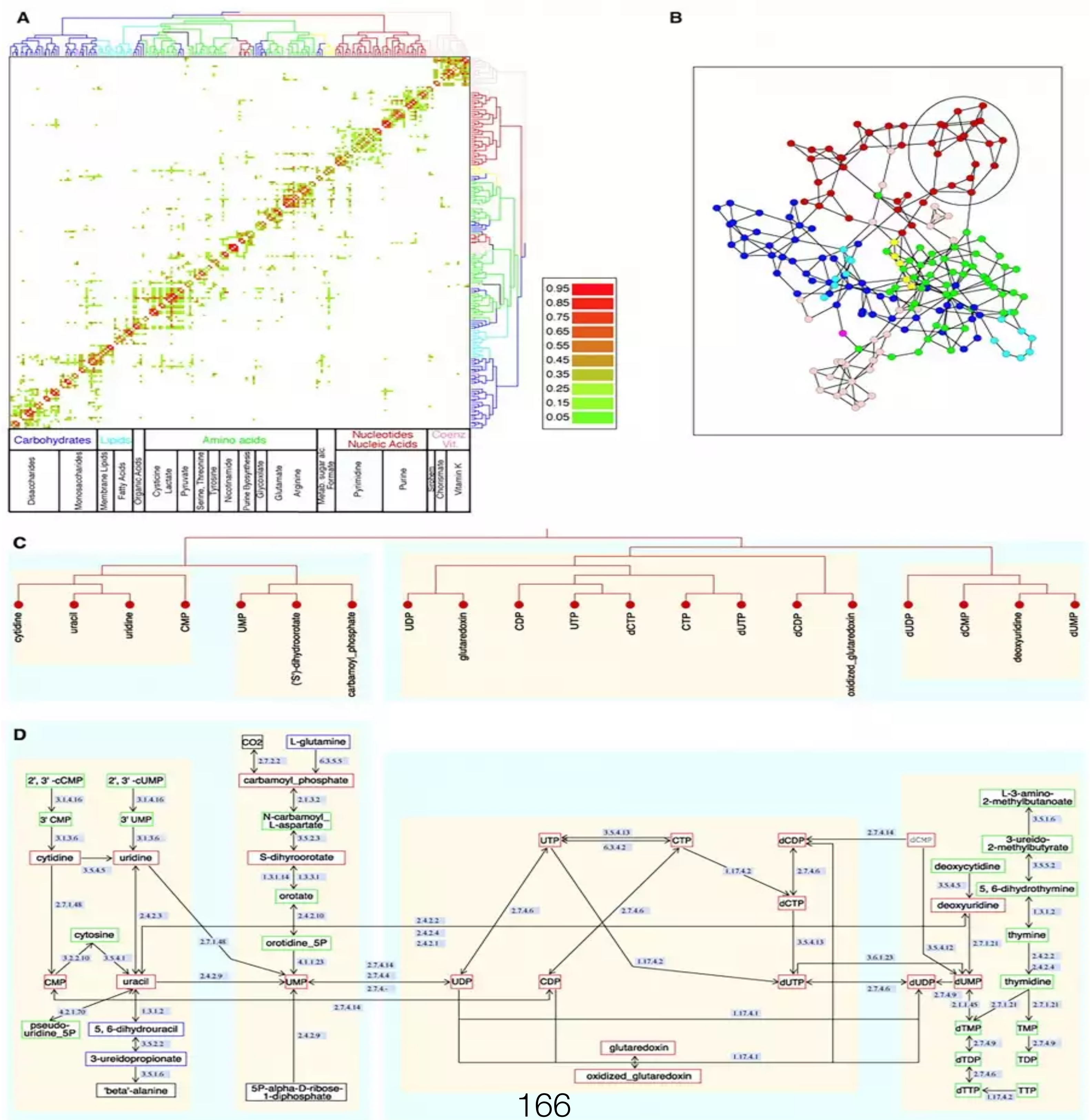


5. Communities

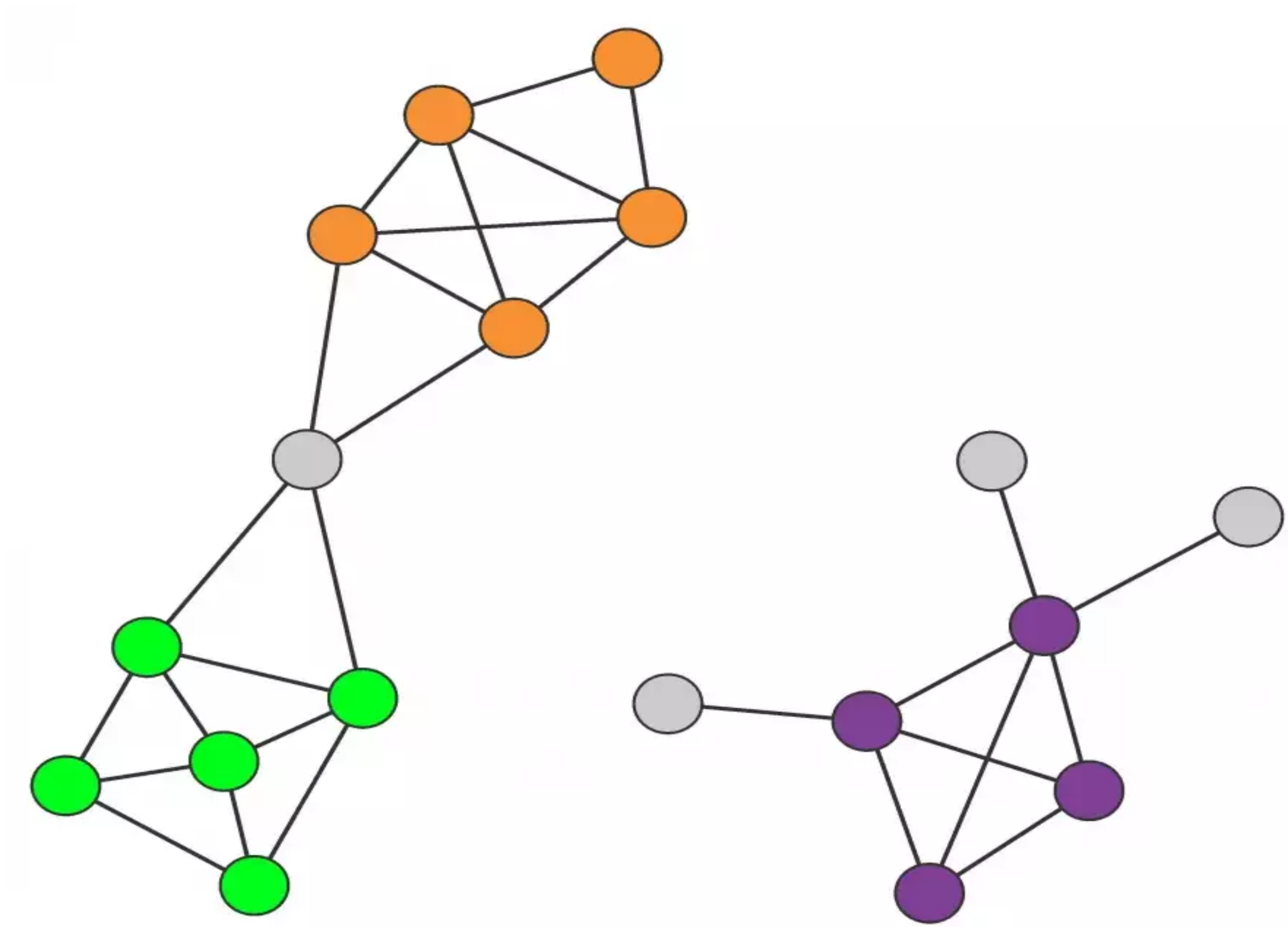
Zachary’s Karate Club



Communities in Metabolic Networks



Connectedness and Density Hypothesis



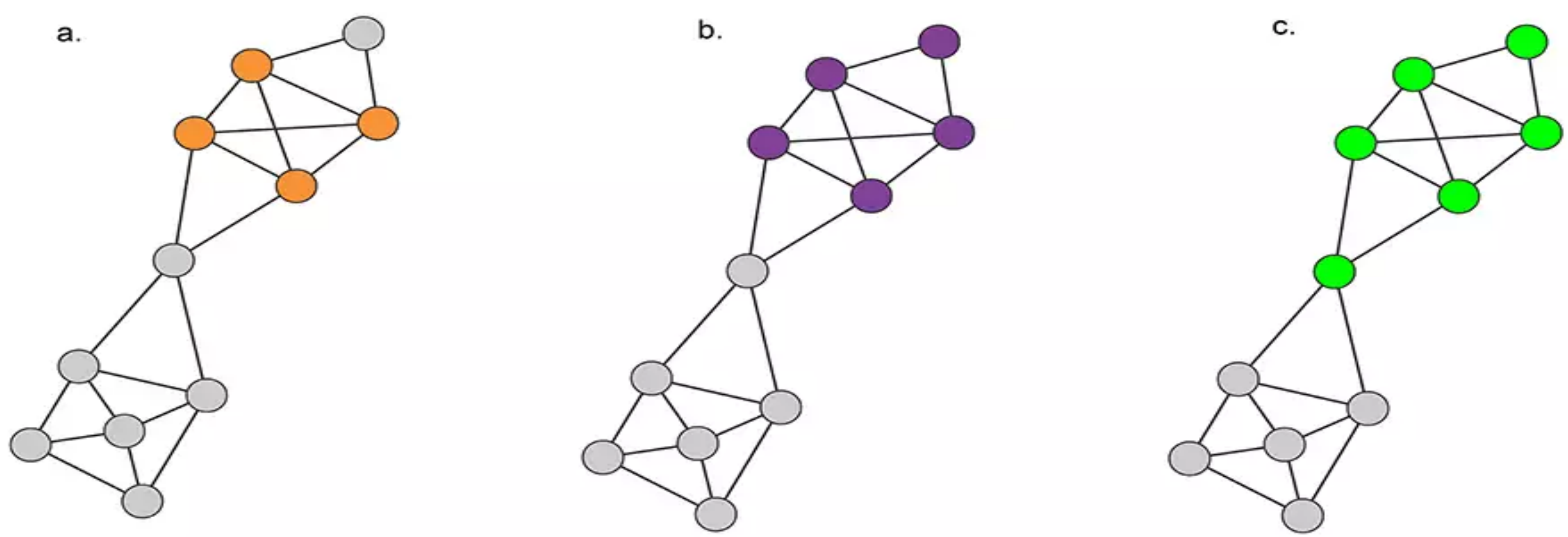
Communities are locally dense connected subgraphs in a network

Defining Communities

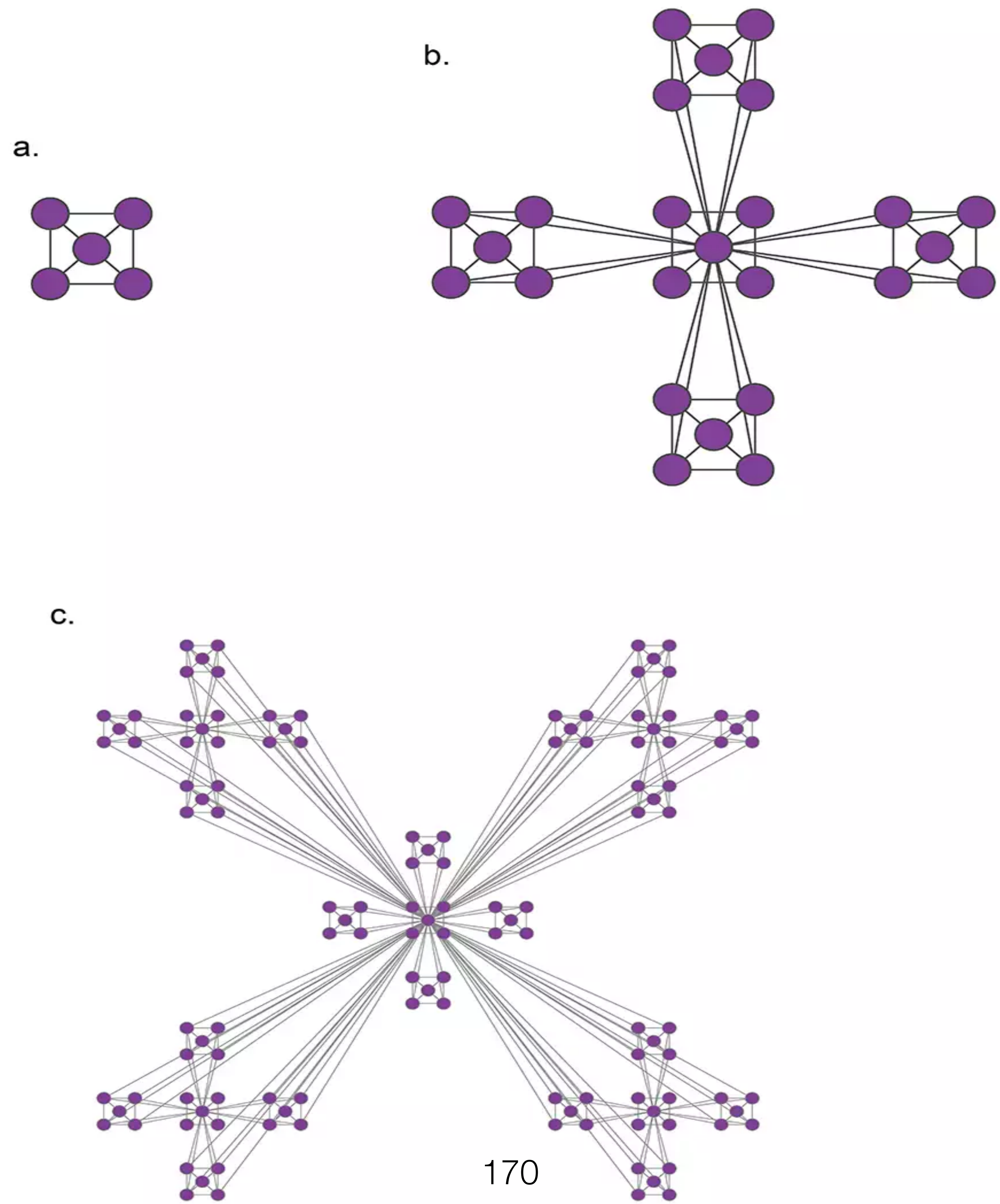
1. **Cliques** A *clique* corresponds to a complete subgraph.
2. **Strong Communities** A *strong community* is a connected subgraph whose nodes have more links to other nodes in the same community that to nodes that belong to other communities.
3. **Weak Communities** A weak community is a subgraph whose nodes' total internal degree exceeds their total external degree.

