A Partial Overview of Complex Networks

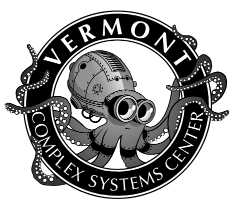
Last updated: 2023/08/22, 11:48:21 EDT

Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 6701, 6713, & a pretend number,
2023–2024| @pocsvox

Prof. Peter Sheridan Dodds | @peterdodds

Computational Story Lab | Vermont Complex Systems Center
Santa Fe Institute | University of Vermont

Outline



Describe | Explain | Create | Share | Ethos: Play

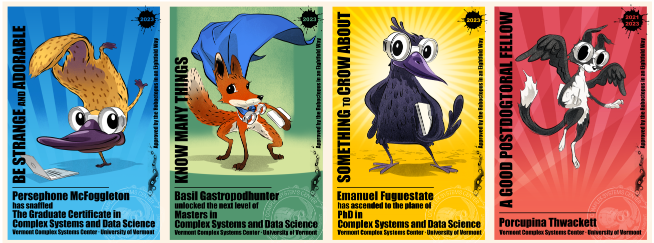
vermontcomplexsystems.org

Leveling up—Scaffolded educational mission:

- Data Science Undergrad.
- Graduate Certificate in Complex Systems and Data Science
- Fall, 2015–: MS in Complex Systems and Data Science
- Fall, 2018–: PhD in ~~The Study of Interesting Things~~ Complex Systems and Data Science

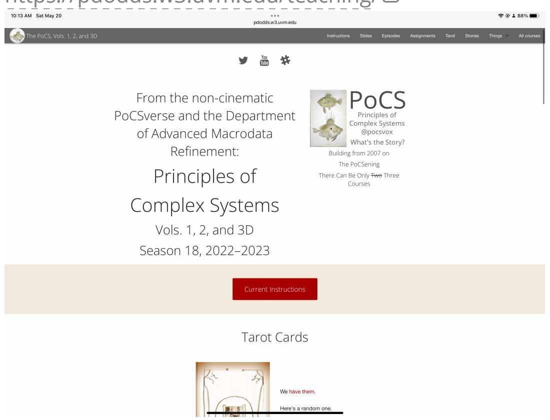
All the words: http://vermontcomplexsystems.org.

Dipoloma-posters:



Principles of Complex Systems, Vols. 1, 2, and 3D

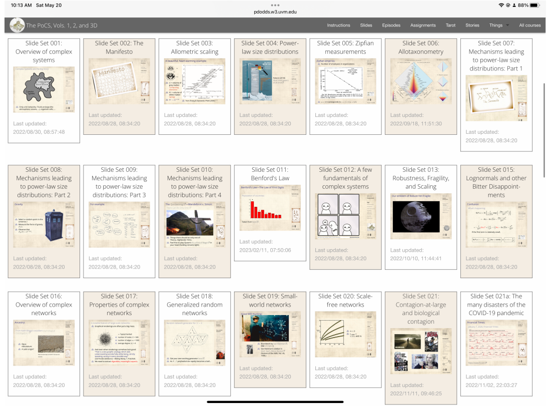
https://pdodds.w3.uvm.edu/teaching/



150,000 lines of LaTeX ...

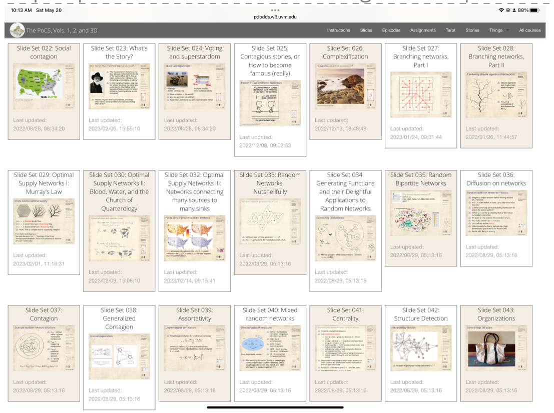
Principles of Complex Systems, Vols. 1, 2, and 3D

https://pdodds.w3.uvm.edu/teaching/courses/pocsverse/slides/



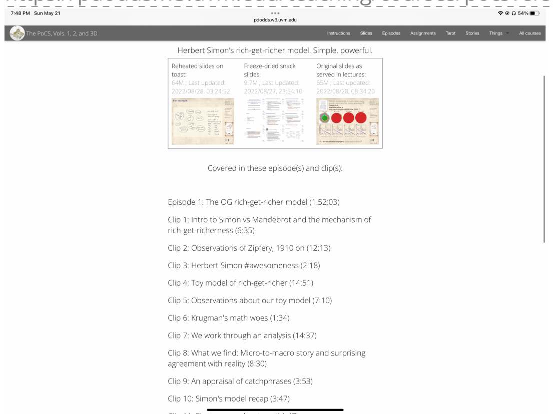
Principles of Complex Systems, Vols. 1, 2, and 3D

https://pdodds.w3.uvm.edu/teaching/courses/pocsverse/slides/



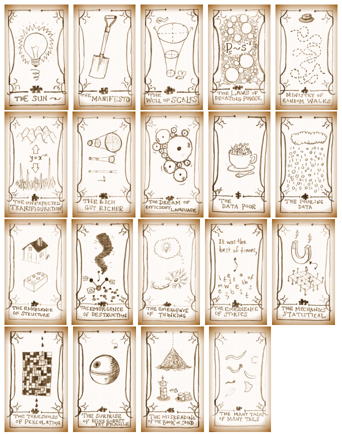
Principles of Complex Systems, Vols. 1, 2, and 3D

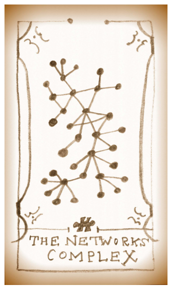
https://pdodds.w3.uvm.edu/teaching/courses/pocsverse/slides/



Exciting details regarding these slides:

- Three servings (all in pdf):
  1. Fresh: For in-class Deliveration.
  2. On toast: Flattened for page-turning joy.
  3. Freeze-dried: Pack-and-go, 3x3 slides per page.
- ~~Presentation versions are hyperly navigable:~~
  ~~back + search + forward.~~
- Web links look like this.
- References in slides link to full citation at end. [4]
- Citations contain links to pdfs for papers (if available).
- Some books will be linked to on Amazon.
- Brought to you by a frightening melange of XƎLATEX, Beamer, perl, PerlTeX, fevered command-line madness, and an almost fanatical devotion to the indomitable emacs.
  #totallynormal





The Science of Complex Systems Manifesto:

1. Systems are ubiquitous and systems matter.
2. Consequently, much of science is about understanding how pieces dynamically fit together.
3. 1700 to 2000 = Golden Age of Reductionism: Atoms!, sub-atomic particles, DNA, genes, people, ...
4. Understanding and creating systems (including new 'atoms') is the greater part of science and engineering.
5. Universality: systems with quantitatively different micro details exhibit qualitatively similar macro behavior (fate, but real and limited)
6. Computing advances make the Science of Complex Systems possible:
   6.1 We can measure and record enormous amounts of data, research areas continue to transition from data scarce to data rich.
   6.2 We can simulate, model, and create complex systems in extraordinary detail.

net•work |ˈnetˌwərk|
noun
1 an arrangement of intersecting horizontal and vertical lines.
• a complex system of roads, railroads, or other transportation routes : *a network of railroads.*
2 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 adj.] *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.

verb [trans.]
connect as or operate with a network : *the stock exchanges have proven to be resourceful in networking these deals.*
• link (machines, esp. computers) to operate interactively : [as adj.] (**networked**) *networked workstations.*
• [intrans.] [often as n.] (**networking**) interact with other people to exchange information and develop contacts, esp. to further one's career : *the skills of networking, bargaining, and negotiation.*

Thesaurus deliciousness:

network
noun
1 *a network of arteries* WEB, lattice, net, matrix, mesh, crisscross, grid, reticulum, reticulation; Anatomy plexus.
2 *a network of lanes* MAZE, labyrinth, warren, tangle.
3 *a network of friends* SYSTEM, complex, nexus, web, webwork.

Ancestry:

From Keith Briggs's etymological investigation:

- Opus reticulatum:
- A Latin origin?

[http://serialconsign.com/2007/11/we-put-net-network]

Ancestry:

First known use: Geneva Bible, 1560

'And thou shalt make unto it a grate like networke of brass (Exodus xxvii 4).'

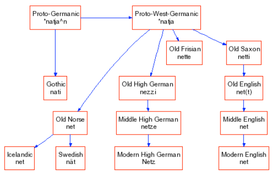
From the OED via Briggs:

- 1658–: reticulate structures in animals
- 1839–: rivers and canals
- 1869–: railways
- 1883–: distribution network of electrical cables
- 1914–: wireless broadcasting networks

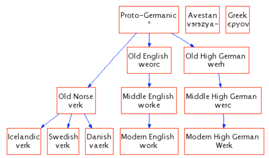
Ancestry:

Net and Work are venerable old words:

- 'Net' first used to mean spider web (King Ælfréd, 888).
- 'Work' appear to have long meant purposeful action.



The network of Germanic `net' words

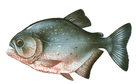


The network of `work' words

- 'Network' = something built based on the idea of natural, flexible lattice or web.
- c.f., ironwork, stonework, fretwork.

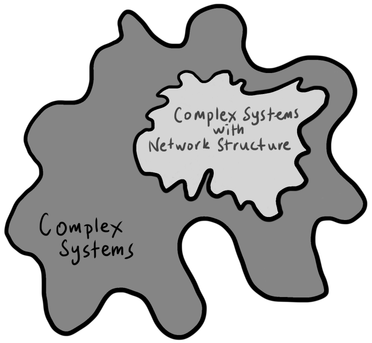
Key Observation:

- Many complex systems can be viewed as complex networks of physical or abstract interactions.
- Opens door to mathematical and numerical analysis.
- Dominant approach of the first decade was of a theoretical-physics/stat-mechish flavor.
- Mindboggling amount of work published on complex networks since 1998 ...
- ... largely due to your typical theoretical physicist:



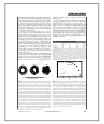
  - *Piranha physicus*
  - Hunt in packs.
  - Feast on new and interesting ideas (see chaos, cellular automata, ...)
  - See also: https://xkcd.com/793/

Complex Systems is the Big Story:



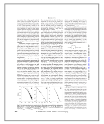
- Only a bit networky: Fluids-at-large (the atmosphere, oceans, ...), organism cells, ...