Defining Communities

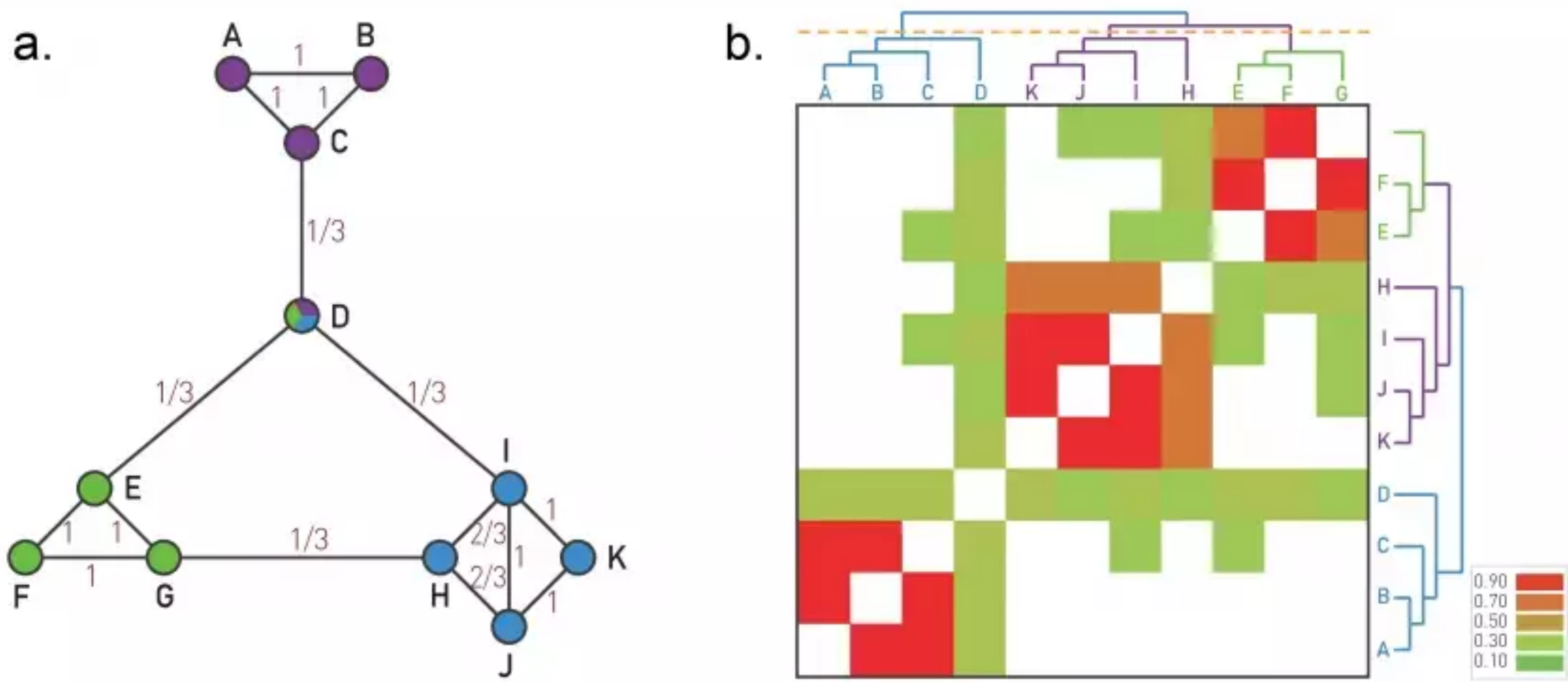
1. **Cliques** A *clique* corresponds to a complete subgraph.
2. **Strong Communities** A *strong community* is a connected subgraph whose nodes have more links to other nodes in the same community that to nodes that belong to other communities.
3. **Weak Communities** A weak community is a subgraph whose nodes' total internal degree exceeds their total external degree.