Popularity (according to Google Scholar)



"Collective dynamics of 'small-world' networks"
Watts and Strogatz,
Nature, **393**, 440–442, 1998. [112]

Times cited: $\sim 51,622$ (as of May 19, 2023)



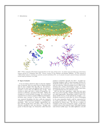
"Emergence of scaling in random networks"
Barabási and Albert,
Science, **286**, 509–511, 1999. [8]

Times cited: $\sim 43,853$ (as of May 19, 2023)

Review articles:

"Complex Networks: Structure and Dynamics"
Boccaletti et al.,
Physics Reports, **424**, 175–308, 2006. [14]

Times cited: ~ 12,318 (as of May 9, 2023)



"The structure and function of complex networks"
M. E. J. Newman,
SIAM Rev., **45**, 167–256, 2003. [77]

Times cited: ~ 23,611 (as of May 9, 2023)

"Statistical mechanics of complex networks"
Albert and Barabási,
Rev. Mod. Phys., **74**, 47–97, 2002. [3]

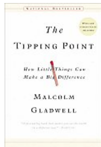
Times cited: ~ 26,636 (as of May 9, 2023)

Popularity according to textbooks:

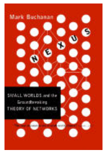
Textbooks:

- Mark Newman (Physics, Michigan)
  "Networks: An Introduction"
- David Easley and Jon Kleinberg (Economics and Computer Science, Cornell)
  "Networks, Crowds, and Markets: Reasoning About a Highly Connected World"

Popularity according to popular books:

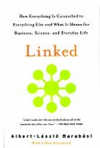


The Tipping Point: How Little Things can make a Big Difference—Malcolm Gladwell [43]

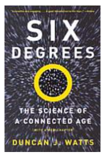


Nexus: Small Worlds and the Groundbreaking Science of Networks—Mark Buchanan

Popularity according to popular books:



Linked: How Everything Is Connected to Everything Else and What It Means—Albert-Laszlo Barabási



Six Degrees: The Science of a Connected Age—Duncan Watts [107]

Numerous others ...

- Complex Social Networks—F. Vega-Redondo [105]
- Fractal River Basins: Chance and Self-Organization—I. Rodríguez-Iturbe and A. Rinaldo [84]
- Random Graph Dynamics—R. Durette
- Scale-Free Networks—Guido Caldarelli
- Evolution and Structure of the Internet: A Statistical Physics Approach—Romu Pastor-Satorras and Alessandro Vespignani
- Complex Graphs and Networks—Fan Chung
- Social Network Analysis—Stanley Wasserman and Kathleen Faust
- Handbook of Graphs and Networks—Eds: Stefan Bornholdt and H. G. Schuster [19]
- Evolution of Networks—S. N. Dorogovtsev and J. F. F. Mendes [34]

More observations

- But surely networks aren't new ...
- Graph theory was well established ...
- Study of social networks started in the 1930's ...
- So why all this 'new' research on networks?
- Answer: Oodles of Easily Accessible Data.
- We can now inform (alas) our theories with a much more measurable reality.*
- Graph theory missed "becoming": Stories = Characters + Time
- A worthy goal: establish mechanistic explanations.

If this is upsetting, maybe string theory is for you ...

More observations

- Internet-scale data sets can be overly exciting.

Witness:

- The End of Theory: The Data Deluge Makes the Scientific Theory Obsolete (Anderson, Wired)
- "The Unreasonable Effectiveness of Data," Halevy et al. [51].
- c.f. Wigner's "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" [114]

But:

- For scientists, description is only part of the battle.
- We still need to understand.

Super Basic definitions

Nodes = A collection of entities which have properties that are somehow related to each other

- e.g., people, forks in rivers, proteins, webpages, organisms, ...

Links = Connections between nodes

- Links may be directed or undirected.
- Links may be binary or weighted.

Other spiffing words: vertices and edges.

Super Basic definitions

Node degree = Number of links per node

- Notation: Node i's degree = k_i.
- k_i = 0,1,2,....
- Notation: the average degree of a network = $\langle k \rangle$ (and sometimes z)
- Connection between number of edges m and average degree:

$$\langle k \rangle = \frac{2m}{N}.$$

- Defn: $\mathcal{N}_i$ = the set of i's k_i neighbors

Super Basic definitions

Adjacency matrix:

- We can represent a network by a matrix A with link weight a_{ij} for nodes i and j in entry (i, j).
- e.g.,

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

- For numerical work, we always use sparse matrices.
- For many real networks, A is a function of time.

Examples

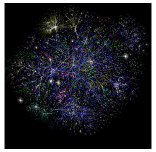
So what passes for a complex network?

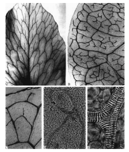
- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks
- The internet (pipes)
- Road networks
- Power grids





- Distribution (branching) versus redistribution (cyclical)

Examples

Interaction networks

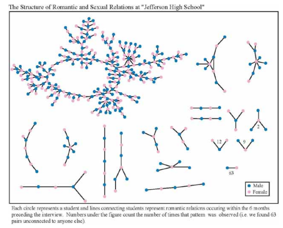
- The Blogosphere (RIP)
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- Airline networks
- Call networks (AT&T)
- The Media
- The internet (World Wide Web)

datamining.typepad.com

Examples

Interaction networks: social networks

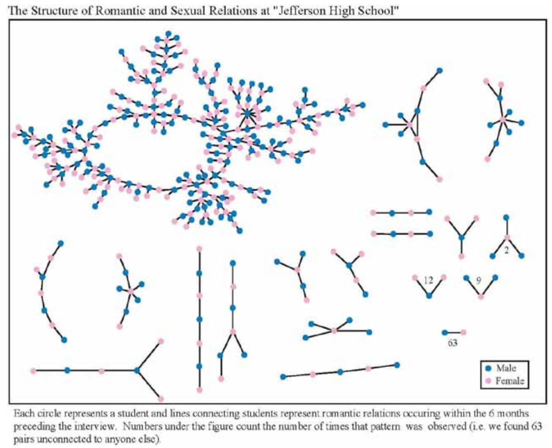
- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- facebook twitter,



(Bearman *et al.*, 2004)

- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

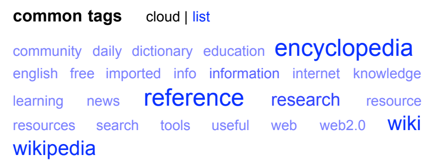
Examples



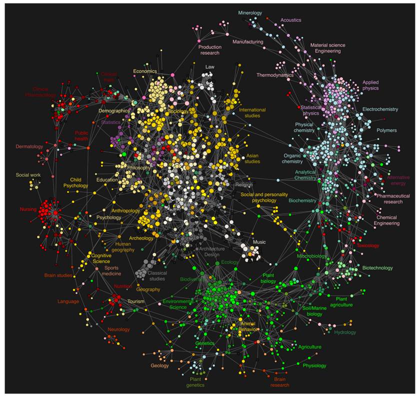
Examples

Relational networks

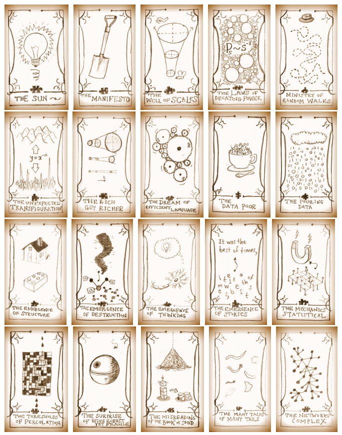
- Consumer purchases (Walmart, Target, Amazon, ...)
- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging, Keywords bit.ly flickr
- Large Language Models



Clickworthy Science:

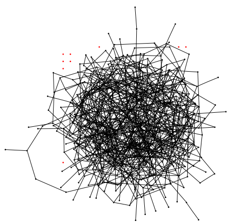


"Clickstream Data Yields High-Resolution Maps of Science", Bollen et al. [18], 2009.



A notable feature of large-scale networks:

- Graphical renderings are often just a big mess.



⇐ Typical hairball
- number of nodes N = 500
- number of edges m = 1000
- average degree $\langle k \rangle$ = 4

- And even when renderings somehow look good: "That is a very graphic analogy which aids understanding wonderfully while being, strictly speaking, wrong in every possible way" said Ponder [Stibbons] —*Making Money*, T. Pratchett.
- We need to extract digestible, meaningful aspects.

Some key aspects of real complex networks:

- degree distribution*
- assortativity
- homophily
- clustering
- motifs
- modularity
- concurrency
- hierarchical scaling
- network distances
- centrality
- efficiency
- interconnectedness
- robustness

- Plus coevolution of network structure and processes on networks.

* Degree distribution is the elephant in the room that we are now all very aware of ...

Properties

1. degree distribution P_k

- P_k is the probability that a randomly selected node has degree k.
- k = node degree = number of connections.
- ex 1: Erdős-Rényi random networks have Poisson degree distributions:

$$P_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

- ex 2: "Scale-free" networks: $P_k \propto k^{-\gamma} \Rightarrow$ 'hubs'.
- link cost controls skew.
- hubs may facilitate or impede contagion.

Properties

Note:

- Erdős-Rényi random networks are a *mathematical construct*.
- 'Scale-free' networks are growing networks that form according to a plausible mechanism.
- Randomness is out there, just not to the degree of a completely random network.
- "Becoming": Stories = Characters + Time

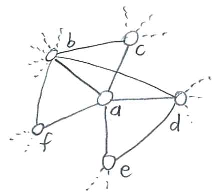
Properties

2. Assortativity/3. Homophily:

- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network: [74] similar degree nodes connecting to each other.
  Often social: company directors, coauthors, actors.
- Disassortative network: high degree nodes connecting to low degree nodes.
  Often techological or biological: internet, WWW, protein interactions, neural networks, food webs.

Local socialness:

4. Clustering:



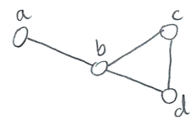
- Your friends tend to know each other.
- Two measures (explained on following slides):
  1. Watts & Strogatz [112]

$$C_1 = \left\langle \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2} \right\rangle_i$$

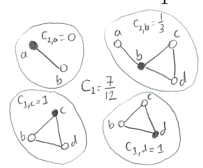
  2. Newman [77]

$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}$$

Example network:



Calculation of C_1:



- C_1 is the average fraction of pairs of neighbors who are connected.
- Fraction of pairs of neighbors who are connected is

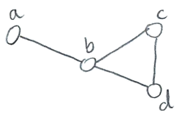
$$\frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2}$$

where k_i is node i's degree, and $\mathcal{N}_i$ is the set of i's neighbors.
- Averaging over all nodes, we have:

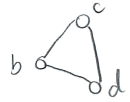
$$C_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2} = \left\langle \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2} \right\rangle_i$$

Triples and triangles

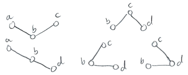
Example network:



Triangles:



Triples:



- Nodes i_1, i_2, and i_3 form a triple around i_1 if i_1 is connected to i_2 and i_3.
- Nodes i_1, i_2, and i_3 form a triangle if each pair of nodes is connected
- The definition $C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}$ measures the fraction of closed triples
- The '3' appears because for each triangle, we have 3 closed triples.
- Social Network Analysis (SNA): fraction of transitive triples.