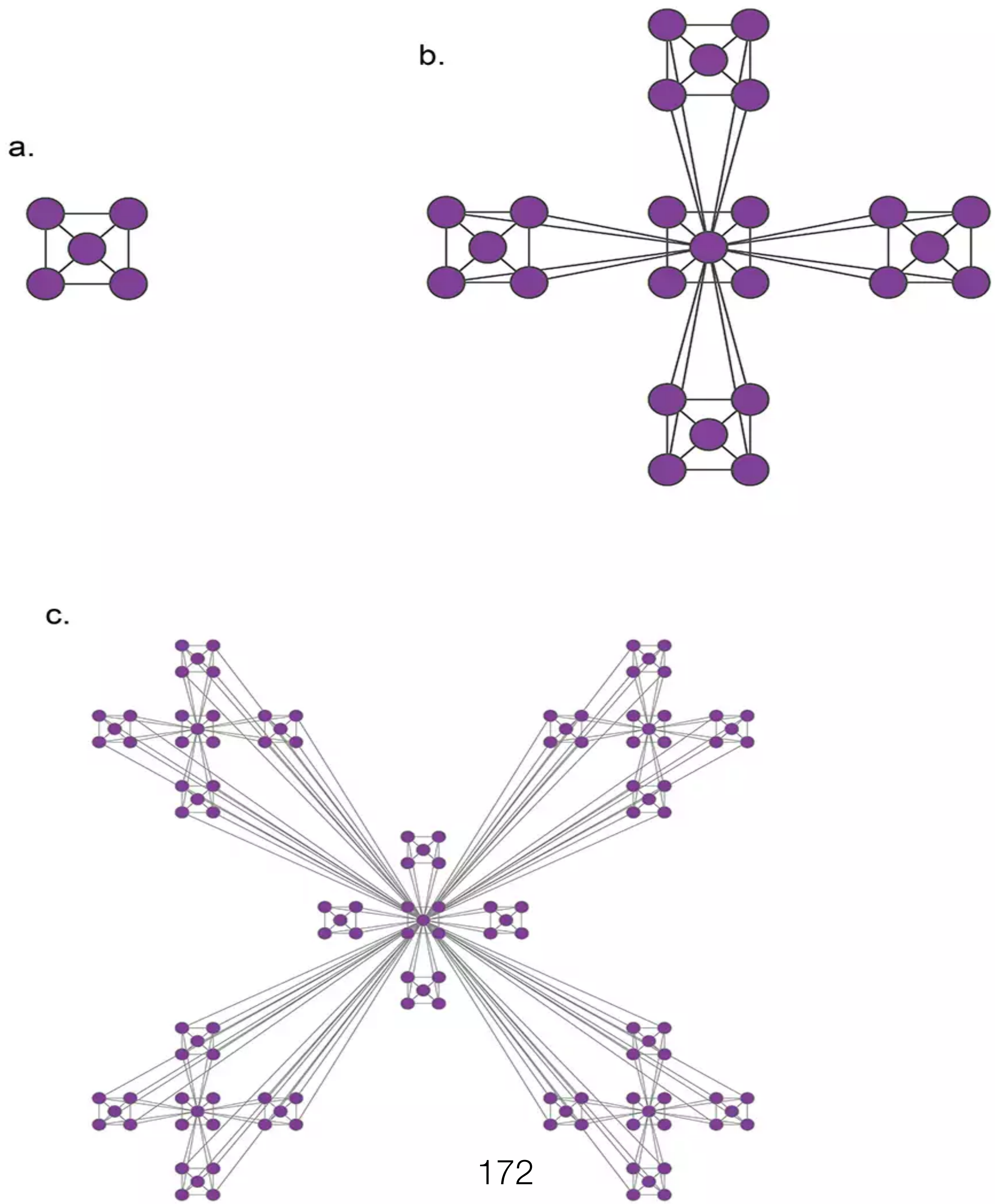
Hierarchy in Real Networks



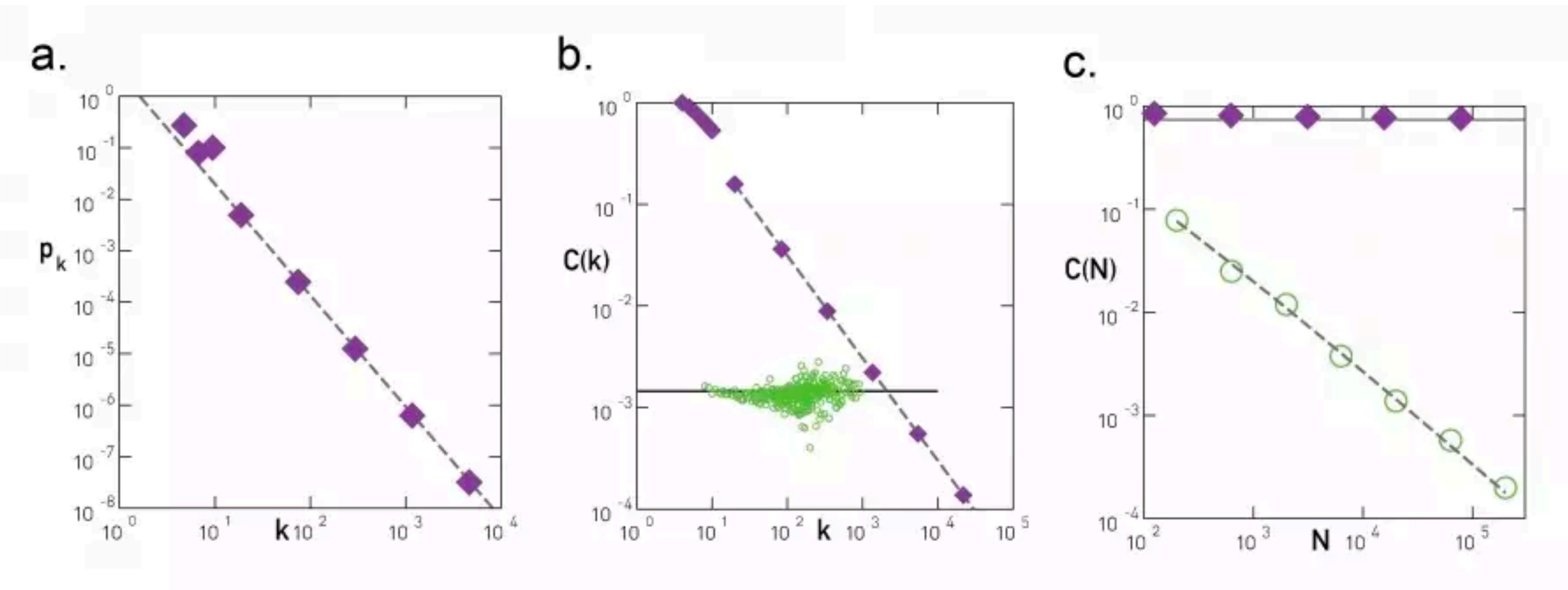
Hierarchical Clustering



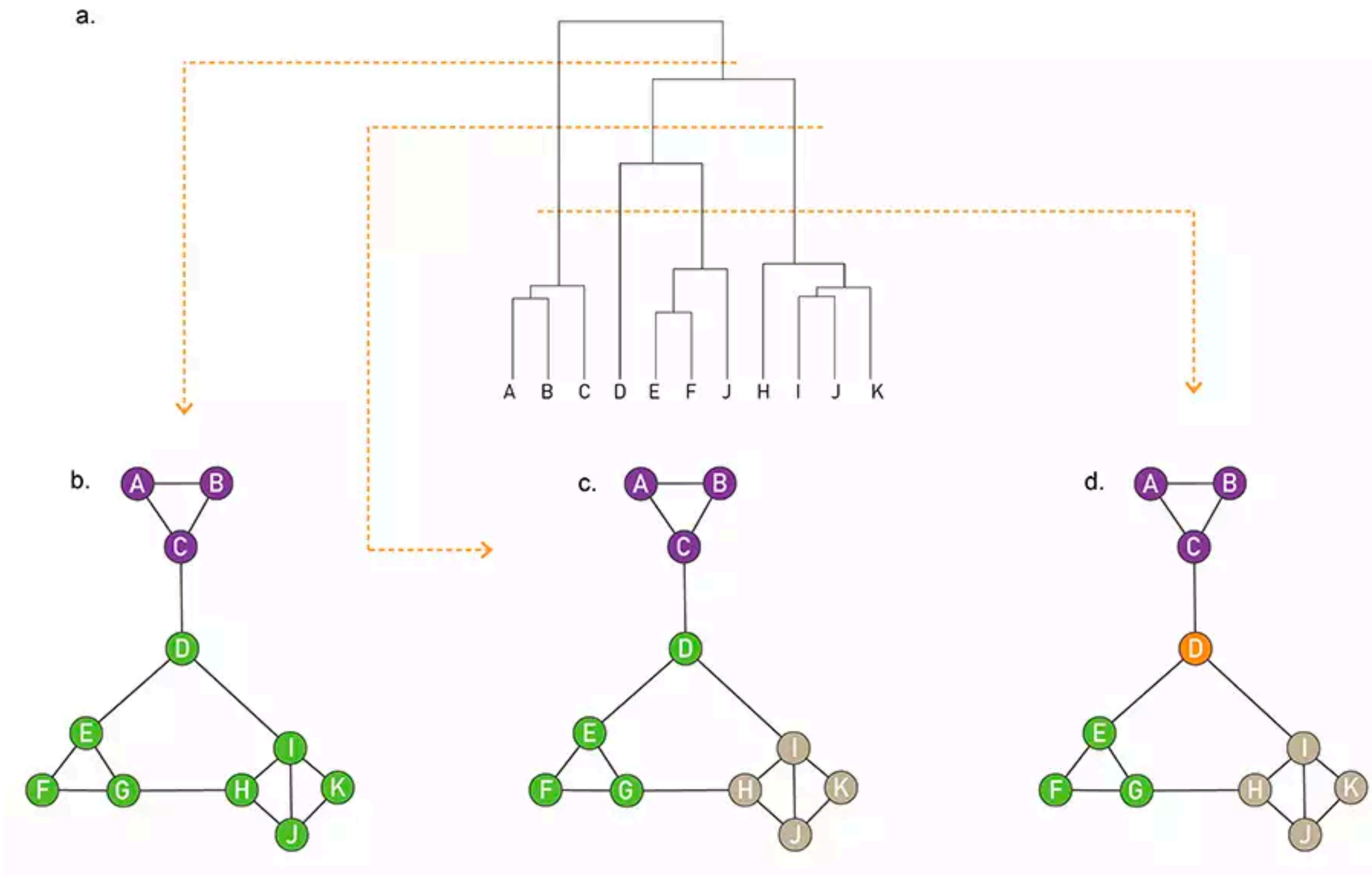
Hierarchy in Real Networks



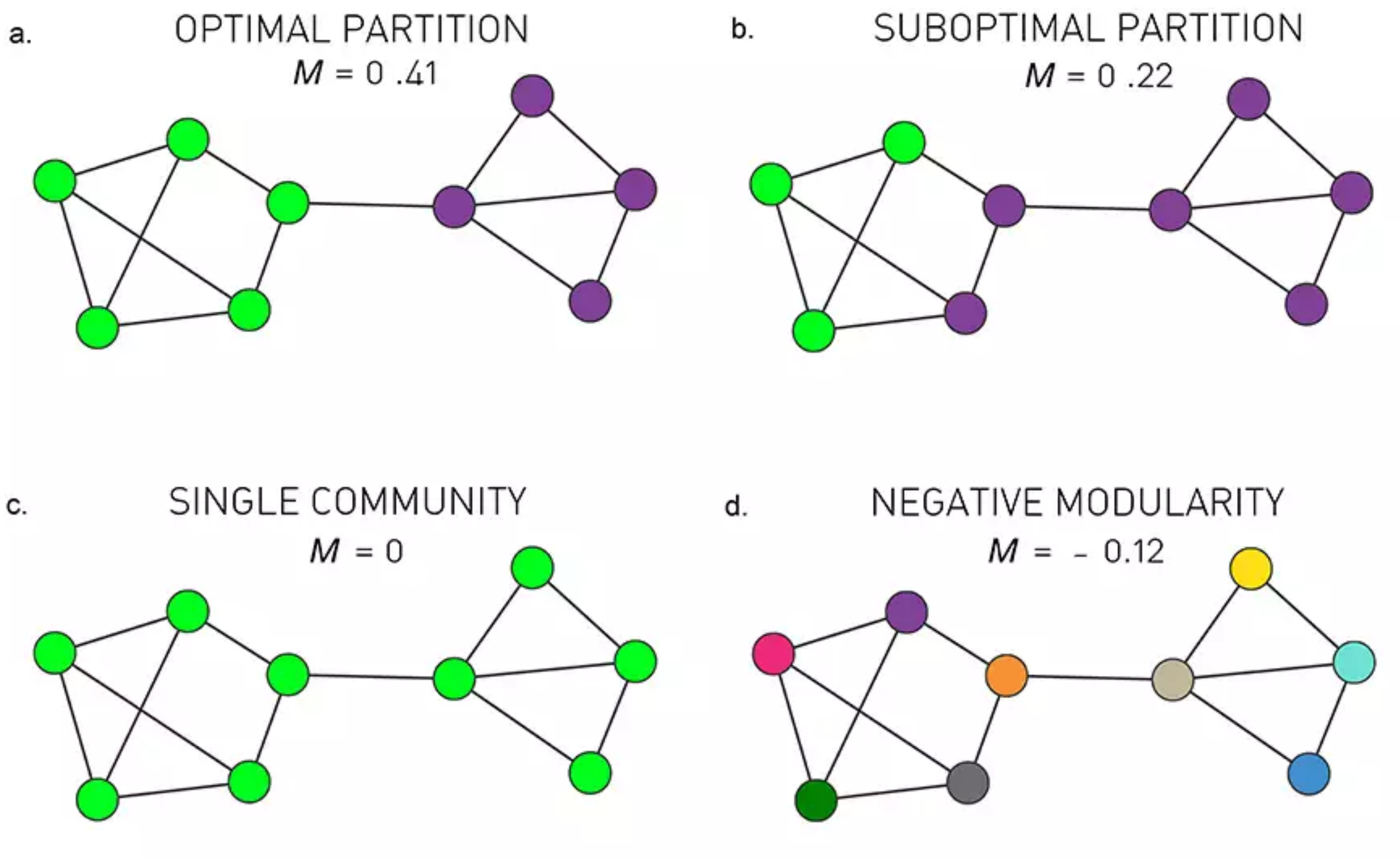
Scaling in Hierachical Networks



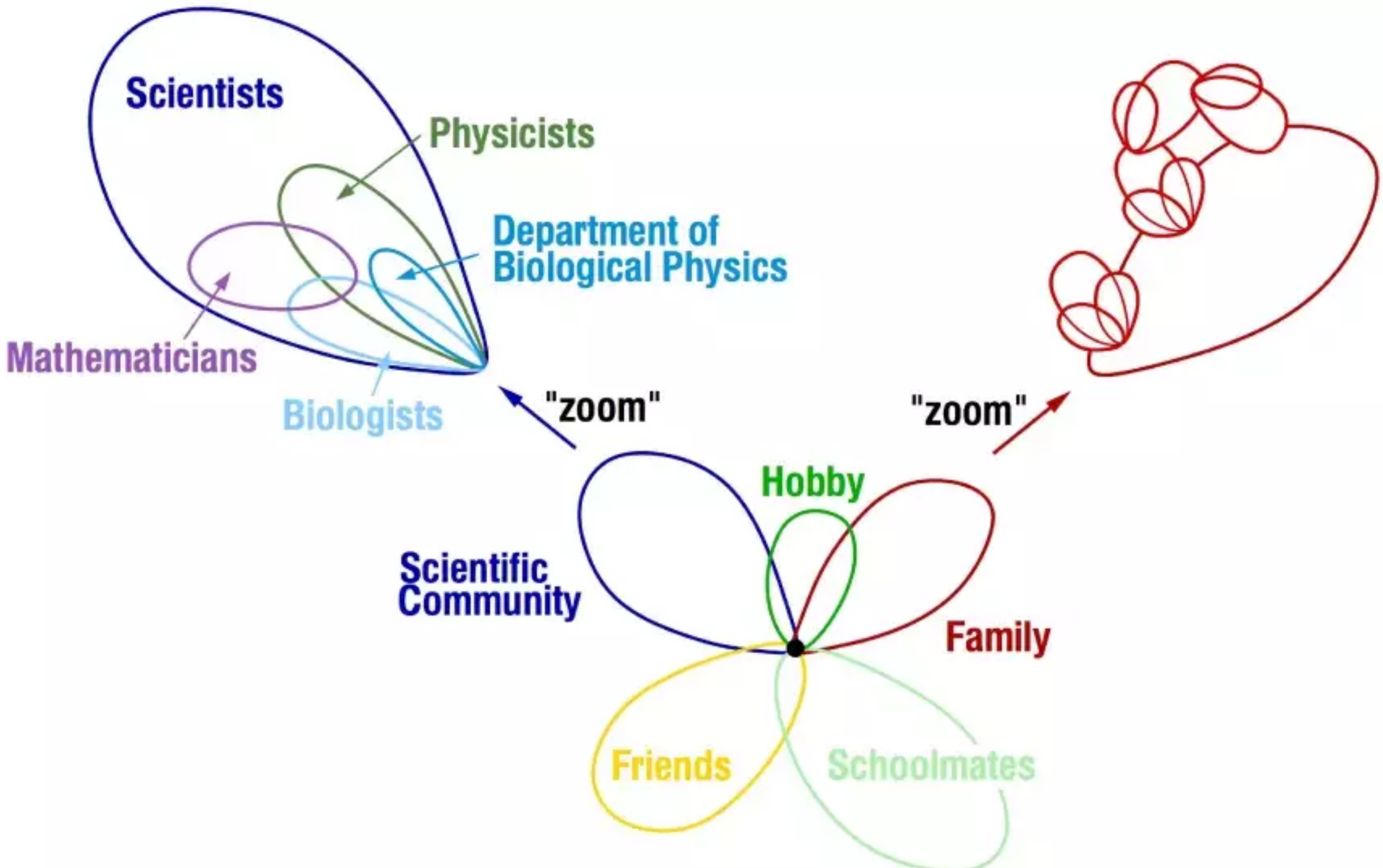
Ambiguity in Hierarchical Clustering



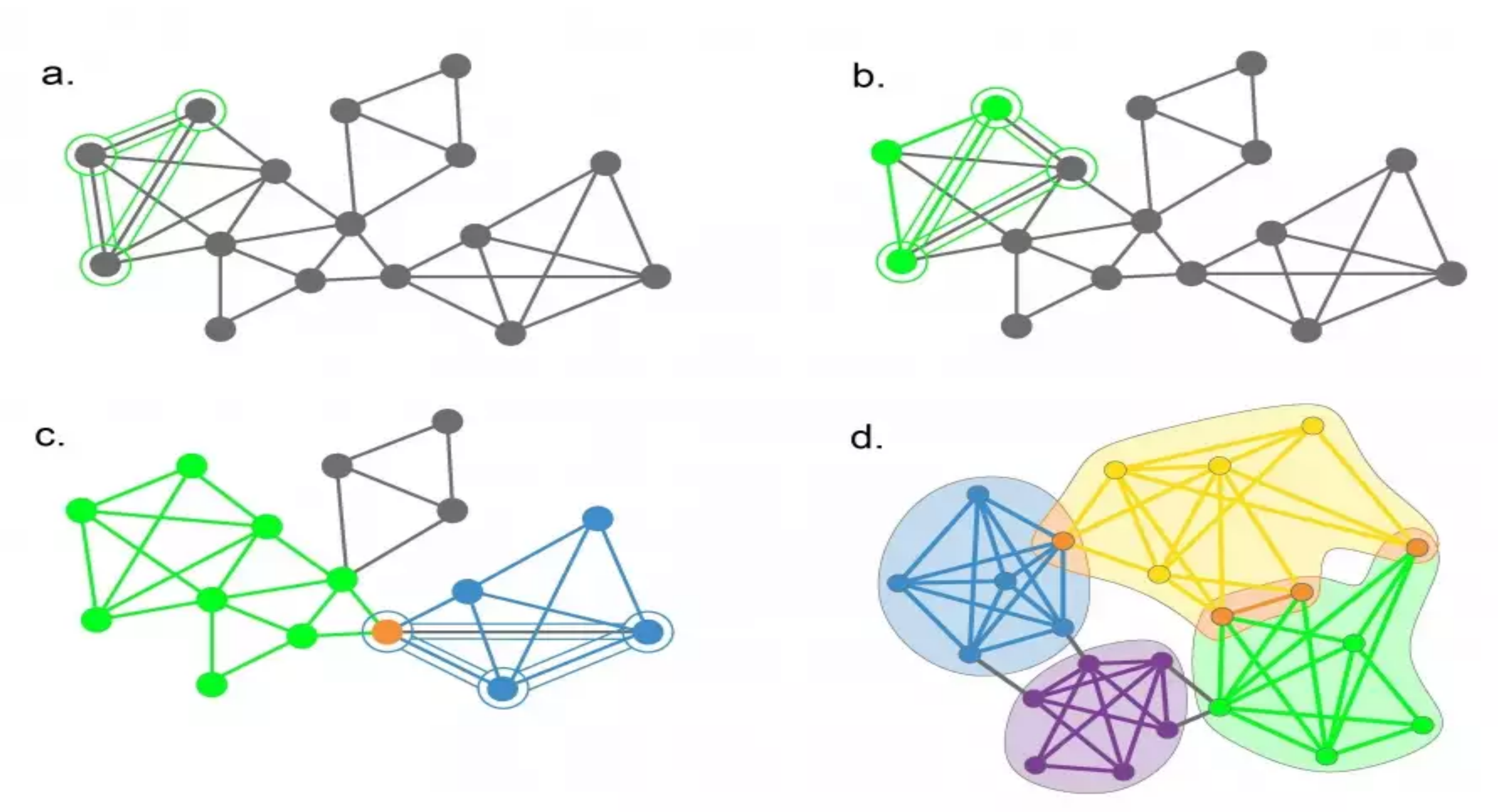
Modularity


a.
OPTIMAL PARTITION
M = 0 .41
b.
SUBOPTIMAL PARTITION
M = 0 .22
c.
SINGLE COMMUNITY
M = 0
d.
NEGATIVE MODULARITY
M = – 0.12