Clustering:

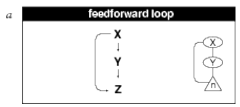
Sneaky counting for undirected, unweighted networks:

- If the path i–j–ℓ exists then $a_{ij}a_{j\ell} = 1$.
- Otherwise, $a_{ij}a_{j\ell} = 0$.
- We want $i \neq \ell$ for good triples.
- In general, a path of n edges between nodes i_1 and i_n travelling through nodes i_2, i_3, ...i_{n-1} exists $\iff a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} \cdots a_{i_{n-2} i_{n-1}} a_{i_{n-1} i_n} = 1$.
- $$\#\text{triples} = \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{\ell=1}^{N}[A^2]_{i\ell} - \text{Tr}A^2\right)$$
- $$\#\text{triangles} = \frac{1}{6}\text{Tr}A^3$$

Properties

5. motifs:

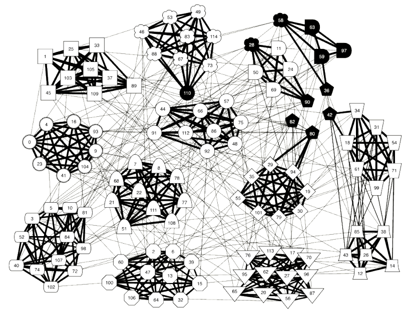
- small, recurring functional subnetworks
- e.g., Feed Forward Loop:



Shen-Orr, Uri Alon, *et al.* [89]

Properties

6. modularity and structure/community detection:



Clauset *et al.*, 2006 [24]: NCAA football

Properties

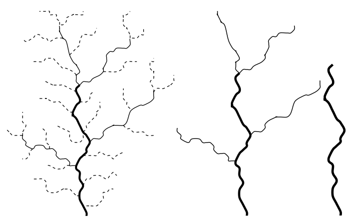
7. concurrency:

- transmission of a contagious element only occurs during contact
- rather obvious but easily missed in a simple model
- dynamic property—static networks are not enough
- knowledge of previous contacts crucial
- beware cumulated network data
- Kretzschmar and Morris, 1996 [58]
- "Temporal networks" become a concrete area of study for Piranha Physicus in 2013.

Properties

8. Horton-Strahler ratios:

- Metrics for branching networks:
  - Method for ordering streams hierarchically
  - Number: $R_n = N_\omega / N_{\omega+1}$
  - Segment length: $R_l = \langle l_{\omega+1} \rangle / \langle l_\omega \rangle$
  - Area/Volume: $R_a = \langle a_{\omega+1} \rangle / \langle a_\omega \rangle$



Properties

9. network distances:

(a) shortest path length d_{ij}:

- Fewest number of steps between nodes i and j.
- (Also called the chemical distance between i and j.)

(b) average path length $\langle d_{ij} \rangle$:

- Average shortest path length in whole network.
- Good algorithms exist for calculation.
- Weighted links can be accommodated.

Properties

9. network distances:

- network diameter d_{max}: Maximum shortest path length between any two nodes.
- closeness $d_{\text{cl}} = [\sum_{ij} d_{ij}^{-1} / \binom{n}{2}]^{-1}$: Average 'distance' between any two nodes.
- Closeness handles disconnected networks ($d_{ij} = \infty$)
- $d_{\text{cl}} = \infty$ only when all nodes are isolated.
- Closeness perhaps compresses too much into one number

Properties

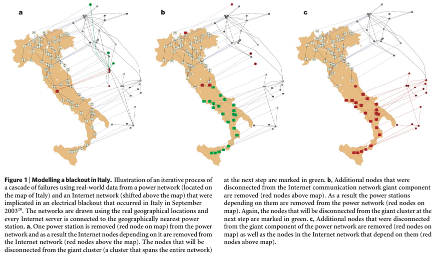
10. centrality:

- Many such measures of a node's 'importance.'
- ex 1: Degree centrality: k_i.
- ex 2: Node i's betweenness = fraction of shortest paths that pass through i.
- ex 3: Edge ℓ's betweenness = fraction of shortest paths that travel along ℓ.
- ex 4: Recursive centrality: Hubs and Authorities (Jon Kleinberg [56])

Properties

Interconnected networks and robustness (two for one deal):

"Catastrophic cascade of failures in interdependent networks"[21]. Buldyrev et al., Nature 2010.



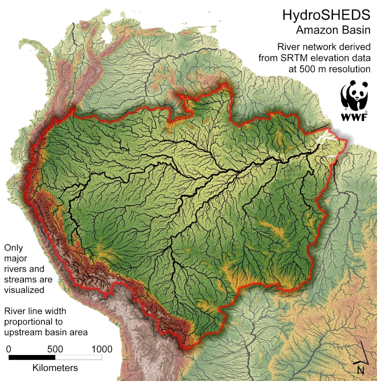
Branching networks are useful things:

- Fundamental to material supply and collection
- Supply: From one source to many sinks in 2- or 3-d.
- Collection: From many sources to one sink in 2- or 3-d.
- Typically observe hierarchical, recursive self-similar structure

Examples:

- River networks
- Cardiovascular networks
- Plants
- Evolutionary trees
- Organizations (only in theory ...)

Branching networks are everywhere ...

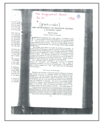


http://hydrosheds.cr.usgs.gov/

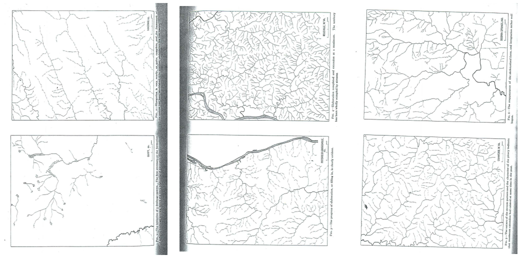
Branching networks are everywhere ...

http://en.wikipedia.org/wiki/Image:Applebox.JPG

An early thought piece: Extension and Integration



"The Development of Drainage Systems: A Synoptic View"
Waldo S. Glock,
The Geographical Review, **21**, 475–482, 1931.[45]



Initiation, Elongation — Elaboration, Piracy. — Abstraction, Absorption.

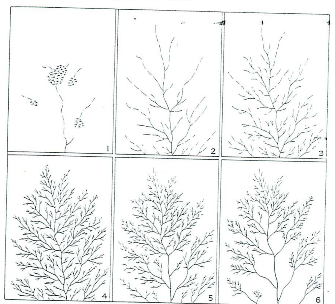


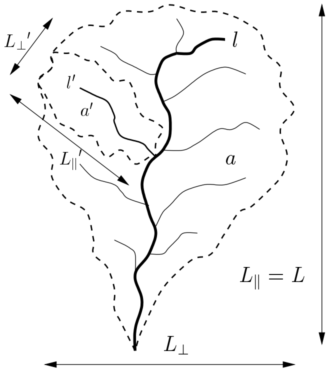
FIG. 6—An ideal diagrammatic summary of the development of a drainage system given for purposes of comparison only. The first four parts show extension, thus: 1, initiation; 2, elongation; 3, elaboration; and 4, maximum extension. Parts 5 and 6 represent steps during integration.

The sequential stages recognized in the evolution of a drainage system are "extension" and "integration"; the first, a stage of increasing complexity; the second, of simplification.

Allometry

- Isometry: dimensions scale linearly with each other.
- Allometry: dimensions scale nonlinearly.

Basin allometry



Allometric relationships:

- $$\ell \propto a^h$$
- $$\ell \propto L^d$$
- Combine above:

$$a \propto L^{d/h} \equiv L^D$$

'Laws'

- Hack's law (1957)[50]:

$$\boxed{\ell \propto a^h}$$

reportedly $0.5 < h < 0.7$

- Scaling of main stream length with basin size:

$$\boxed{\ell \propto L_\parallel^d}$$

reportedly $1.0 < d < 1.1$

- Basin allometry:

$$\boxed{L_\parallel \propto a^{h/d} \equiv a^{1/D}}$$

$D < 2 \rightarrow$ basins elongate.

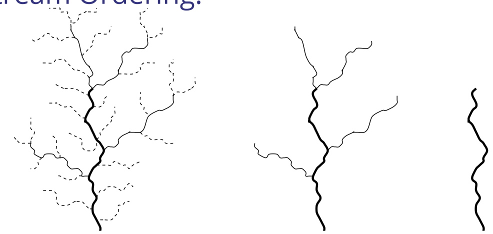
There are a few more 'laws': [31]

Relation:	Name or description:
$T_k = T_1(R_T)^{k-1}$	Tokunaga's law
$\ell \sim L^d$	self-affinity of single channels
$n_\omega/n_{\omega+1} = R_n$	Horton's law of stream numbers
$\bar{\ell}_{\omega+1}/\bar{\ell}_\omega = R_\ell$	Horton's law of main stream lengths
$\bar{a}_{\omega+1}/\bar{a}_\omega = R_a$	Horton's law of basin areas
$\bar{s}_{\omega+1}/\bar{s}_\omega = R_s$	Horton's law of stream segment lengths
$L_\perp \sim L^H$	scaling of basin widths
$P(a) \sim a^{-\tau}$	probability of basin areas
$P(\ell) \sim \ell^{-\gamma}$	probability of stream lengths
$\ell \sim a^h$	Hack's law
$a \sim L^D$	scaling of basin areas
$\Lambda \sim a^\beta$	Langbein's law
$\lambda \sim L^\varphi$	variation of Langbein's law

Reported parameter values: [31]

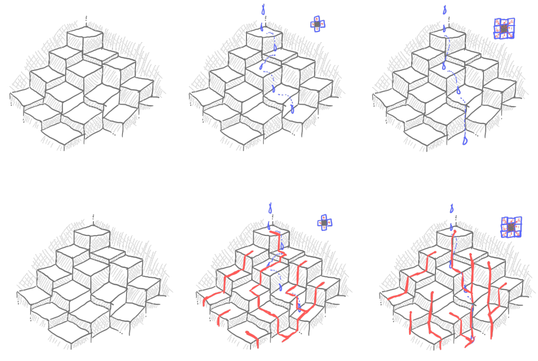
Parameter:	Real networks:
R_n	3.0–5.0
R_a	3.0–6.0
$R_\ell = R_T$	1.5–3.0
T_1	1.0–1.5
d	1.1 ± 0.01
D	1.8 ± 0.1
h	0.50–0.70
τ	1.43 ± 0.05
γ	1.8 ± 0.1
H	0.75–0.80
β	0.50–0.70
φ	1.05 ± 0.05

Stream Ordering:



1. Label all source streams as order $\omega = 1$ and remove.
2. Label all new source streams as order $\omega = 2$ and remove.
3. Repeat until one stream is left (order = Ω)
4. Basin is said to be of the order of the last stream removed.
5. Example above is a basin of order $\Omega = 3$.

Basic algorithm for extracting networks from Digital Elevation Models (DEMs):



- Also: /Users/dodds/work/rivers/1998dems/kevinlakewaster.c

Horton's laws

Self-similarity of river networks

- First quantified by Horton (1945) [53], expanded by Schumm (1956) [88]

Three laws:

- Horton's law of stream numbers:

$$\boxed{n_\omega/n_{\omega+1} = R_n > 1}$$

- Horton's law of stream lengths:

$$\boxed{\bar{\ell}_{\omega+1}/\bar{\ell}_\omega = R_\ell > 1}$$

- Horton's law of basin areas:

$$\boxed{\bar{a}_{\omega+1}/\bar{a}_\omega = R_a > 1}$$

Network Architecture

Tokunaga's law [101, 102, 103]

- Property 1: Scale independence—depends only on difference between orders:

$$T_{\mu,\nu} = T_{\mu-\nu}$$

- Property 2: Number of side streams grows exponentially with difference in orders:

$$T_{\mu,\nu} = T_1(R_T)^{\mu-\nu-1}$$

- We usually write Tokunaga's law as:

$$\boxed{T_k = T_1(R_T)^{k-1}} \quad \text{where } R_T \simeq 2$$

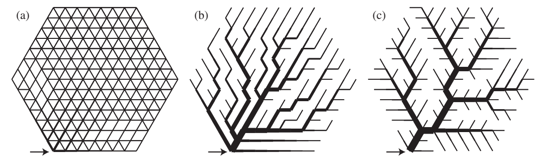
.