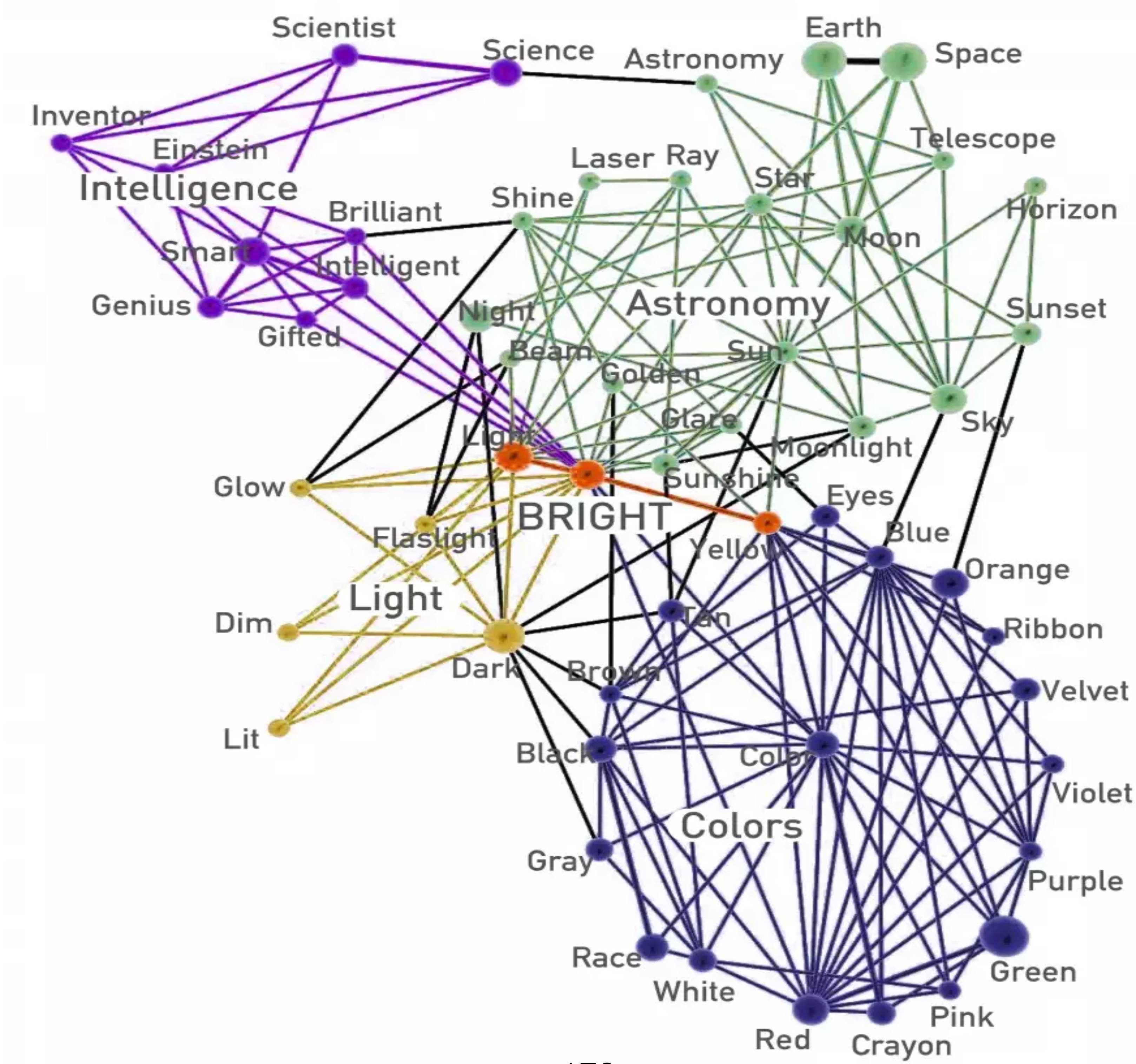
Overlapping Communities



Clique Percolation



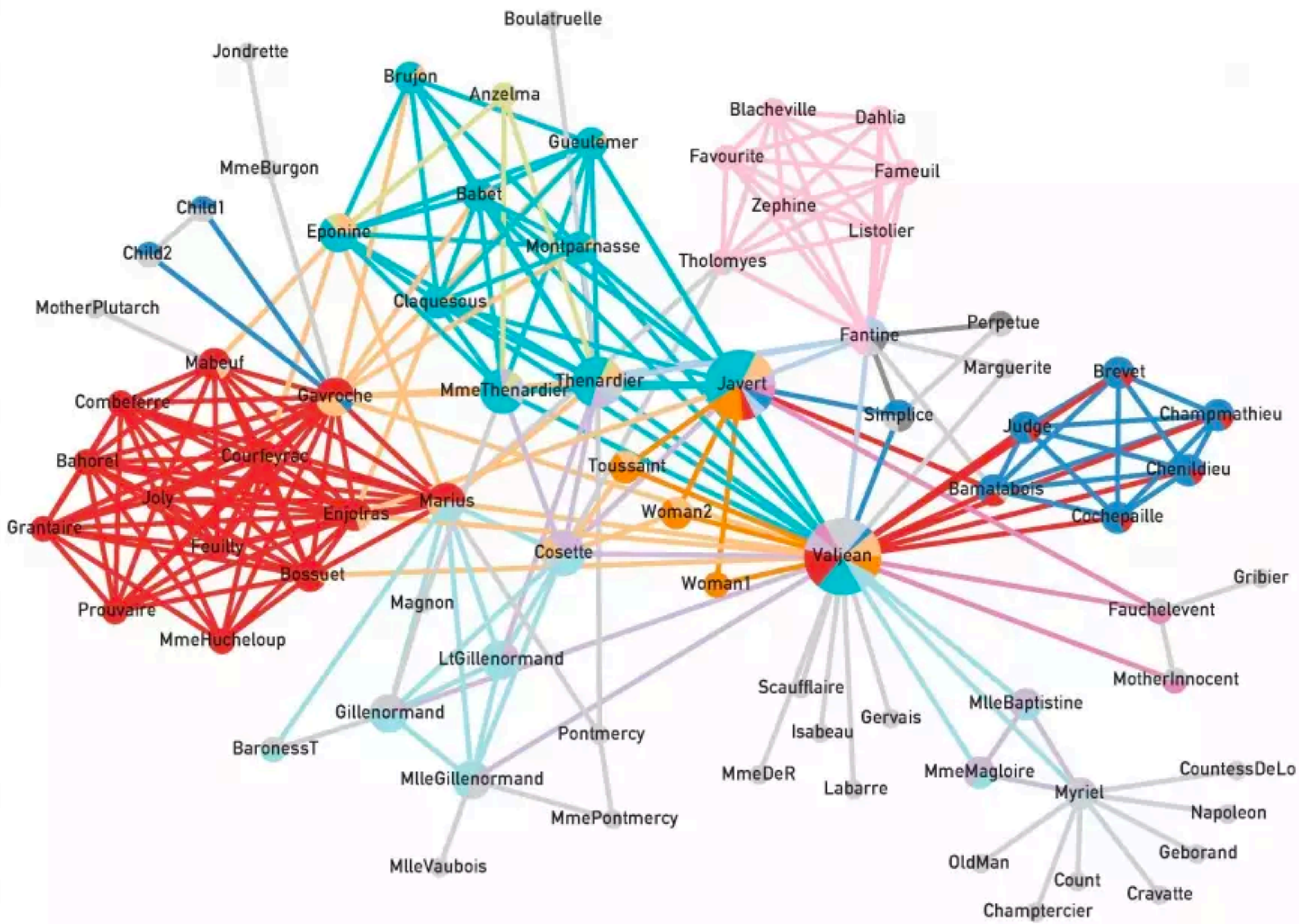
Overlapping Communities



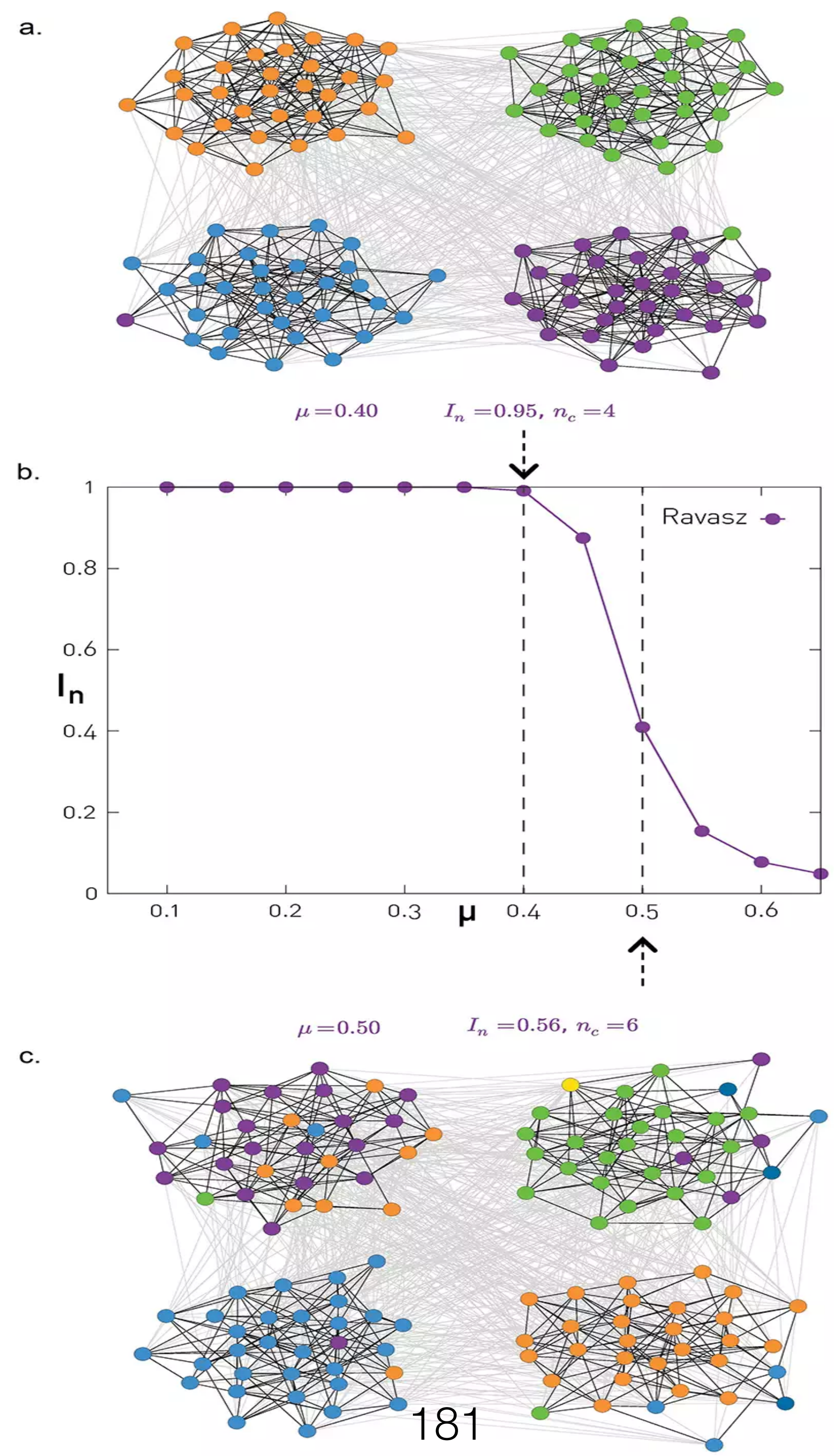
http://www.mapequation.org/apps/NetworkNavigator.html

Infomap

http://barabasi.com/networksciencebook/images/ch-09/video-9-3.webm


Boulatruelle
Jondrette
Brujon
Anzelma
Blacheville
Dahlia
Gueulemer
Favourite
Fameuil
MmeBurgon
Babet
Zephine
Child1
Eponine
Listolier
Montparnasse
Tholomyes
Child2
Claquesous
MotherPlutarch
Perpetue
Fantine
Mabeuf
Marguerite
Thenardier
Brevet
Javert
Combeferre
Gavroche
MmeThenardier
Simplice
Champmathieu
Judge
Courfeyrac
Bahorel
Toussaint
Chenildieu
Joly
Marius
Bamatabois
Enjolras
Woman2
Cochepaille
Grantaire
Feuilly
Cosette
Valjean
Bossuet
Gribier
Woman1
Magnon
Fauchelevent
Prouvaire
MmeHucheloup
LtGillenormand
MotherInnocent
Scaufflaire
Gervais
MlleBaptistine
Gillenormand
Isabeau
Pontmercy
BaronessT
MmeDeR
Labarre
MmeMagloire
CountessDeLo
MlleGillenormand
Myriel
Napoleon
MmePontmercy
Geborand
OldMan
MlleVaubois
Count
Cravatte
Champtercier

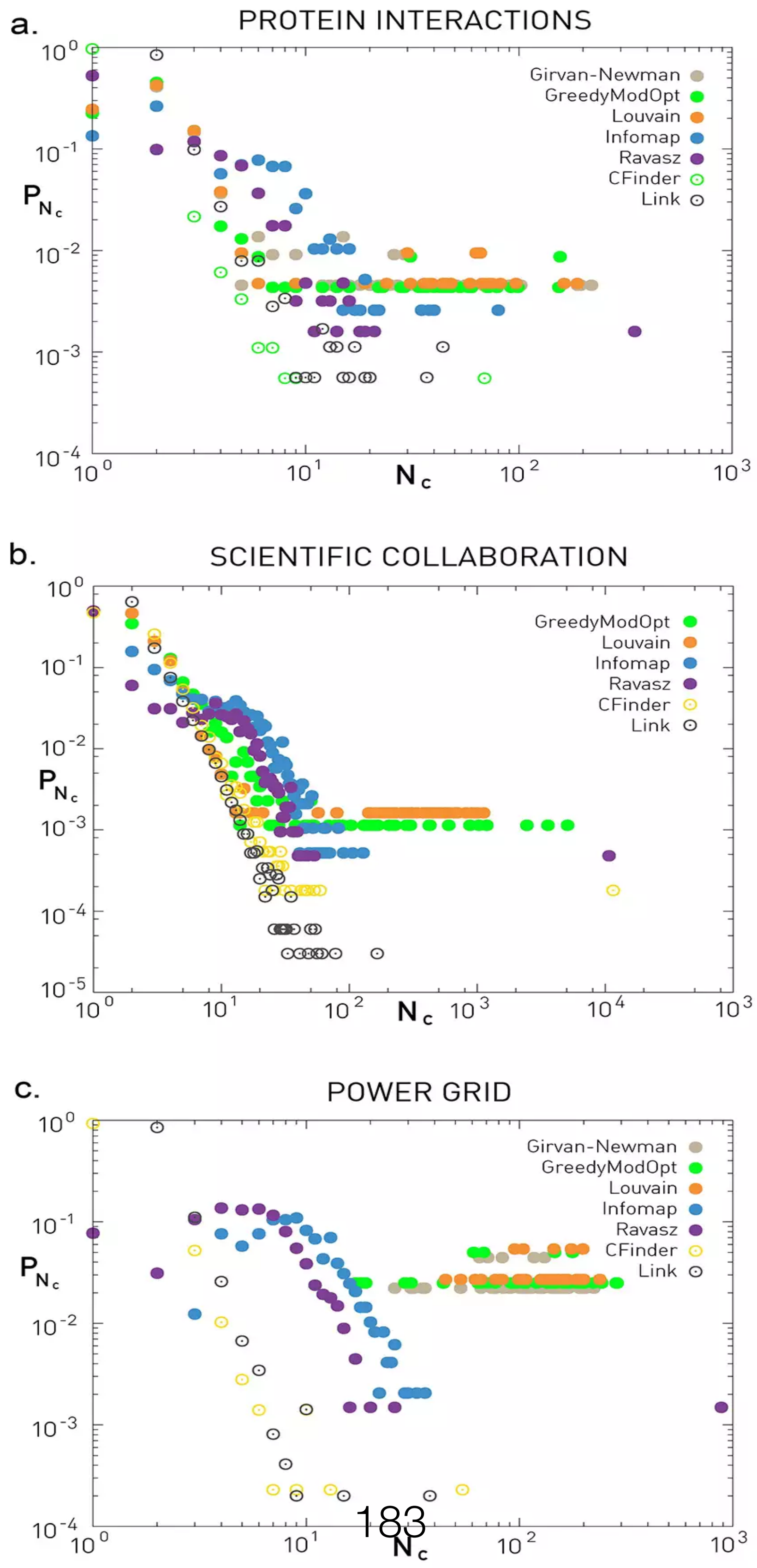
Testing Accuracy with the NG Benchmark



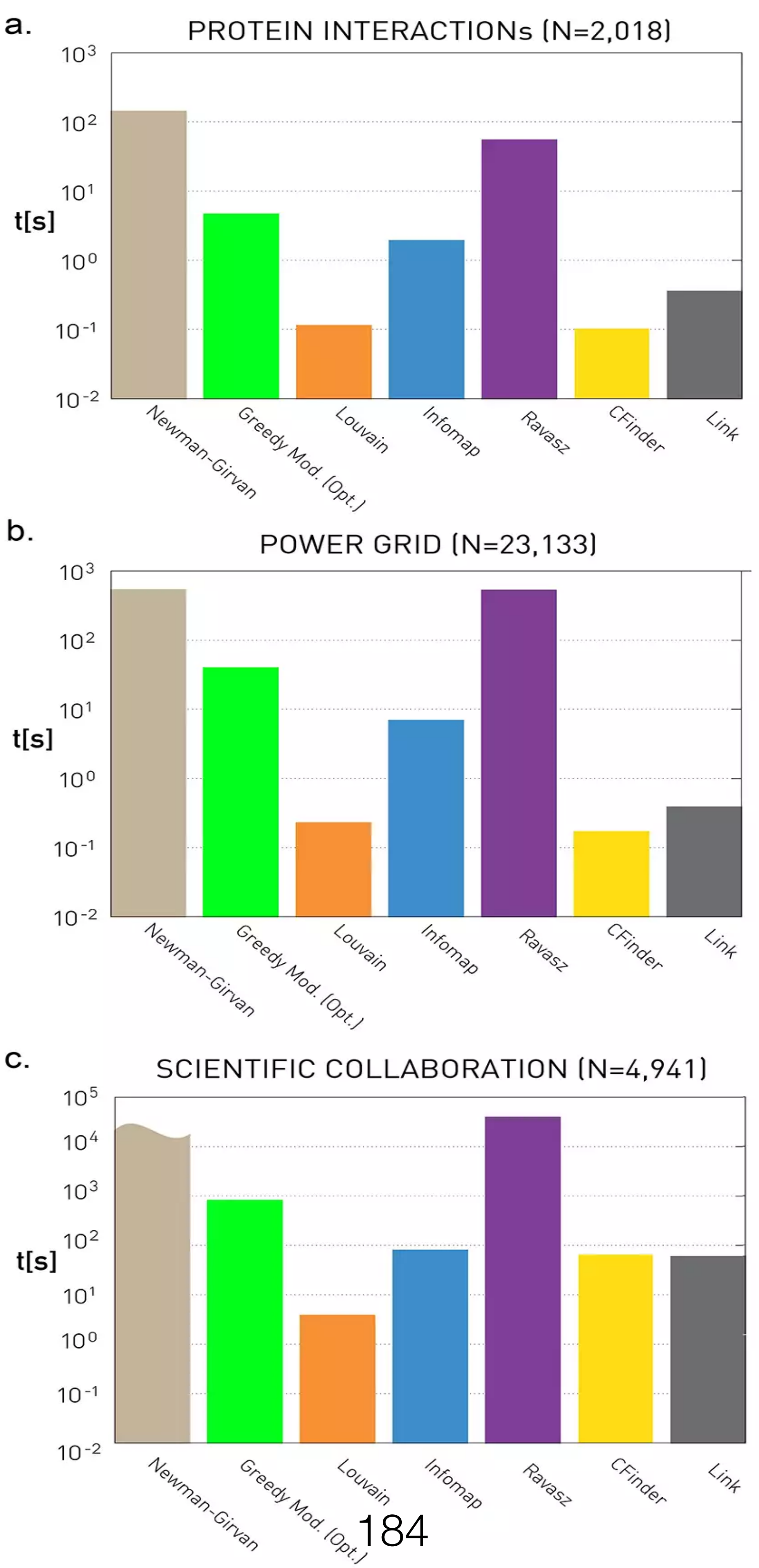
Algorithmic Complexity

Name	Nature	Comp.	REF
Ravasz	Hierarchical Agglomerative	$O(N^2)$	[11]
Girvan-Newman	Hierarchical Divisive	$O(N^2)$	[9]
Greedy Modularity	Modularity Optimization	$O(N^2)$	[33]
Greedy Modularity (Optimized)	Modularity Optimization	$O(N\log^2 N)$	[35]
Louvain	Modularity Optimization	$O(L)$	[2]
Infomap	Flow Optimization	$O(N\log N)$	[44]
Clique Percolation (CFinder)	Overlapping Communities	$Exp(N)$	[48]
Link Clustering	Hierarchical Agglomerative; Overlapping Communities	$O(N^2)$	[51]