Connecting exponents

Only 3 parameters are independent: e.g., take d, R_n, and R_s

relation:	scaling relation/parameter: [31]
$\ell \sim L^d$	d
$T_k = T_1(R_T)^{k-1}$	$T_1 = R_n - R_s - 2 + 2R_s/R_n$
	$R_T = R_s$
$n_\omega/n_{\omega+1} = R_n$	R_n
$\bar{a}_{\omega+1}/\bar{a}_\omega = R_a$	$R_a = R_n$
$\bar{\ell}_{\omega+1}/\bar{\ell}_\omega = R_\ell$	$R_\ell = R_s$
$\ell \sim a^h$	$h = \ln R_s / \ln R_n$
$a \sim L^D$	$D = d/h$
$L_\perp \sim L^H$	$H = d/h - 1$
$P(a) \sim a^{-\tau}$	$\tau = 2 - h$
$P(\ell) \sim \ell^{-\gamma}$	$\gamma = 1/h$
$\Lambda \sim a^\beta$	$\beta = 1 + h$
$\lambda \sim L^\varphi$	$\varphi = d$

Single source optimal supply



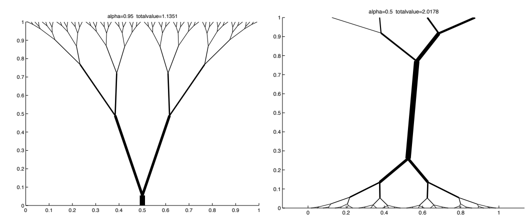
(a) $\gamma > 1$: Braided (bulk) flow
(b) $\gamma < 1$: Local minimum: Branching flow
(c) $\gamma < 1$: Global minimum: Branching flow

- Note: This is a single source supplying a region.

From Bohn and Magnasco [16]
See also Banavar *et al.* [6]: "Topology of the Fittest Transportation Network"; focus is on presence or absence of loops—same story

Single source optimal supply

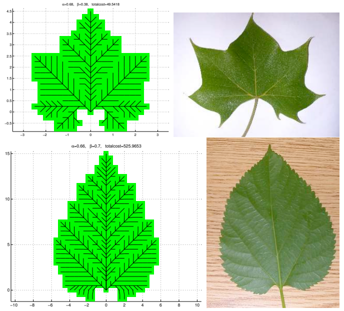
Optimal paths related to transport (Monge) problems:



"Optimal paths related to transport problems"
Qinglan Xia,
Communications in Contemporary Mathematics, **5**, 251–279, 2003. [116]

Growing networks: [117]

FIGURE 3. A maple leaf



- Top: $\alpha = 0.66$, $\beta = 0.38$; Bottom: $\alpha = 0.66$, $\beta = 0.70$

Single source optimal supply

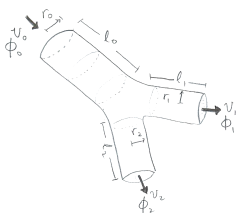
An immensely controversial issue ...

- The form of natural branching networks: Random, optimal, or some combination? [55, 113, 7, 33, 27]
- River networks, blood networks, trees, ...

Two observations:

- Self-similar networks appear everywhere in nature for single source supply/single sink collection.
- Real networks differ in details of scaling but reasonably agree in scaling relations.

Optimization—Murray's law



- Murray's law (1926) connects branch radii at forks: [72, 71, 73, 59, 100]

$$\boxed{r^3_{\text{parent}} = r^3_{\text{offspring1}} + r^3_{\text{offspring2}}}$$

where r_{parent} = radius of 'parent' branch, and $r_{\text{offspring1}}$ and $r_{\text{offspring2}}$ are radii of the two 'offspring' sub-branches.

- Holds up well for outer branchings of blood networks [90].
- Also found to hold for trees [73, 66] when xylem is not a supporting structure [67].
- See D'Arcy Thompson's "On Growth and Form" for background and general inspiration [99, 100].

Animal power

Fundamental biological and ecological constraint:

$$\boxed{P = c\,M^{\alpha}}$$

P = basal metabolic rate

M = organismal body mass

Stories—The Fraction Assassin:

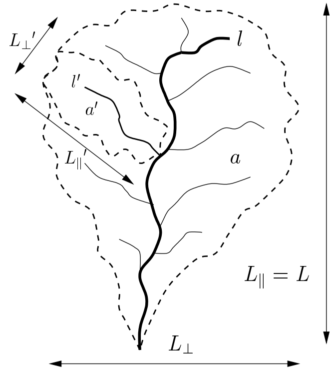
Quarterology spreads throughout the land:

The Cabal assassinates 2/3-scaling:

- 1964: Troon, Scotland.
- 3rd Symposium on Energy Metabolism.
- $\alpha = 3/4$ made official ... ... 29 to zip.

- But the Cabal slipped up by publishing the conference proceedings ...
- "Energy Metabolism; Proceedings of the 3rd symposium held at Troon, Scotland, May 1964," Ed. Sir Kenneth Blaxter [13]

Somehow, optimal river networks are connected:



- a = drainage basin area
- ℓ = length of longest (main) stream
- $L = L_\parallel$ = longitudinal length of basin

Mysterious allometric scaling in river networks

- 1957: J. T. Hack [50]
  "Studies of Longitudinal Stream Profiles in Virginia and Maryland"

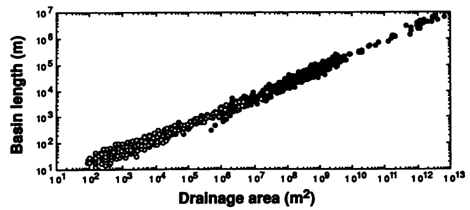
$$\ell \sim a^h$$

$$h \sim 0.6$$

- Anomalous scaling: we would expect $h = 1/2$ ...
- Subsequent studies: $0.5 \lesssim h \lesssim 0.6$
- Another quest to find universality/god ...
- A catch: studies done on small scales.

Large-scale networks:

(1992) Montgomery and Dietrich [69]:

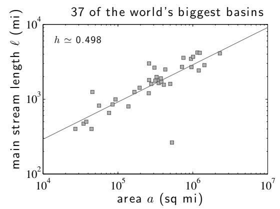


- Composite data set: includes everything from unchanneled valleys up to world's largest rivers.
- Estimated fit:

$$L \simeq 1.78 a^{0.49}$$

- Mixture of basin and main stream lengths.

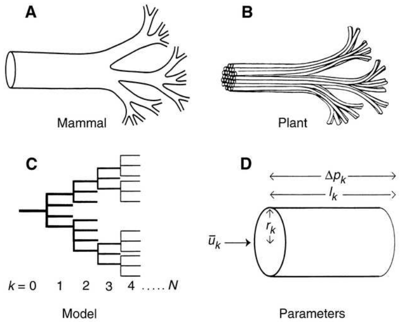
World's largest rivers only:



- Data from Leopold (1994) [60, 32]
- Estimate of Hack exponent: $h = 0.50 \pm 0.06$

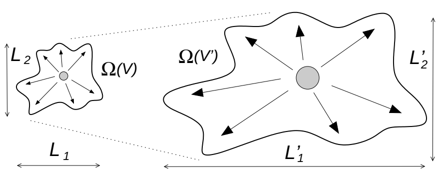
Nutrient delivering networks:

- 1960's: Rashevsky considers blood networks and finds a $2/3$ scaling.
- 1997: West *et al.* [113] use a network story to find $3/4$ scaling.



Geometric argument

- Allometrically growing regions:

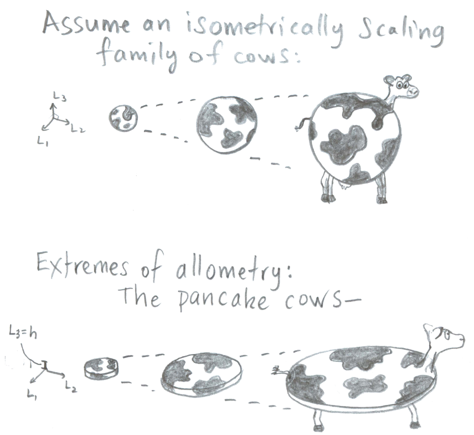


- Have d length scales which scale as

$$L_i \propto V^{\gamma_i} \text{ where } \gamma_1 + \gamma_2 + \ldots + \gamma_d = 1.$$

- For isometric growth, $\gamma_i = 1/d$.
- For allometric growth, we must have at least two of the $\{\gamma_i\}$ being different

Spherical cows and pancake cows:



Minimal network volume:

Real supply networks are close to optimal:

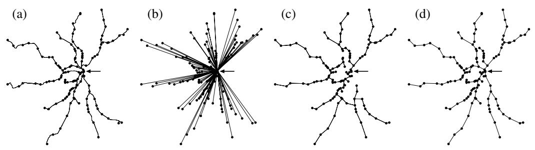
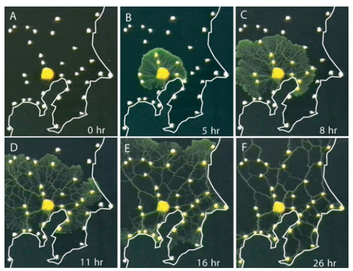


Figure 1. (a) Commuter rail network in the Boston area. The arrow marks the assumed root of the network. (b) Star graph. (c) Minimum spanning tree. (d) The model of equation (3) applied to the same set of stations.

Gastner and Newman (2006): "Shape and efficiency in spatial distribution networks" [41]

"Rules for Biologically Inspired Adaptive Network Design"
Tero et al.,
Science, **327**, 439-442, 2010. [98]



Urban deslime in action:
https://www.youtube.com/watch?v=GwKuFREOgmo

Blood networks

- Then P, the rate of overall energy use in Ω, can at most scale with volume as

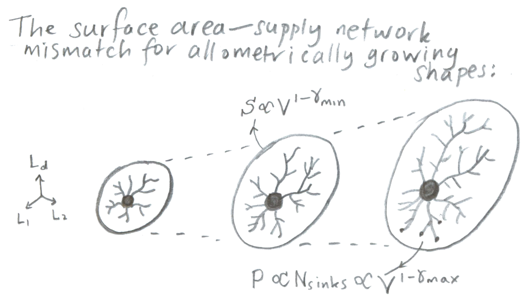
$$P \propto \rho V \propto \rho M \propto M^{(d-1)/d}$$

- For $d = 3$ dimensional organisms, we have

$$\boxed{P \propto M^{2/3}}$$

- Including other constraints may raise scaling exponent to a higher, less efficient value.

- Exciting bonus: Scaling obtained by the supply network story and the surface-area law only match for isometrically growing shapes.



Hack's law

- Volume of water in river network can be calculated by adding up basin areas
- Flows sum in such a way that

$$V_{\text{net}} = \sum_{\text{all pixels}} a_{\text{pixel } i}$$

- Hack's law again:

$$\ell \sim a^h$$

- Can argue

$$V_{\text{net}} \propto V_{\text{basin}}^{1+h} = a_{\text{basin}}^{1+h}$$

where h is Hack's exponent.
- $\therefore$ minimal volume calculations gives

$$\boxed{h = 1/2}$$

Real data:

- Banavar et al.'s approach [7] is okay because ρ really is constant.
- The irony: shows optimal basins are isometric
- Optimal Hack's law: $\ell \sim a^h$ with $h = 1/2$
- (Zzzzz)

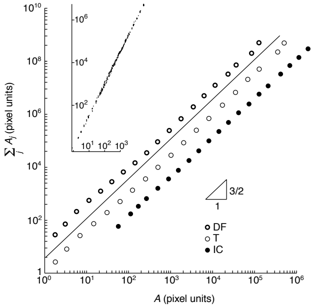
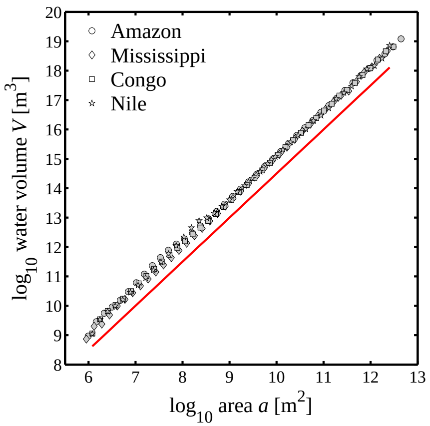
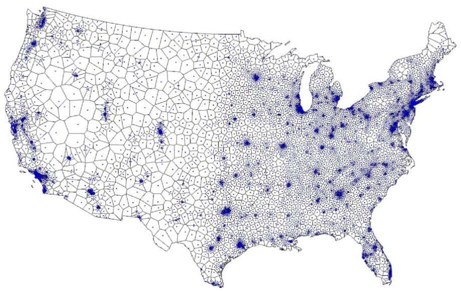


Figure 2 Allometric scaling in river networks. Double logarithmic plot of $C \times \Sigma_{i \in c} A_i$ versus A for three river networks characterized by different climates, geology and geographic locations (Dry Fork, West Virginia, 586 km², digital terrain map (DTM) size 30 × 30 m²; Island Creek, Idaho, 260 km², DTM size 30 × 30 m²; Tirso, Italy, 2,024 km², DTM size 237 × 237 m²). The experimental points are obtained by binning total contributing areas, and computing the ensemble average of the sum of the inner areas for each sub-basin within the binned interval. The figure uses pixel units in which the smallest area element is assigned a unit value. Also plotted is the predicted scaling relationship with slope 3/2. The inset shows the raw data from the Tirso basin before any binning has been done.

Even better—prefactors match up:

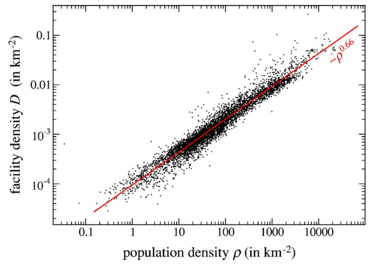


"Optimal design of spatial distribution networks"
Gastner and Newman,
Phys. Rev. E, **74**, 016117, 2006. [40]



- Approximately optimal location of 5000 facilities.
- Based on 2000 Census data.
- Simulated annealing + Voronoi tessellation.

Optimal source allocation



- Optimal facility density ρ_{fac} vs. population density ρ_{pop}.
- Fit is $\rho_{\text{fac}} \propto \rho_{\text{pop}}^{0.66}$ with $r^2 = 0.94$.
- Looking good for a 2/3 power ...

Deriving the optimal source distribution:

- Basic idea: Minimize the average distance from a random individual to the nearest facility. [40]
- Assume given a fixed population density ρ_{pop} defined on a spatial region Ω.
- Formally, we want to find the locations of n sources $\{\vec{x}_1, \ldots, \vec{x}_n\}$ that minimizes the cost function

$$F(\{\vec{x}_1, \ldots, \vec{x}_n\}) = \int_{\Omega} \rho_{\text{pop}}(\vec{x}) \min_i ||\vec{x} - \vec{x}_i|| \mathrm{d}\vec{x}\,.$$

- Also known as the p-median problem, and connected to cluster analysis.
- Not easy ...in fact this one is an NP-hard problem. [40]
- Approximate solution originally due to Gusein-Zade [49].

Global redistribution networks

One more thing:

- How do we supply these facilities?
- How do we best redistribute mail? People?
- How do we get beer to the pubs?
- Gastner and Newman model: cost is a function of basic maintenance and travel time:

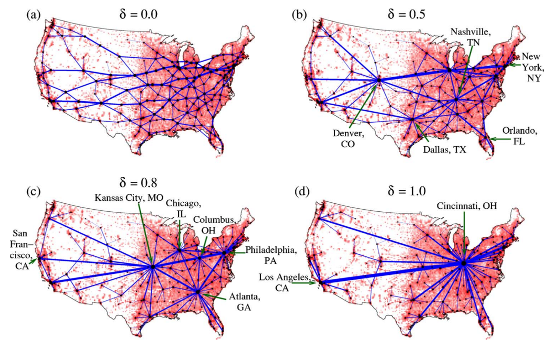
$$C_{\text{maint}} + \gamma C_{\text{travel}}.$$

- Travel time is more complicated: Take 'distance' between nodes to be a composite of shortest path distance ℓ_{ij} and number of legs to journey:

$$(1 - \delta)\ell_{ij} + \delta(\#\text{hops}).$$

- When $\delta = 1$, only number of hops matters.

Global redistribution networks



From Gastner and Newman (2006) [40]

Public versus private facilities

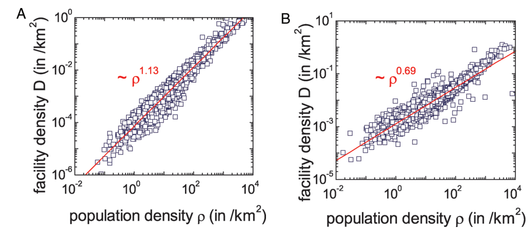
Beyond minimizing distances:

- "Scaling laws between population and facility densities" by Um *et al.*, Proc. Natl. Acad. Sci., 2009. [104]
- Um *et al.* find empirically and argue theoretically that the connection between facility and population density

$$\rho_{\text{fac}} \propto \rho_{\text{pop}}^{\alpha}$$

does not universally hold with $\alpha = 2/3$.
- Two idealized limiting classes:
  1. For-profit, commercial facilities: $\alpha = 1$;
  2. Pro-social, public facilities: $\alpha = 2/3$.
- Um *et al.* investigate facility locations in the United States and South Korea.

Public versus private facilities: evidence



- Left plot: ambulatory hospitals in the U.S.
- Right plot: public schools in the U.S.
- Note: break in scaling for public schools. Transition from $\alpha \simeq 2/3$ to $\alpha = 1$ around $\rho_{\text{pop}} \simeq 100$.

Public versus private facilities: evidence

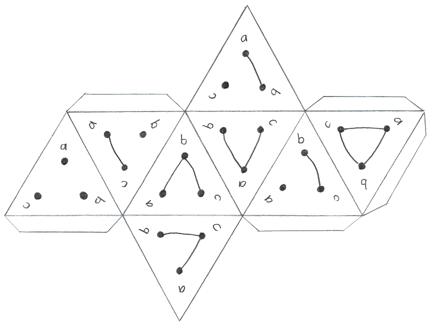
US facility	α (SE)	R^2
Ambulatory hospital	1.13(1)	0.93
Beauty care	1.08(1)	0.86
Laundry	1.05(1)	0.90
Automotive repair	0.99(1)	0.92
Private school	0.95(1)	0.82
Restaurant	0.93(1)	0.89
Accommodation	0.89(1)	0.70
Bank	0.88(1)	0.89
Gas station	0.86(1)	0.94
Death care	0.79(1)	0.80
* Fire station	0.78(3)	0.93
* Police station	0.71(6)	0.75
Public school	0.69(1)	0.87

SK facility	α (SE)	R^2
Bank	1.18(2)	0.96
Parking place	1.13(2)	0.91
* Primary clinic	1.09(2)	1.00
* Hospital	0.96(5)	0.97
* University/college	0.93(9)	0.89
Market place	0.87(2)	0.90
* Secondary school	0.77(3)	0.98
* Primary school	0.77(3)	0.97
Social welfare org.	0.75(2)	0.84
* Police station	0.71(5)	0.94
Government office	0.70(1)	0.93
* Fire station	0.60(4)	0.93
* Public health center	0.09(5)	0.19

Rough transition between public and private at $\alpha \simeq 0.8$.

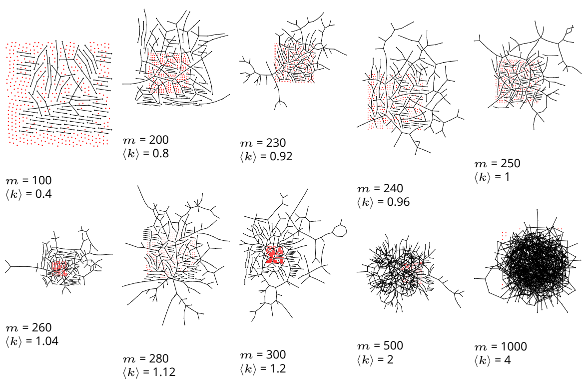
Note: * indicates analysis is at state/province level; otherwise county level.

Random network generator for $N = 3$:

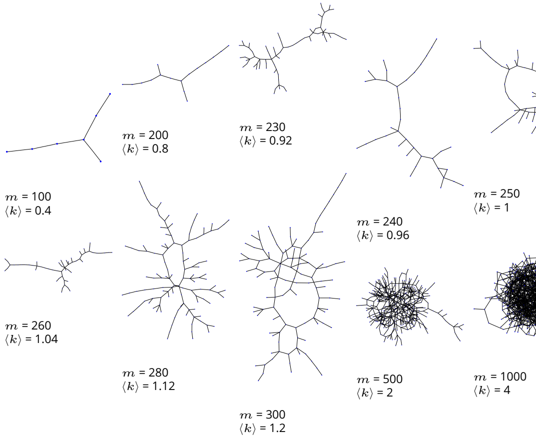


- Get your own exciting generator here.
- As $N \nearrow$, polyhedral die rapidly becomes a ball ...

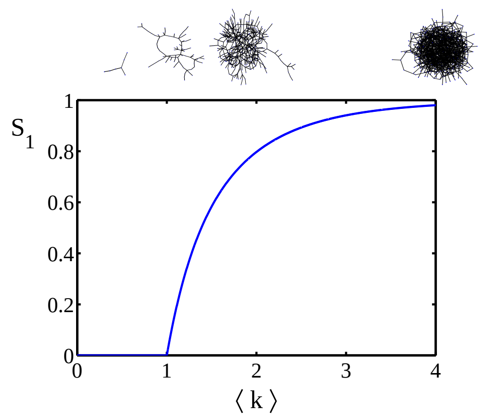
Random networks: examples for N=500



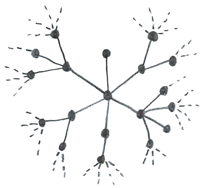
Random networks: largest components



Giant component



Clustering in random networks:



- So for large random networks ($N \to \infty$), clustering drops to zero.
- Key structural feature of random networks is that they locally look like pure branching networks
- No small loops.