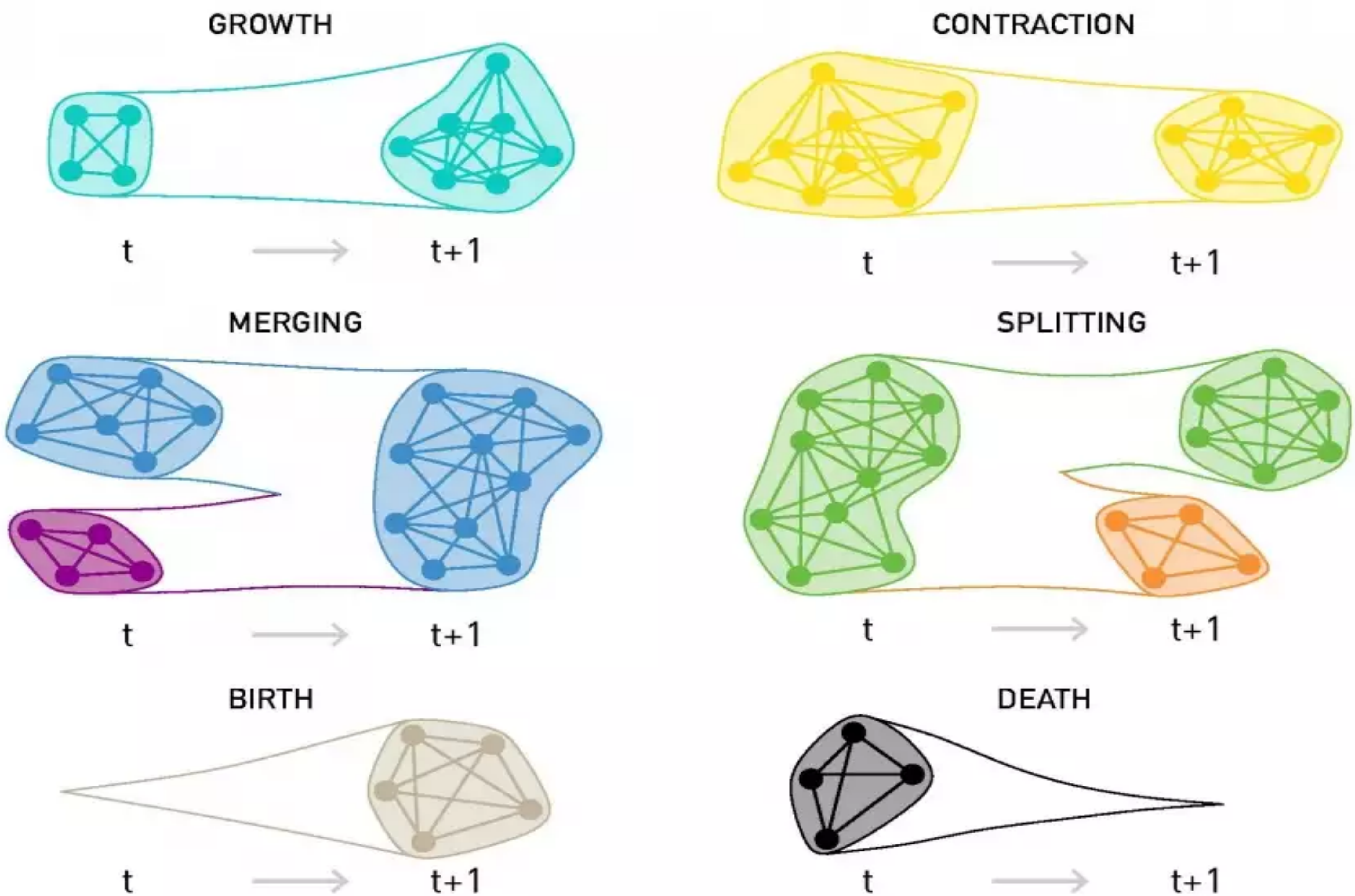
Community Size Distribution



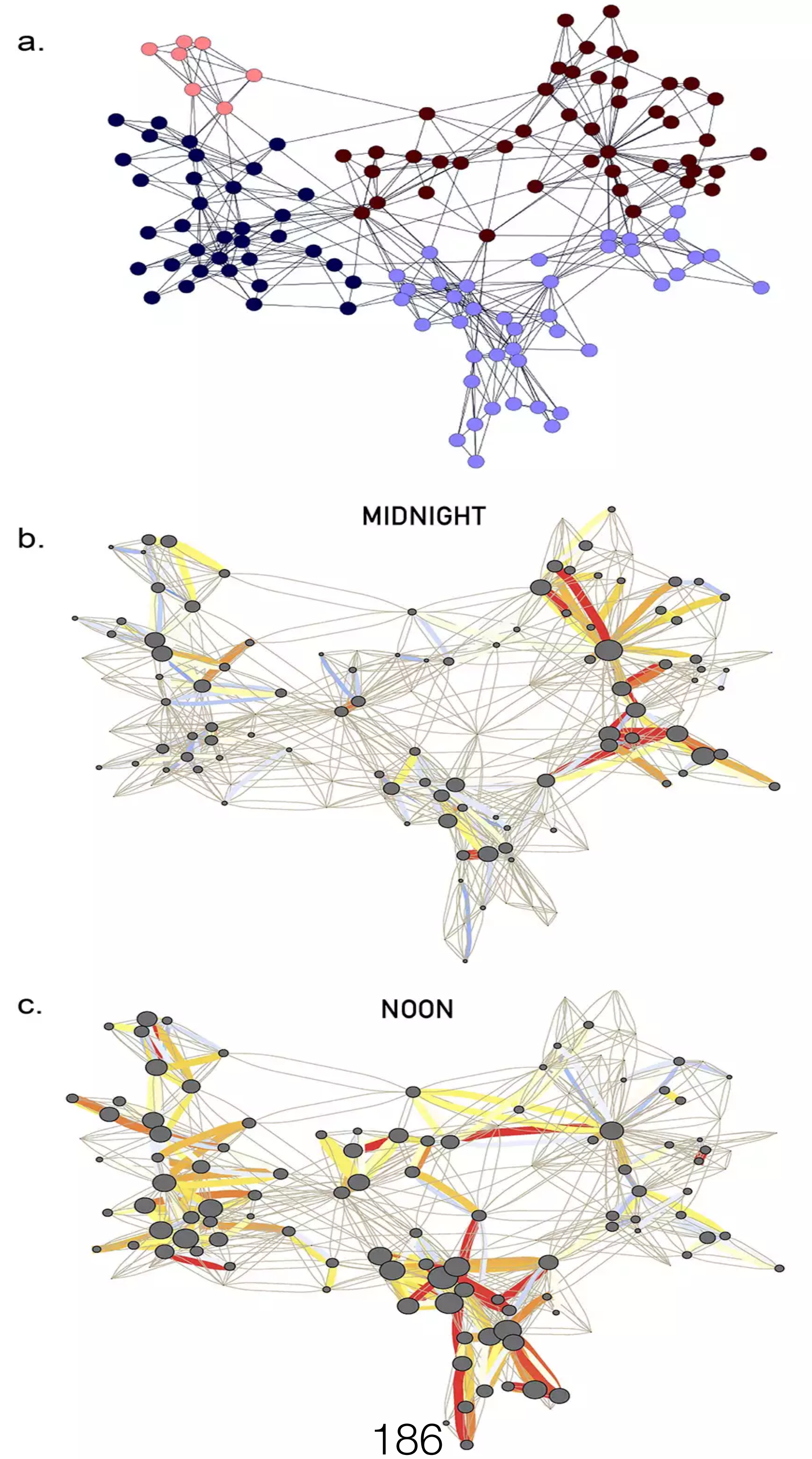
The Running Time



Evolving Communities

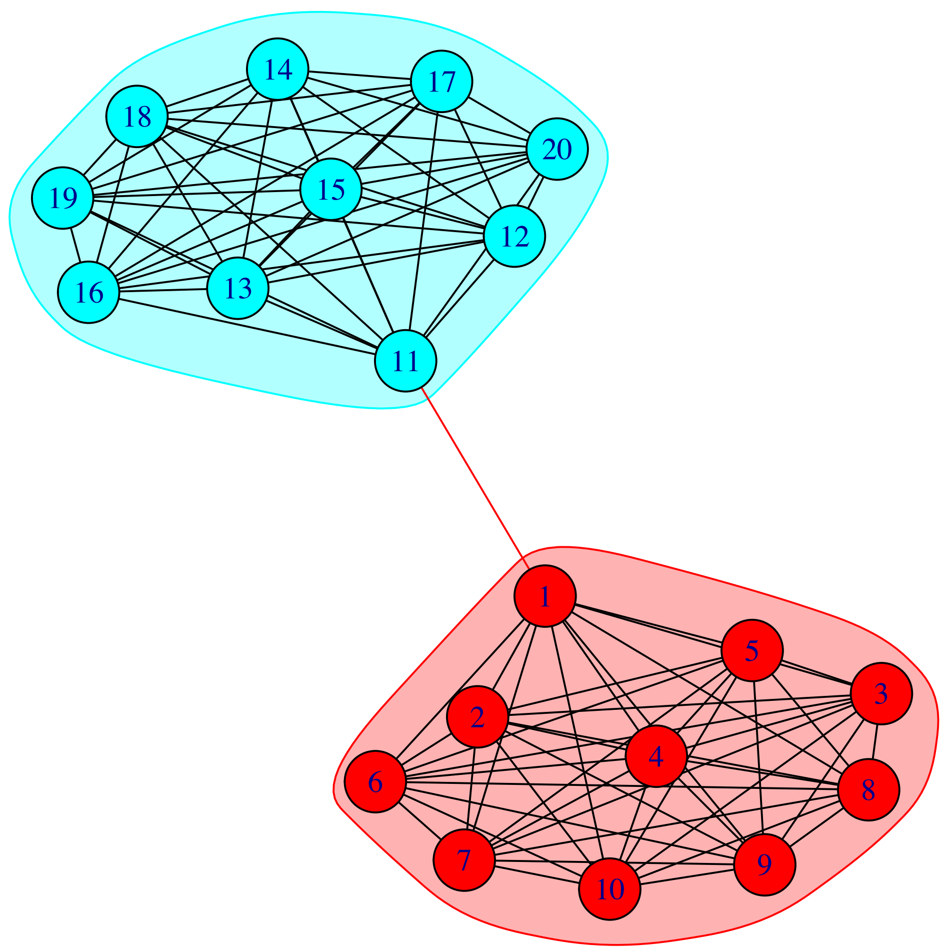


Communities and Call Patterns

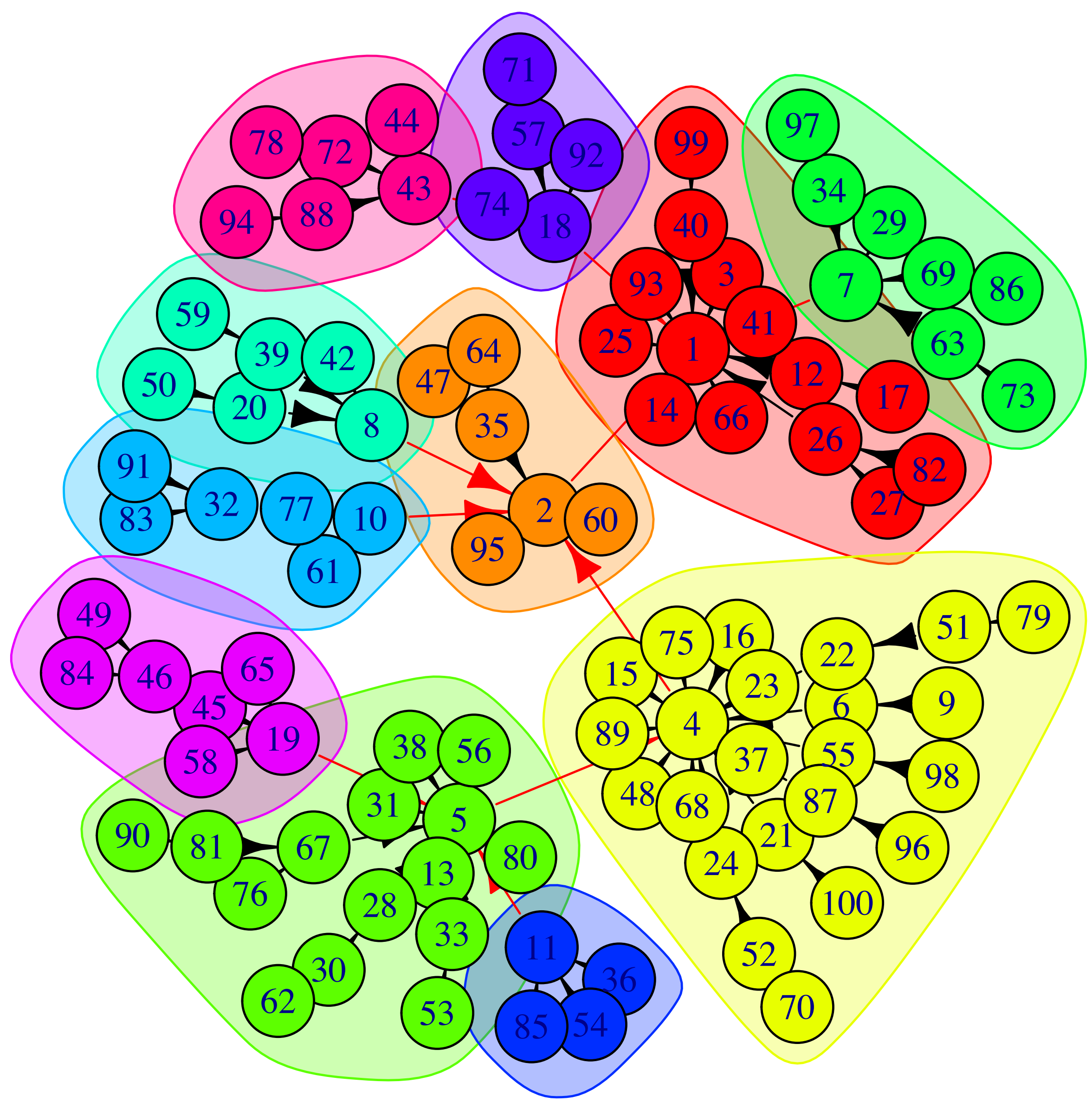


Edge betweenness

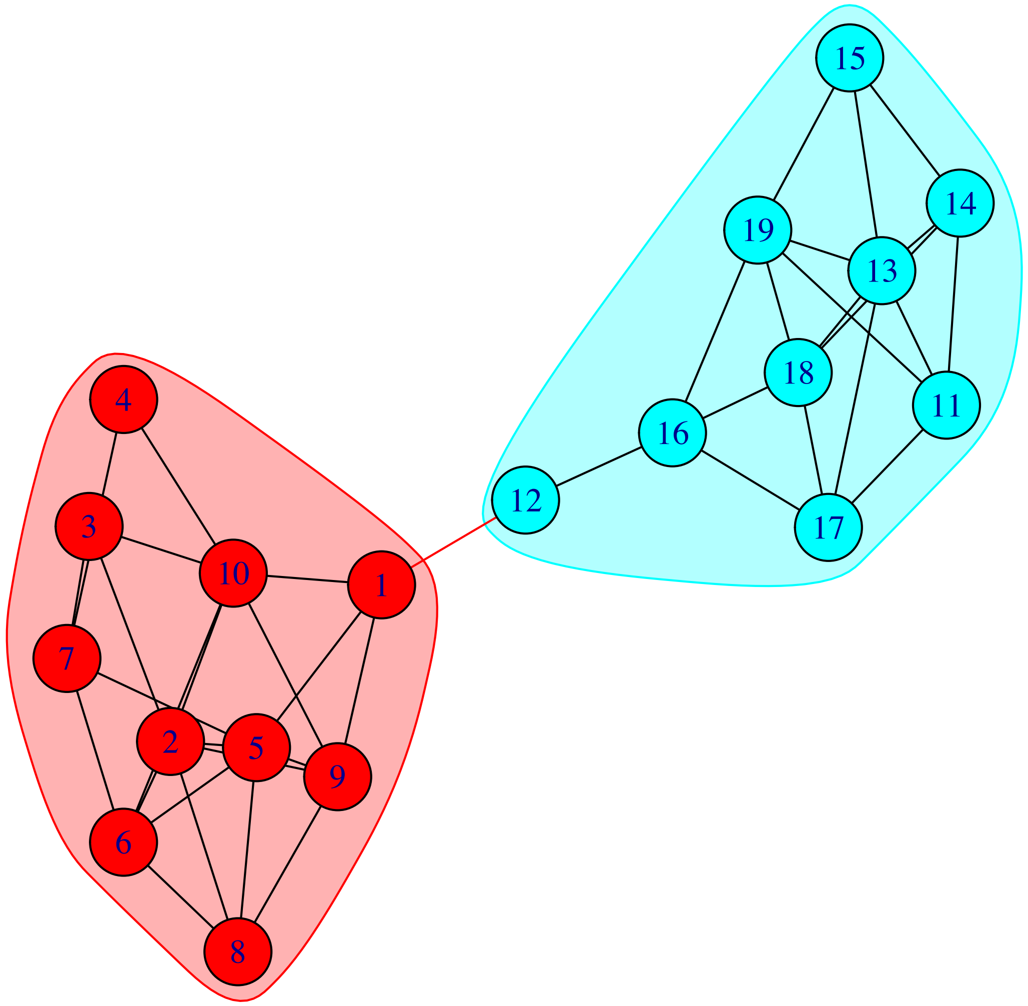
M Newman and M Girvan: Finding and evaluating community structure in networks, Physical Review E 69, 026113 (2004)

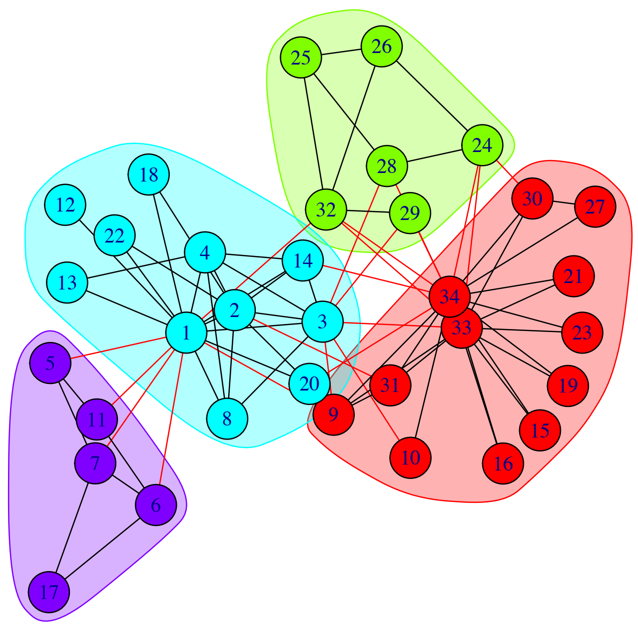
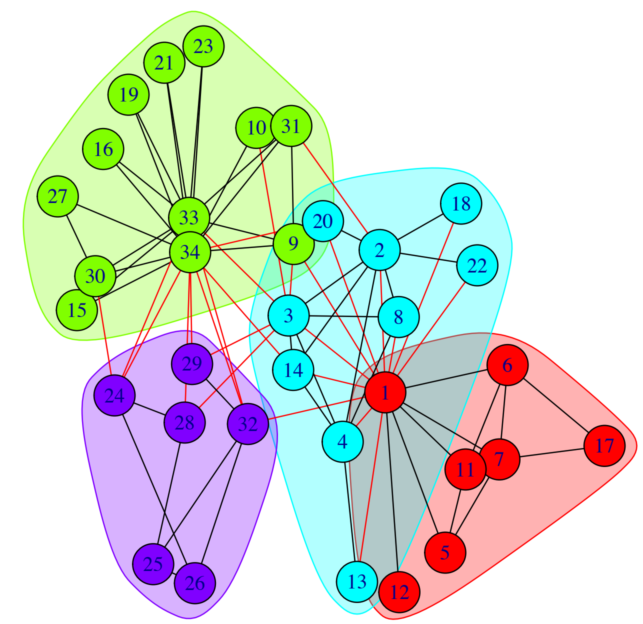