Degree distribution:

- Recall P_k = probability that a randomly selected node has degree k.
- Consider method 1 for constructing random networks: each possible link is realized with probability p.
- Now consider one node: there are '$N-1$ choose k' ways the node can be connected to k of the other $N-1$ nodes.
- Each connection occurs with probability p, each non-connection with probability $(1-p)$.
- Therefore have a binomial distribution:

$$P(k; p, N) = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

Limiting form of $P(k; p, N)$:

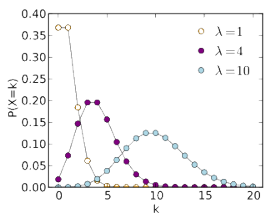
- Our degree distribution: $P(k; p, N) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$.
- What happens as $N \to \infty$?
- We must end up with the normal distribution right?
- If p is fixed, then we would end up with a Gaussian with average degree $\langle k \rangle \simeq pN \to \infty$.
- But we want to keep $\langle k \rangle$ fixed ...
- So examine limit of $P(k; p, N)$ when $p \to 0$ and $N \to \infty$ with $\langle k \rangle = p(N-1)$ = constant.

$$P(k; p, N) \simeq \frac{\langle k \rangle^k}{k!} \left(1 - \frac{\langle k \rangle}{N-1}\right)^{N-1-k} \to \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

- This is a Poisson distribution with mean $\langle k \rangle$.

Poisson basics:

$$\boxed{P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}}$$



- $\lambda > 0$
- $k = 0, 1, 2, 3, \ldots$
- Classic use: probability that an event occurs k times in a given time period, given an average rate of occurrence.
- e.g.: phone calls/minute, horse-kick deaths.
- 'Law of small numbers'

Models

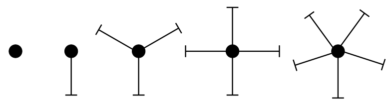
Generalized random networks:

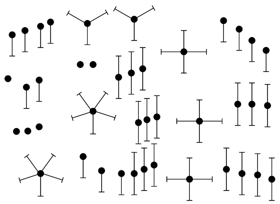
- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Building random networks: Stubs

Phase 1:

- Idea: start with a soup of unconnected nodes with stubs (half-edges):



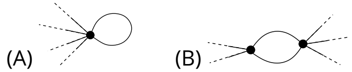


- Randomly select stubs (not nodes!) and connect them.
- Must have an even number of stubs.
- Initially allow self- and repeat connections.

Building random networks: First rewiring

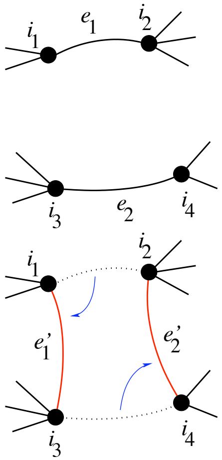
Phase 2:

- Now find any (A) self-loops and (B) repeat edges and randomly rewire them.



- Being careful: we can't change the degree of any node, so we can't simply move links around.
- Simplest solution: randomly rewire two edges at a time.

General random rewiring algorithm



- Randomly choose two edges. (Or choose problem edge and a random edge)
- Check to make sure edges are disjoint.
- Rewire one end of each edge.
- Node degrees do not change.
- Works if e_1 is a self-loop or repeated edge.
- Same as finding on/off/on/off 4-cycles. and rotating them.

Sampling random networks

Phase 2:

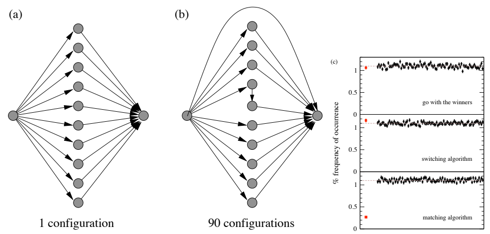
- Use rewiring algorithm to remove all self and repeat loops.

Phase 3:

- Randomize network wiring by applying rewiring algorithm liberally.
- Rule of thumb: # Rewirings $\simeq 10 \times$ # edges [68].

Random sampling

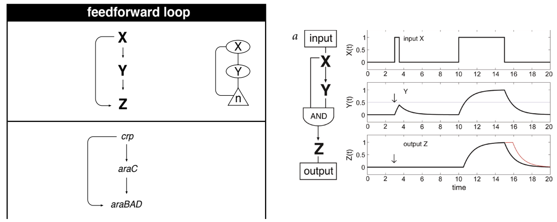
- Problem with only joining up stubs is failure to randomly sample from all possible networks.
- Example from Milo et al. (2003) [68]:



Network motifs

- Idea of motifs [89] introduced by Shen-Orr, Alon et al. in 2002.
- Looked at gene expression within full context of transcriptional regulation networks.
- Specific example of Escherichia coli.
- Directed network with 577 interactions (edges) and 424 operons (nodes).
- Used network randomization to produce ensemble of alternate networks with same degree frequency N_k.
- Looked for certain subnetworks (motifs) that appeared more or less often than expected

Network motifs



- Z only turns on in response to sustained activity in X.
- Turning off X rapidly turns off Z.
- Analogy to elevator doors.

The edge-degree distribution:

- The degree distribution P_k is fundamental for our description of many complex networks
- Again: P_k is the degree of randomly chosen node.
- A second very important distribution arises from choosing randomly on edges rather than on nodes.
- Define Q_k to be the probability the node at a random end of a randomly chosen edge has degree k.
- Now choosing nodes based on their degree (i.e., size):

$$\boxed{Q_k \propto k P_k}$$

- Normalized form:

$$Q_k = \frac{k P_k}{\sum_{k'=0}^{\infty} k' P_{k'}} = \frac{k P_k}{\langle k \rangle}.$$

- Big deal: Rich-get-richer mechanism is built into this selection process.

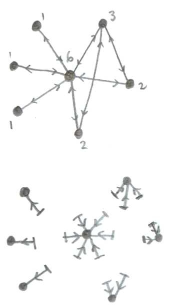
The edge-degree distribution:

- For networks, Q_k is also the probability that a friend (neighbor) of a random node has k friends.
- Useful variant on Q_k:

  R_k = probability that a friend of a random node has k other friends.
-

$$R_k = \frac{(k+1)P_{k+1}}{\sum_{k'=0}(k'+1)P_{k'+1}} = \frac{(k+1)P_{k+1}}{\langle k \rangle}$$

- Equivalent to friend having degree $k+1$.
- Natural question: what's the expected number of other friends that one friend has?



- Probability of randomly selecting a node of degree k by choosing from nodes: $P_1 = 3/7$, $P_2 = 2/7$, $P_3 = 1/7$, $P_6 = 1/7$.
- Probability of landing on a node of degree k after randomly selecting an edge and then randomly choosing one direction to travel: $Q_1 = 3/16$, $Q_2 = 4/16$, $Q_3 = 3/16$, $Q_6 = 6/16$.
- Probability of finding # outgoing edges = k after randomly selecting an edge and then randomly choosing one direction to travel: $R_0 = 3/16$ $R_1 = 4/16$, $R_2 = 3/16$, $R_5 = 6/16$.

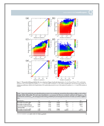
Two reasons why this matters

Reason #1:

- Average # friends of friends per node is

$$\langle k_2 \rangle = \langle k \rangle \times \langle k \rangle_R = \langle k \rangle \frac{1}{\langle k \rangle}\left(\langle k^2 \rangle - \langle k \rangle\right) = \langle k^2 \rangle - \langle k \rangle.$$

- Key: Average depends on the 1st and 2nd moments of P_k and not just the 1st moment.
- Three peculiarities:
  1. We might guess $\langle k_2 \rangle = \langle k \rangle(\langle k \rangle - 1)$ but it's actually $\langle k(k-1) \rangle$.
  2. If P_k has a large second moment, then $\langle k_2 \rangle$ will be big. (e.g., in the case of a power-law distribution)
  3. Your friends really are different from you ... [37, 76]
  4. See also: class size paradoxes (nod to: Gelman)



"Generalized friendship paradox in complex networks: The case of scientific collaboration"
Eom and Jo,
Nature Scientific Reports, **4**, 4603, 2014. [35]

Your friends really are ~~monsters~~ #winners:[1]

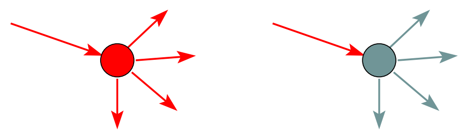
- Go on, hurt me: Friends have more coauthors, citations, and publications.
- Other horrific studies: your connections on Twitter have more followers than you, are happier than you [17], more sexual partners than you, ...
- The hope: Maybe they have more enemies and diseases too.
- Research possibility: The Frenemy Paradox.

[1] Some press here [MIT Tech Review].

Spreading on Random Networks

- For random networks, we know local structure is pure branching.
- Successful spreading is ∴ contingent on single edges infecting nodes.

Success Failure:



- Focus on binary case with edges and nodes either infected or not.
- First big question: for a given network and contagion process, can global spreading from a single seed occur?

Global spreading condition

- We need to find: [30]
  R = the average # of infected edges that one random infected edge brings about.
- Call **R** the gain ratio.
- Define B_{k1} as the probability that a node of degree k is infected by a single infected edge.
-

$$\mathbf{R} = \sum_{k=0}^{\infty} \underbrace{\frac{kP_k}{\langle k \rangle}}_{\substack{\text{prob. of}\\ \text{connecting to}\\ \text{a degree } k \text{ node}}} \bullet \underbrace{(k-1)}_{\substack{\text{\# outgoing}\\ \text{infected}\\ \text{edges}}} \bullet \underbrace{B_{k1}}_{\substack{\text{Prob. of}\\ \text{infection}}}$$

$$+ \sum_{k=0}^{\infty} \overbrace{\frac{kP_k}{\langle k \rangle}} \bullet \underbrace{0}_{\substack{\text{\# outgoing}\\ \text{infected}\\ \text{edges}}} \bullet \underbrace{(1-B_{k1})}_{\substack{\text{Prob. of}\\ \text{no infection}}}$$

Global spreading condition

- Our global spreading condition is then:

$$\boxed{\mathbf{R} = \sum_{k=0}^{\infty} \frac{kP_k}{\langle k \rangle} \bullet (k-1) \bullet B_{k1} > 1.}$$

- Case 1–Rampant spreading: If $B_{k1} = 1$ then

$$\mathbf{R} = \sum_{k=0}^{\infty} \frac{kP_k}{\langle k \rangle} \bullet (k-1) = \frac{\langle k(k-1) \rangle}{\langle k \rangle} > 1.$$

- Good: This is just our giant component condition again.