Edge betweenness

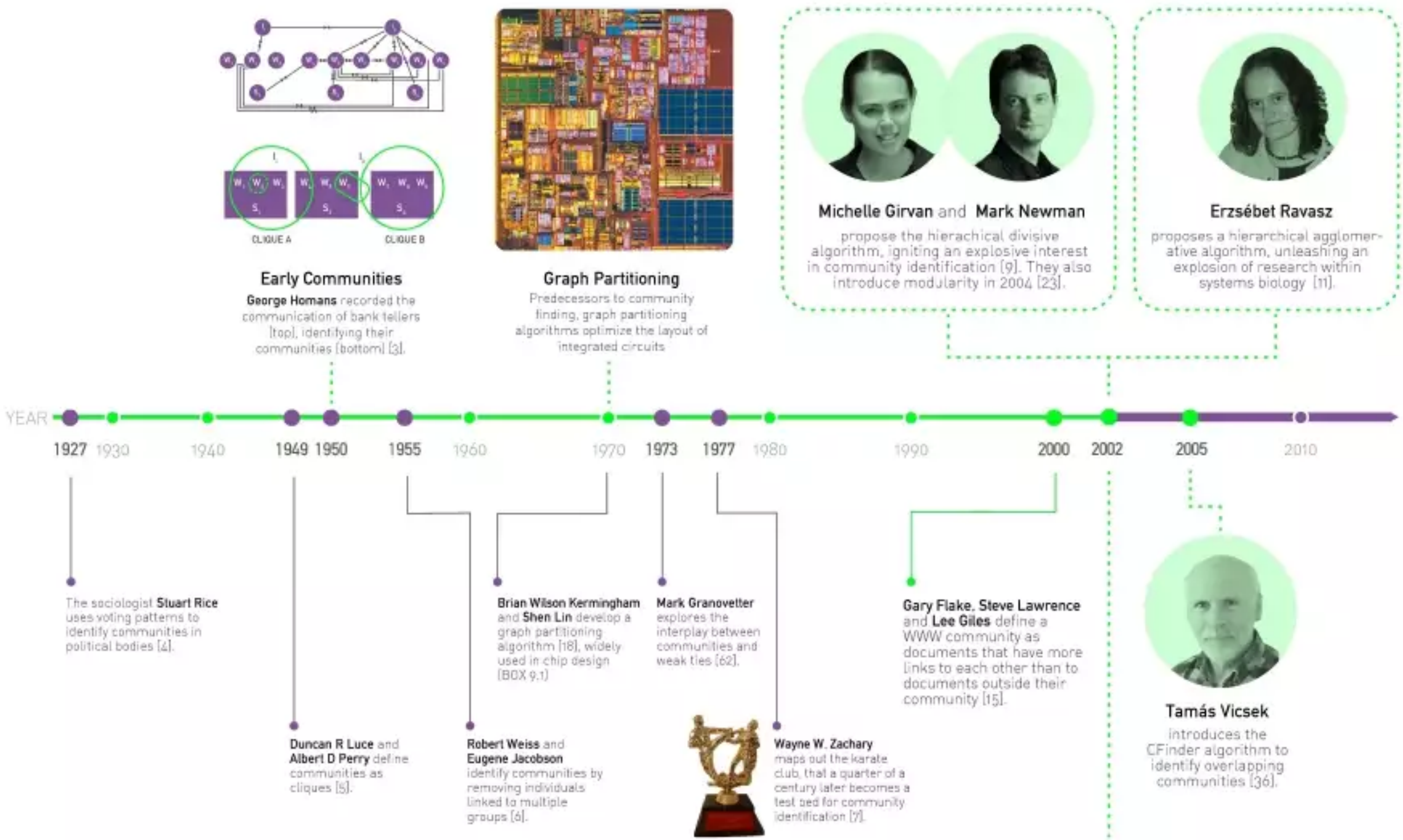


Label propagation

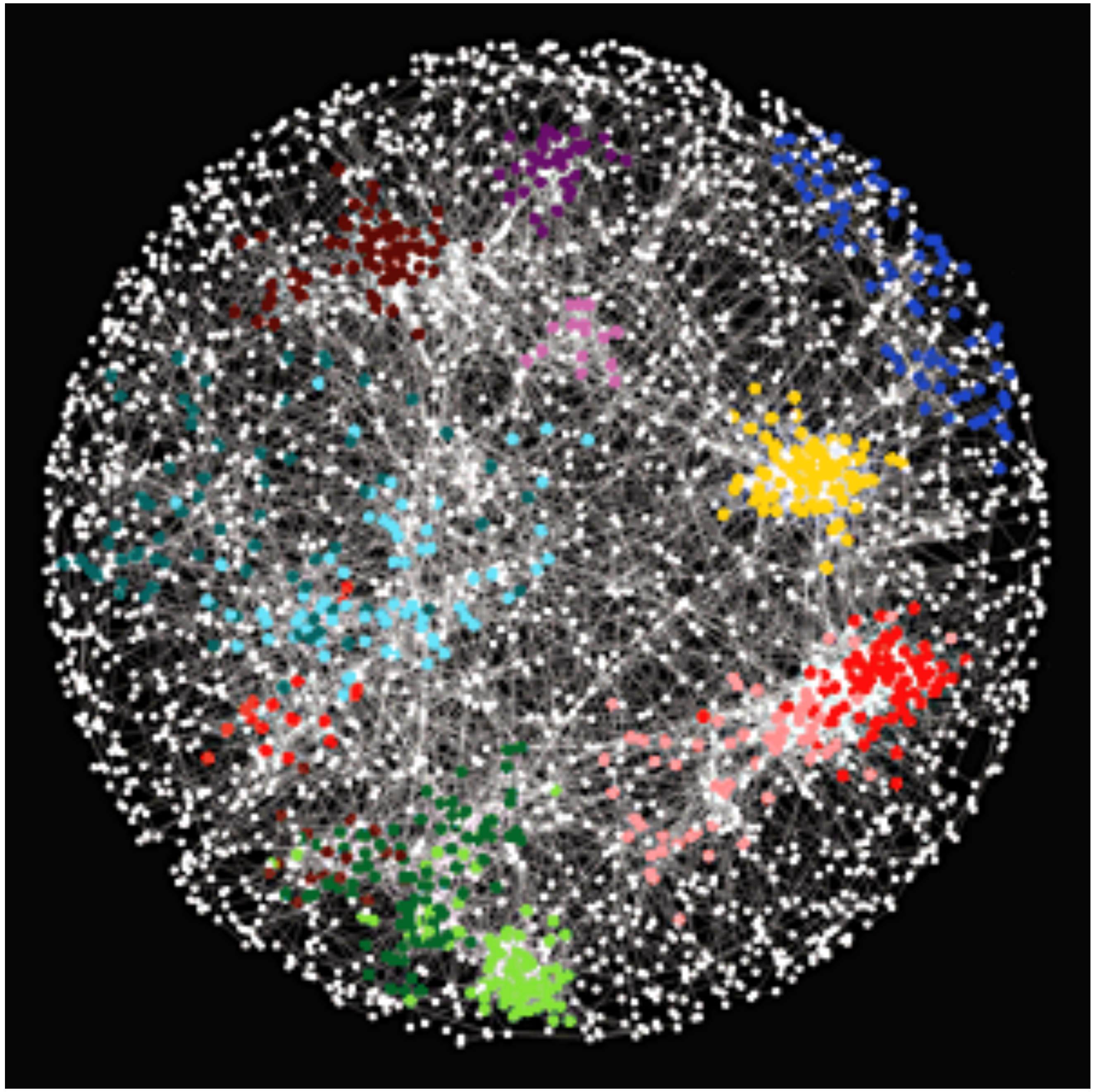


Community Finding: a Brief History



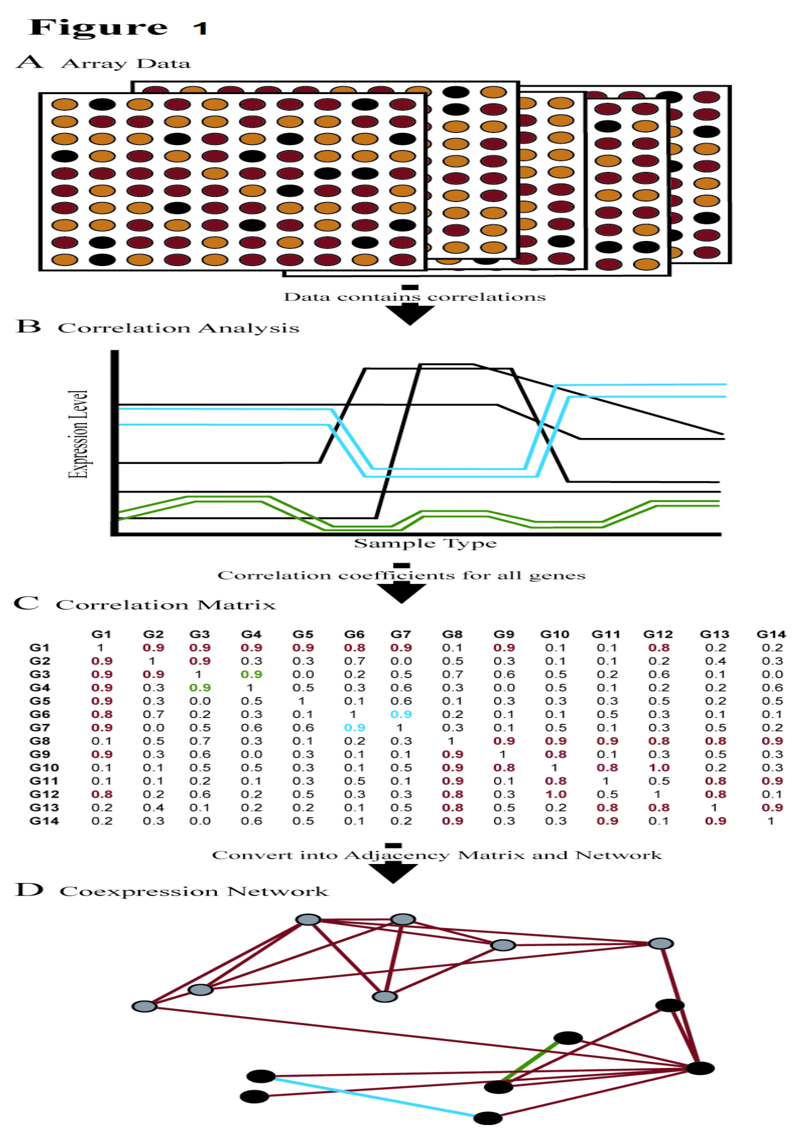
Exercises 6
Communities



6. Inferring Biological Networks

How we can build them?

Co-expression Networks



Correlación

Definición

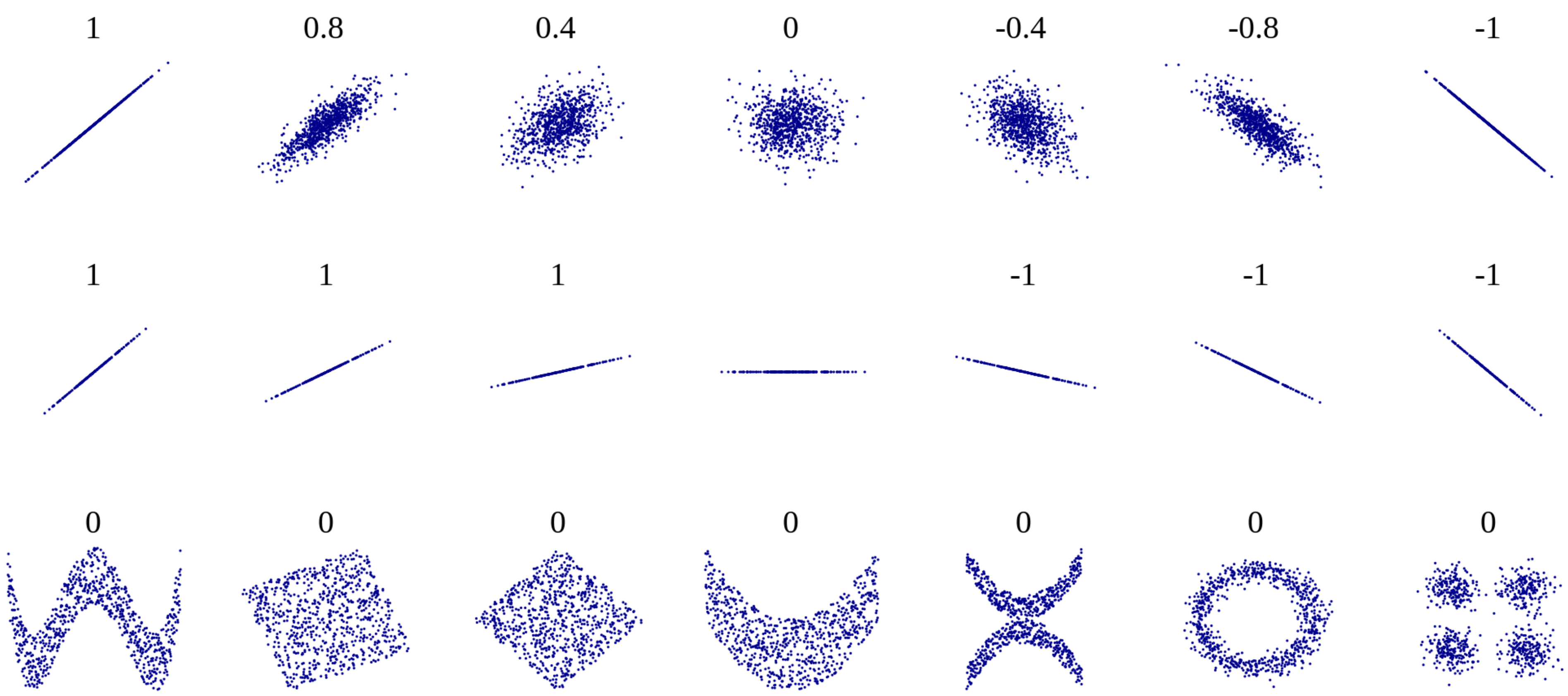
$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y},$$

Coeficiente de correlación de Pearson
Para una muestra de valores de dos variables
$X = X_1, X_2, X_3 \ldots X_n$ y $Y = Y_1, Y_2, Y_3 \ldots, Y_n$

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$$-1 \leq r \leq 1$$

Correlación (de Pearson) y linealidad



Teoría de la información

Entropía Shanon

Sea X una variable aleatoria discreta, con valores $X = X_1, X_2, X_3 \ldots, X_r$, con probabilidad $p_1, p_2, p_3 \ldots, p_r$

$$H(X) = \sum_i^r p_i \log p_i$$

$$0 \leq H(X) \leq \log r$$

Información mutua

Definición

Sean X , Y dos variables aleatorias discretas

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$0 \leq I(X;Y) < \infty$$

Información mutua

Definición

Sean X , Y dos variables aleatorias discretas

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$0 \leq I(X;Y) < \infty$$

ARACNe

ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks

Basso *et al* Nature Genetics (2005)

1. Calcular la MI entre genes.
2. Remover interacciones a terceros vecinos.

ARACNE

Cálculo de la información mutua

$$\begin{pmatrix} Genes & Expresion1 & \dots & ExpresionM \\ G1 & E_{11} & \dots & E_{1M} \\ G2 & E_{21} & \dots & E_{2M} \\ G3 & E_{31} & \dots & E_{3M} \\ \vdots & \vdots & \vdots & \vdots \\ Gn & E_{N1} & \dots & E_{NM} \end{pmatrix}$$

ARACNE

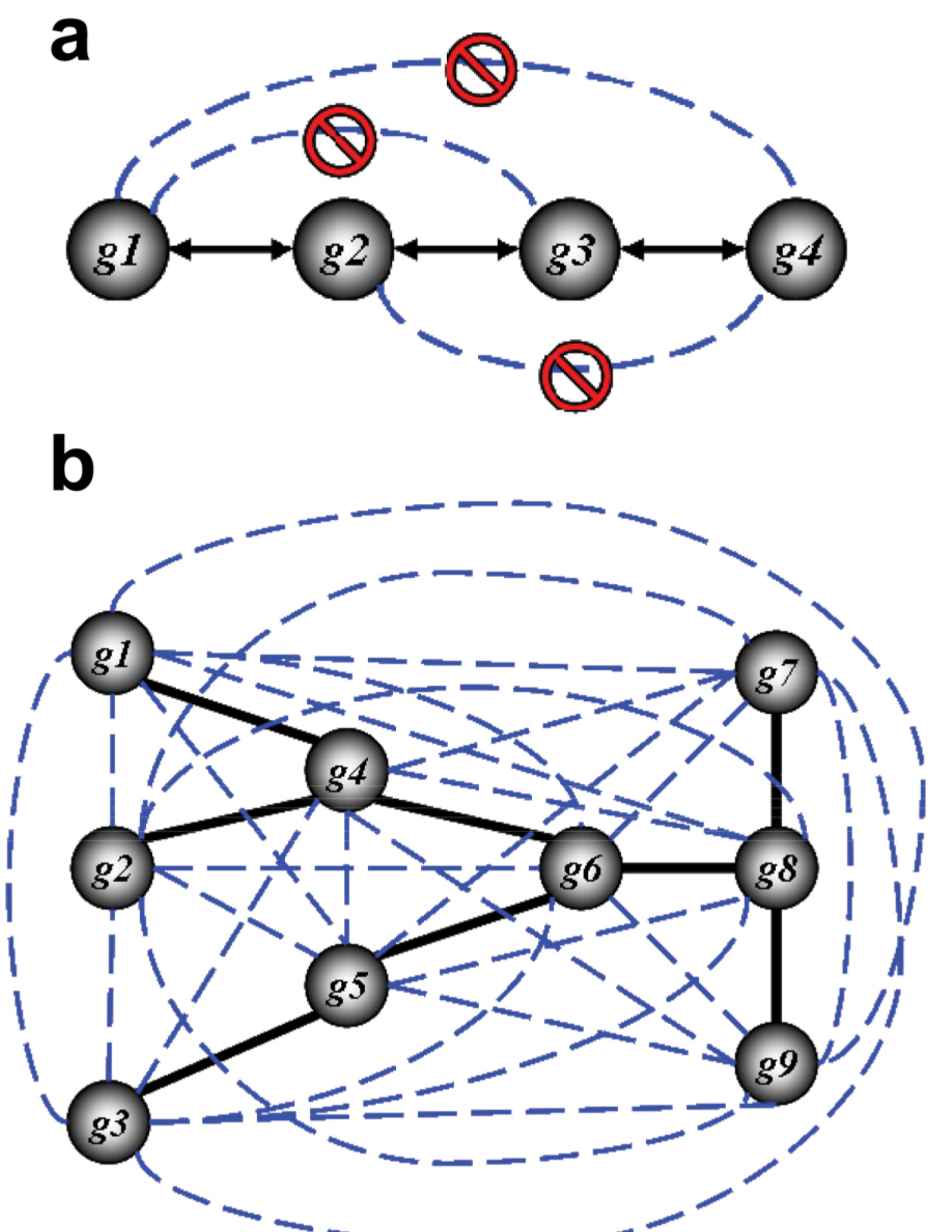
Cálculo de la información mutua

$$\begin{pmatrix} MI(G1,G1) & \dots & MI(G1,GN) \\ MI(G2,G1) & \dots & MI(G2,GN) \\ \vdots & \vdots & \vdots \\ MI(GN,G1) & \dots & MI(GN,GN) \end{pmatrix}$$