Global spreading condition

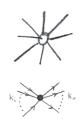
- Case 2—Simple disease-like: If $B_{k1} = \beta < 1$ then

$$\mathbf{R} = \sum_{k=0}^{\infty} \frac{kP_k}{\langle k \rangle} \bullet (k-1) \bullet \beta > 1.$$

- A fraction (1-β) of edges do not transmit infection.
- Analogous phase transition to giant component case but critical value of $\langle k \rangle$ is increased.
- Aka bond percolation.
- Resulting degree distribution $\tilde{P}_k$:

$$\tilde{P}_k = \beta^k \sum_{i=k}^{\infty} \binom{i}{k} (1-\beta)^{i-k} P_i.$$

Random directed networks:



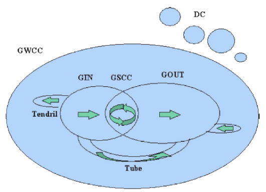
- So far, we've largely studied networks with undirected, unweighted edges.
- Now consider directed, unweighted edges.
- Nodes have k_i and k_o incoming and outgoing edges, otherwise random.
- Network defined by joint in- and out-degree distribution: $P_{k_\mathrm{i},k_\mathrm{o}}$
- Normalization: $\sum_{k_\mathrm{i}=0}^{\infty}\sum_{k_\mathrm{o}=0}^{\infty} P_{k_\mathrm{i},k_\mathrm{o}} = 1$
- Marginal in-degree and out-degree distributions:

$$P_{k_\mathrm{i}} = \sum_{k_\mathrm{o}=0}^{\infty} P_{k_\mathrm{i},k_\mathrm{o}} \text{ and } P_{k_\mathrm{o}} = \sum_{k_\mathrm{i}=0}^{\infty} P_{k_\mathrm{i},k_\mathrm{o}}$$

- Required balance:

$$\langle k_\mathrm{i} \rangle = \sum_{k_\mathrm{i}=0}^{\infty}\sum_{k_\mathrm{o}=0}^{\infty} k_\mathrm{i} P_{k_\mathrm{i},k_\mathrm{o}} = \sum_{k_\mathrm{i}=0}^{\infty}\sum_{k_\mathrm{o}=0}^{\infty} k_\mathrm{o} P_{k_\mathrm{i},k_\mathrm{o}} = \langle k_\mathrm{o} \rangle$$

Directed network structure:



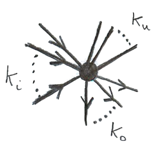
From Boguñá and Serano.[15]

- GWCC = Giant Weakly Connected Component (directions removed);
- GIN = Giant In-Component;
- GOUT = Giant Out-Component;
- GSCC = Giant Strongly Connected Component;
- DC = Disconnected Components (finite).

- When moving through a family of increasingly connected directed random networks, GWCC usually appears before GIN, GOUT, and GSCC which tend to appear together.[80, 15]

Observation:

- Directed and undirected random networks are separate families ...
- ...and analyses are also disjoint.
- Need to examine a larger family of random networks with mixed directed and undirected edges.



- Consider nodes with three types of edges:
  1. k_u undirected edges,
  2. k_i incoming directed edges,
  3. k_o outgoing directed edges.
- Define a node by generalized degree:

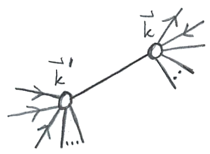
$$\vec{k} = [\, k_u \;\; k_i \;\; k_o \,]^{\mathsf{T}}.$$

Correlations:

- Now add correlations (two point or Markovian) ☐:
  1. $P^{(u)}(\vec{k}\,|\,\vec{k}')$ = probability that an undirected edge leaving a degree $\vec{k}'$ nodes arrives at a degree $\vec{k}$ node.
  2. $P^{(i)}(\vec{k}\,|\,\vec{k}')$ = probability that an edge leaving a degree $\vec{k}'$ nodes arrives at a degree $\vec{k}$ node is an in-directed edge relative to the destination node.
  3. $P^{(o)}(\vec{k}\,|\,\vec{k}')$ = probability that an edge leaving a degree $\vec{k}'$ nodes arrives at a degree $\vec{k}$ node is an out-directed edge relative to the destination node.
- Now require more refined (detailed) balance.
- Conditional probabilities cannot be arbitrary.
  1. $P^{(u)}(\vec{k}\,|\,\vec{k}')$ must be related to $P^{(u)}(\vec{k}'\,|\,\vec{k})$.
  2. $P^{(o)}(\vec{k}\,|\,\vec{k}')$ and $P^{(i)}(\vec{k}\,|\,\vec{k}')$ must be connected.

Correlations—Undirected edge balance:

- Randomly choose an edge, and randomly choose one end.
- Say we find a degree $\vec{k}$ node at this end, and a degree $\vec{k}'$ node at the other end.
- Define probability this happens as $P^{(u)}(\vec{k},\vec{k}')$.
- Observe we must have $P^{(u)}(\vec{k},\vec{k}') = P^{(u)}(\vec{k}',\vec{k})$.



- Conditional probability connection:

$$P^{(u)}(\vec{k},\vec{k}') = P^{(u)}(\vec{k}\,|\,\vec{k}')\frac{k_u' P(\vec{k}')}{\langle k_u \rangle}$$

$$P^{(u)}(\vec{k}',\vec{k}) = P^{(u)}(\vec{k}'\,|\,\vec{k})\frac{k_u P(\vec{k})}{\langle k_u \rangle}.$$

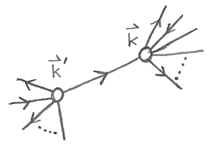
Correlations—Directed edge balance:

- The quantities

$$\frac{k_o P(\vec{k})}{\langle k_o \rangle} \text{ and } \frac{k_i P(\vec{k})}{\langle k_i \rangle}$$

give the probabilities that in starting at a random end of a randomly selected edge, we begin at a degree $\vec{k}$ node and then find ourselves travelling:

1. along an outgoing edge, or
2. against the direction of an incoming edge.



- We therefore have

$$P^{(\mathrm{dir})}(\vec{k},\vec{k}') = P^{(i)}(\vec{k}\,|\,\vec{k}')\frac{k_o' P(\vec{k}')}{\langle k_o \rangle} = P^{(o)}(\vec{k}'\,|\,\vec{k})\frac{k_i P(\vec{k})}{\langle k_i \rangle}.$$

- Note that $P^{(\mathrm{dir})}(\vec{k},\vec{k}')$ and $P^{(\mathrm{dir})}(\vec{k}',\vec{k})$ are in general not related if $\vec{k} \neq \vec{k}'$.

Summary of contagion conditions for uncorrelated networks:

- I. Undirected, Uncorrelated—$f(d+1) = \mathbf{f}(d)$:

$$\mathbf{R} = \sum_{k_u} P^{(u)}(k_u\,|\,*) \bullet (k_u - 1) \bullet B_{k_u,*}$$

- II. Directed, Uncorrelated—$f(d+1) = \mathbf{f}(d)$:

$$\mathbf{R} = \sum_{k_i,k_o} P^{(i)}(k_i,k_o\,|\,*) \bullet k_o \bullet B_{k_i,*}$$

- III. Mixed Directed and Undirected, Uncorrelated—

$$\begin{bmatrix} f^{(u)}(d+1) \\ f^{(o)}(d+1) \end{bmatrix} = \mathbf{R}\begin{bmatrix} f^{(u)}(d) \\ f^{(o)}(d) \end{bmatrix}$$

$$\mathbf{R} = \sum_{\vec{k}} \begin{bmatrix} P^{(u)}(\vec{k}\,|\,*) \bullet (k_u - 1) & P^{(i)}(\vec{k}\,|\,*) \bullet k_u \\ P^{(u)}(\vec{k}\,|\,*) \bullet k_o & P^{(i)}(\vec{k}\,|\,*) \bullet k_o \end{bmatrix} \bullet B_{k_u k_i,*}$$

Summary of contagion conditions for correlated networks:

- IV. Undirected, Correlated—$f_{k_u}(d+1) = \sum_{k_u'} R_{k_u k_u'} f_{k_u'}(d)$

$$R_{k_u k_u'} = P^{(u)}(k_u\,|\,k_u') \bullet (k_u - 1) \bullet B_{k_u k_u'}$$

- V. Directed, Correlated—$f_{k_i k_o}(d+1) = \sum_{k_i',k_o'} R_{k_i k_o k_i' k_o'} f_{k_i' k_o'}(d)$

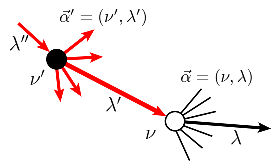
$$R_{k_i k_o k_i' k_o'} = P^{(i)}(k_i, k_o\,|\,k_i', k_o') \bullet k_o \bullet B_{k_i k_o k_i' k_o'}$$

- VI. Mixed Directed and Undirected, Correlated—

$$\begin{bmatrix} f^{(u)}_{\vec{k}}(d+1) \\ f^{(o)}_{\vec{k}}(d+1) \end{bmatrix} = \sum_{\vec{k}'} \mathbf{R}_{\vec{k}\vec{k}'} \begin{bmatrix} f^{(u)}_{\vec{k}'}(d) \\ f^{(o)}_{\vec{k}'}(d) \end{bmatrix}$$

$$\mathbf{R}_{\vec{k}\vec{k}'} = \begin{bmatrix} P^{(u)}(\vec{k}\,|\,\vec{k}') \bullet (k_u - 1) & P^{(i)}(\vec{k}\,|\,\vec{k}') \bullet k_u \\ P^{(u)}(\vec{k}\,|\,\vec{k}') \bullet k_o & P^{(i)}(\vec{k}\,|\,\vec{k}') \bullet k_o \end{bmatrix} \bullet B_{\vec{k}\vec{k}'}$$

Full generalization:



$$f_{\vec{\alpha}}(d+1) = \sum_{\vec{\alpha}'} R_{\vec{\alpha}\vec{\alpha}'} f_{\vec{\alpha}'}(d)$$

$R_{\vec{\alpha}\vec{\alpha}'}$ is the gain ratio matrix and has the form:

$$R_{\vec{\alpha}\vec{\alpha}'} = P_{\vec{\alpha}\vec{\alpha}'} \bullet k_{\vec{\alpha}\vec{\alpha}'} \bullet B_{\vec{\alpha}\vec{\alpha}'}.$$

- $P_{\vec{\alpha}\vec{\alpha}'}$ = conditional probability that a type λ' edge emanating from a type ν' node leads to a type ν node.
- $k_{\vec{\alpha}\vec{\alpha}'}$ = potential number of newly infected edges of type λ emanating from nodes of type ν.
- $B_{\vec{\alpha}\vec{\alpha}'}$ = probability that a type ν node is eventually infected by a single infected type λ' link arriving from a neighboring node of type ν'.
- Generalized contagion condition:

$$\max|\mu| : \mu \in \sigma(\mathbf{R}) > 1$$

Some claims for social networks:

- Social networks yes, but groups, groups, groups
- Sufficiently large social groups are:
  1. Fandoms.
  2. Pyramid Schemes,
  3. Or both.
- Homo narrativus: Storytellers, believers, spreaders.
- Stories ∼ Characters + Time.
- Characters are shortcuts to stories.