ARACNE

DPI (Data Processing Inequality)

$$MI(G1, G3) \leq min[MI(G1, G2); MI(G2, G3)]$$



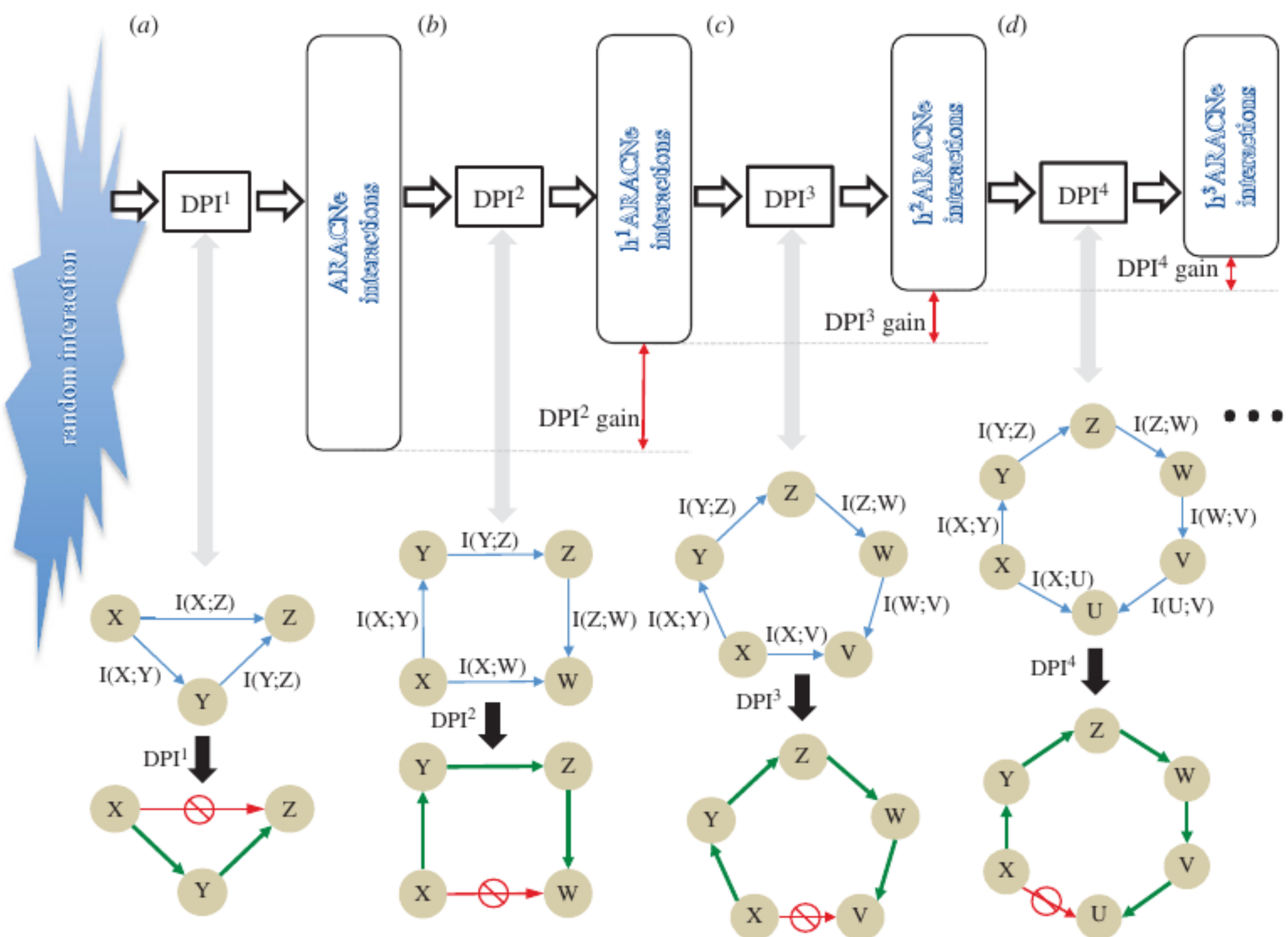
Exercises

Interface Focus

rsfs.royalsocietypublishing.org

hARACNe: improving the accuracy of regulatory model reverse engineering via higher-order data processing inequality tests

In Sock Jang[1], Adam Margolin[1] and Andrea Califano[2]



Maximal information coefficient

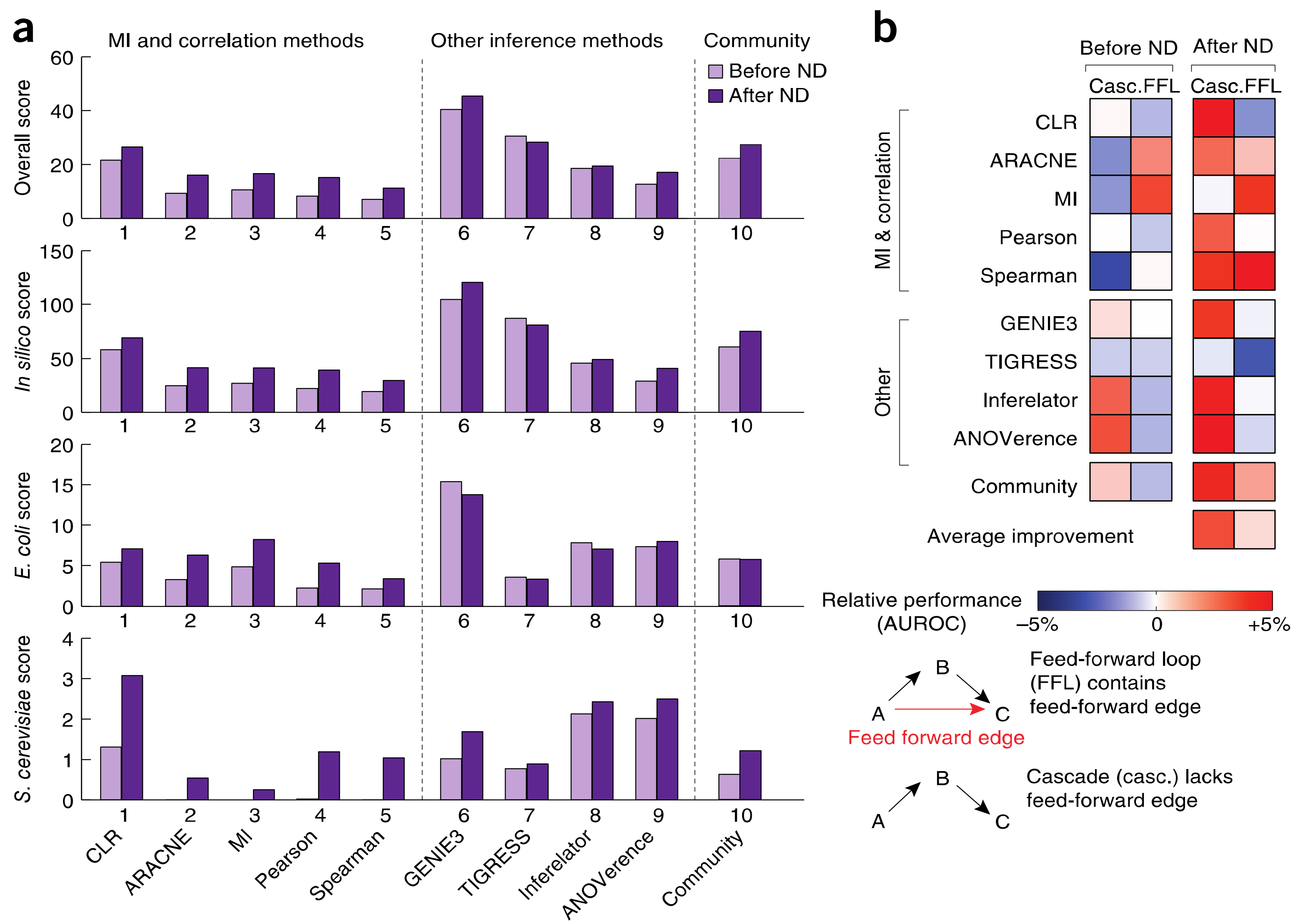
Detecting Novel Associations in Large Data Sets

David N. Reshef,[1,2,3]*† Yakir A. Reshef,[2,4]*† Hilary K. Finucane,[5] Sharon R. Grossman,[2,6]
Gilean McVean,[3,7] Peter J. Turnbaugh,[6] Eric S. Lander,[2,8,9]
Michael Mitzenmacher,[10]‡ Pardis C. Sabeti[2,6]‡

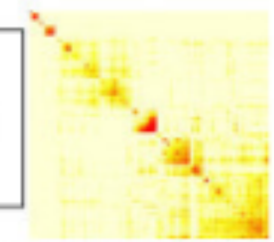
Relationship Type	MIC	Pearson	Spearman	Mutual Information (KDE)	Mutual Information (Kraskov)	CorGC (Principal Curve-Based)	Maximal Correlation
Random	0.18	-0.02	-0.02	0.01	0.03	0.19	0.01
Linear	1.00	1.00	1.00	5.03	3.89	1.00	1.00
Cubic	1.00	0.61	0.69	3.09	3.12	0.98	1.00
Exponential	1.00	0.70	1.00	2.09	3.62	0.94	1.00
Sinusoidal (Fourier frequency)	1.00	-0.09	-0.09	0.01	-0.11	0.36	0.64
Categorical	1.00	0.53	0.49	2.22	1.65	1.00	1.00
Periodic/Linear	1.00	0.33	0.31	0.69	0.45	0.49	0.91
Parabolic	1.00	-0.01	-0.01	3.33	3.15	1.00	1.00
Sinusoidal (non-Fourier frequency)	1.00	0.00	0.00	0.01	0.20	0.40	0.80
Sinusoidal (varying frequency)	1.00	-0.11	-0.11	0.02	0.06	0.38	0.76

Network deconvolution as a general method to distinguish direct dependencies in networks

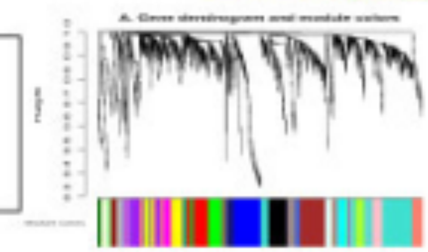
Soheil Feizi[1–3], Daniel Marbach[1,2], Muriel Médard[3] & Manolis Kellis[1,2]



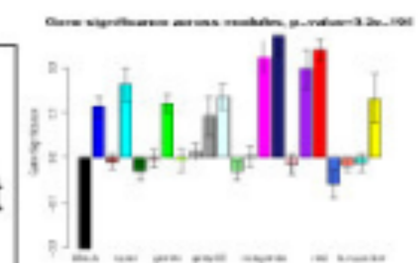
Construct a gene co-expression network
Rationale: make use of interaction patterns among genes
Tools: correlation as a measure of co-expression



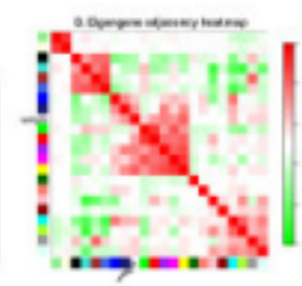
Identify modules
Rationale: module (pathway) based analysis
Tools: hierarchical clustering, Dynamic Tree Cut



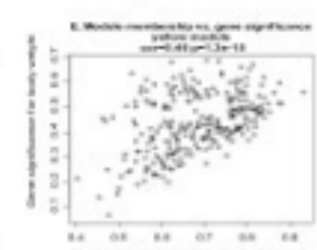
Relate modules to external information
Array Information: clinical data, SNPs, proteomics
Gene Information: ontology, functional enrichement
Rationale: find biologically interesting modules



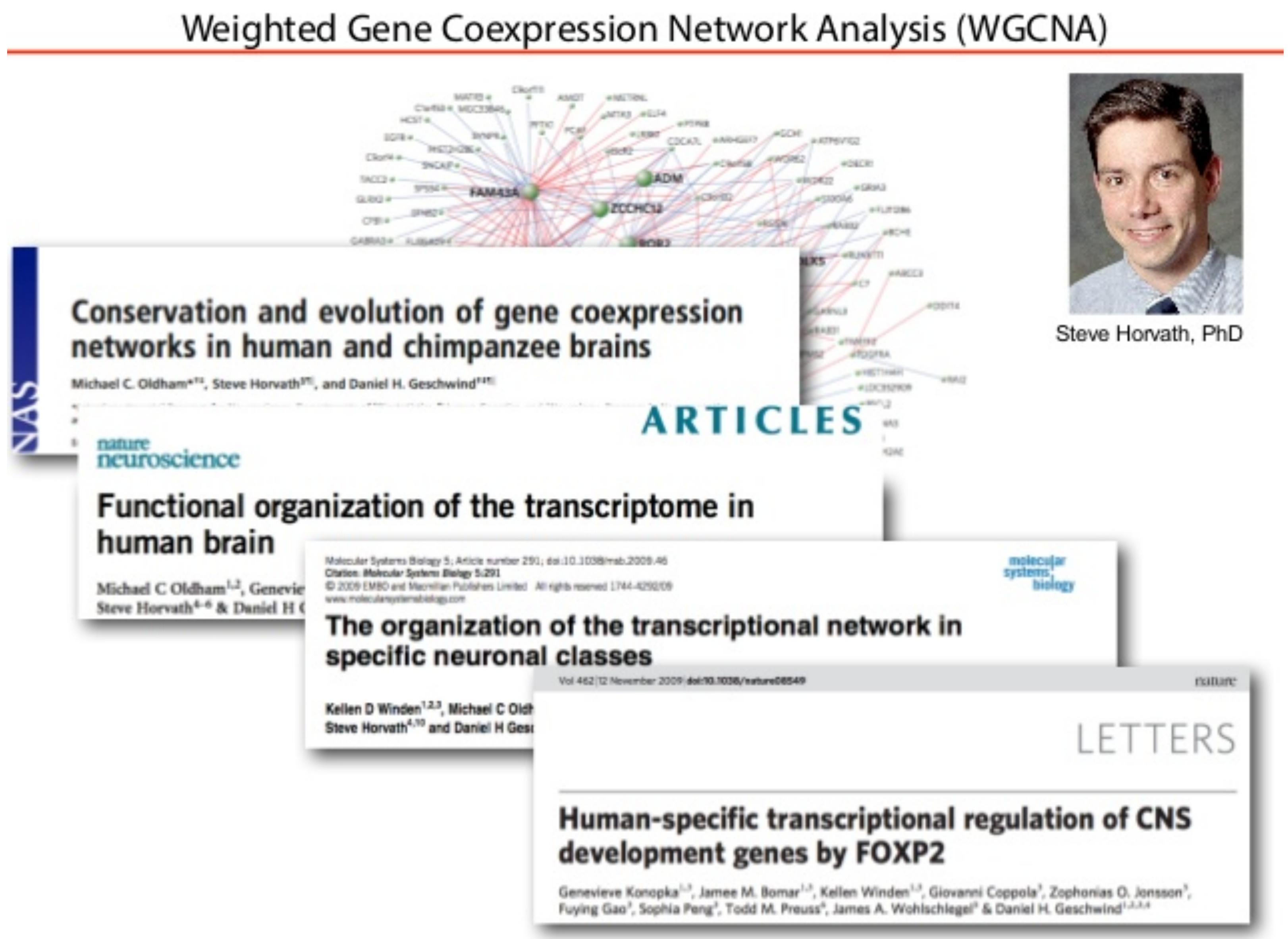
Study module relationships
Rationale: biological data reduction, systems-level view
Tools: Eigengene Networks



Find the key drivers in *interesting* modules
Rationale: experimental validation, biomarkers
Tools: intramodular connectivity, causality testing



Redes genéticas ponderadas



Exercises 7

Inferring Networks