For novel diseases:

1. Can we predict the size of an epidemic?
2. How important is the reproduction number R_0?

R_0 approximately same for all of the following:

- 1918-19 “Spanish Flu” ~ 75,000,000 world-wide, 500,000 deaths in US.
- 1957-58 “Asian Flu” ~ 2,000,000 world-wide, 70,000 deaths in US.
- 1968-69 “Hong Kong Flu” ~ 1,000,000 world-wide, 34,000 deaths in US.
- 2003 “SARS Epidemic” ~ 800 deaths world-wide.

Improving simple models

Idea for social networks: incorporate identity

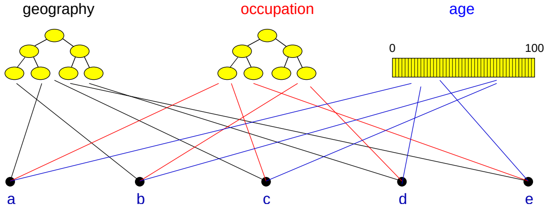
Identity is formed from attributes such as:

- Geographic location
- Type of employment
- Age
- Recreational activities

Groups are crucial ...

- formed by people with at least one similar attribute
- Attributes ⇔ Contexts ⇔ Interactions ⇔ Networks.[110]

Generalized context space



(Blau & Schwartz[12], Simmel[91], Breiger[20])

A toy agent-based model:

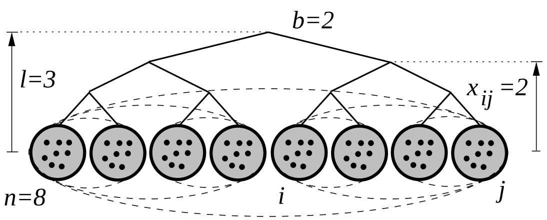
“Multiscale, resurgent epidemics in a hierarchcial metapopulation model”
Watts et al.,
Proc. Natl. Acad. Sci., **102**, 11157–11162, 2005.[111]

Geography: allow people to move between contexts

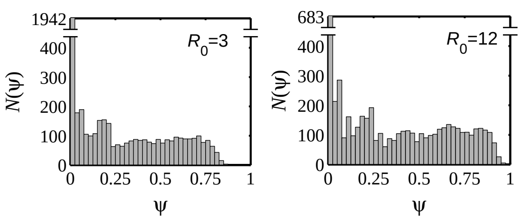
- Locally: standard SIR model with random mixing
- discrete time simulation
- β = infection probability
- γ = recovery probability
- P = probability of travel
- Movement distance: $\Pr(d) \propto \exp(-d/\xi)$
- ξ = typical travel distance

A toy agent-based model

Schematic:

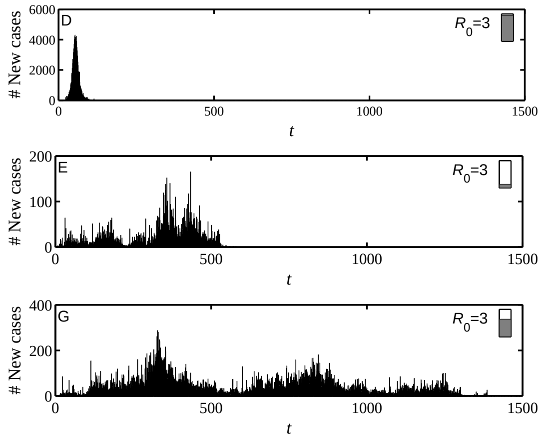


Example model output: size distributions



- Flat distributions are possible for certain ξ and P.
- Different R_0's may produce similar distributions
- Same epidemic sizes may arise from different R_0's

Model output—resurgence

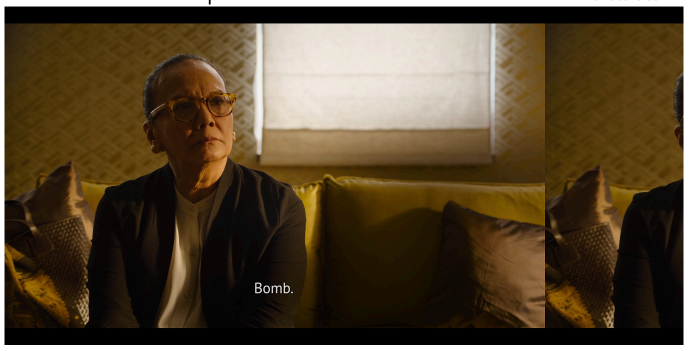


Journal entry, 2020/02/21:

Twitter DMs to Sam Scarpino:

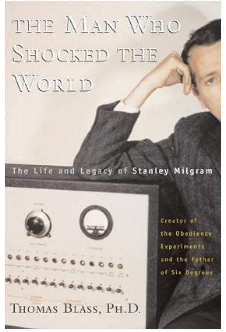
- Okay: The scientists studying pandemics need to be able to present some kind set of numbers that show how bad things are. The whole R_0 disaster has been waiting to happen because people have been ... lazily having fun with math models? Unconcerned about how to communicate vital scientific information? Stupid? I don't know. Maybe a radar plot visualization. I don't know.
- “When these three boundaries are crossed, we are in trouble”
- Measles has an R_0 of 20. We should all have it. Of course, there's no f**king time scale for R_0 so we don't know when that happens.

The Last of Us: Groups.



Understanding distributed social search

Milgram's social search experiment



http://www.stanleymilgram.com

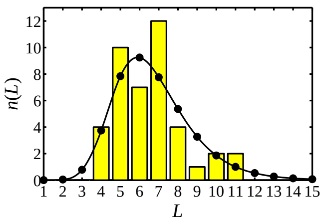
- Target person = Boston stockbroker.
- 296 senders from Boston and Omaha.
- 20% of senders reached target.
- chain length $\simeq 6.5$.

Popular terms:

- The Small World Phenomenon;
- "Six Degrees of Separation."

The model—results

Milgram's Nebraska-Boston data:



Model parameters:

- $N = 10^8$,
- $z = 300$, $g = 100$,
- $b = 10$,
- $\alpha = 1$, $H = 2$;
- $\langle L_{\text{model}} \rangle \simeq 6.7$
- $L_{\text{data}} \simeq 6.5$

Social search—the Columbia experiment

- 60,000+ participants in 166 countries
- 18 targets in 13 countries including
  - a professor at an Ivy League university,
  - an archival inspector in Estonia,
  - a technology consultant in India,
  - a policeman in Australia, and
  - a veterinarian in the Norwegian army.
- 24,000+ chains

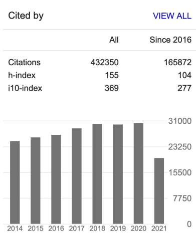
We were lucky and contagious:

"Using E-Mail to Count Connections", Sarah Milstein, New York Times, Circuits Section (December, 2001)

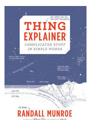
Jonathan Harris's Wordcount:

A word frequency distribution explorer:

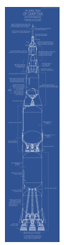
The long tail of knowledge:

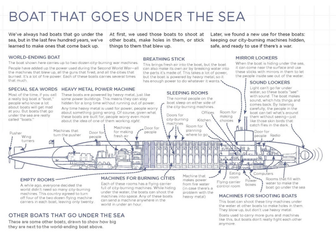


Take a scrolling voyage to the citational abyss, starting at the surface with the lonely, giant citaceans, moving down to the legion of strange, sometimes misplaced, unloved creatures, that dwell in Kahneman's Google Scholar page

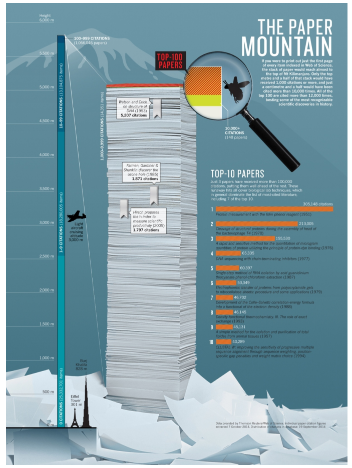


"Thing Explainer: Complicated Stuff in Simple Words" by Randall Munroe (2015).[70]





Up goer five

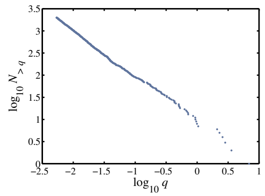


Nature (2014): Most cited papers of all time

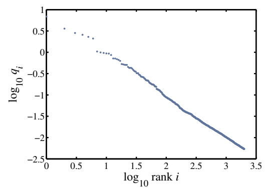
Size distributions:

Brown Corpus (1,015,945 words):

CCDF:



Zipf:



- The, of, and, to, a, ...= 'objects'
- 'Size' = word frequency
- Beep: (Important) CCDF and Zipf plots are related ...

Pre-Zipf's law observations of Zipf's law

- 1910s: Word frequency examined re Stenography (or shorthand and brachygraphy or tachygraphy), Jean-Baptiste Estoup [36].
- 1910s: Felix Auerbach pointed out the Zipfitude of city sizes in "Das Gesetz der Bevölkerungskonzentration" ("The Law of Population Concentration") [5].
- 1924: G. Udny Yule [118]:
  # Species per Genus (offers first theoretical mechanism)
- 1926: Lotka [61]:
  # Scientific papers per author (Lotka's law)

Theoretical Work of Yore:

- 1949: Zipf's "Human Behaviour and the Principle of Least-Effort" is published. [120]
- 1953: Mandelbrot [62]:
  Optimality argument for Zipf's law; focus on language.
- 1955: Herbert Simon [92, 120]:
  Zipf's law for word frequency, city size, income, publications, and species per genus.
- 1965/1976: Derek de Solla Price [26, 83]:
  Network of Scientific Citations.
- 1999: Barabasi and Albert [8]:
  The World Wide Web, networks-at-large.

Essential Extract of a Growth Model:

Random Competitive Replication (RCR):

1. Start with 1 elephant (or element) of a particular flavor at $t = 1$
2. At time $t = 2, 3, 4, \ldots$, add a new elephant in one of two ways:
   - With probability ρ, create a new elephant with a new flavor
     = Mutation/Innovation
   - With probability $1-\rho$, randomly choose from all existing elephants, and make a copy.
     = Replication/Imitation
   - Elephants of the same flavor form a group

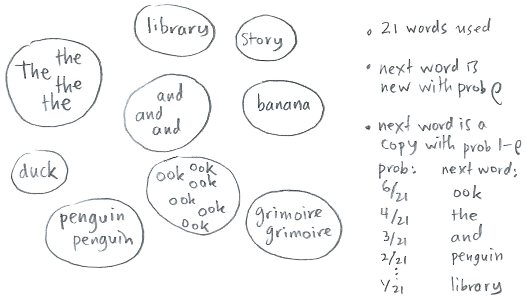
Random Competitive Replication:

Example: Words appearing in a language

- Consider words as they appear sequentially.
- With probability ρ, the next word has not previously appeared
  = Mutation/Innovation
- With probability $1-\rho$, randomly choose one word from all words that have come before, and reuse this word
  = Replication/Imitation

Note: This is a terrible way to write a novel.

For example:



- Micro-to-Macro story with ρ and γ measurable.

$$\boxed{\gamma = \frac{(2-\rho)}{(1-\rho)} = 1 + \frac{1}{(1-\rho)}}$$

- Observe $2 < \gamma < \infty$ for $0 < \rho < 1$.
- For $\rho \simeq 0$ (low innovation rate):

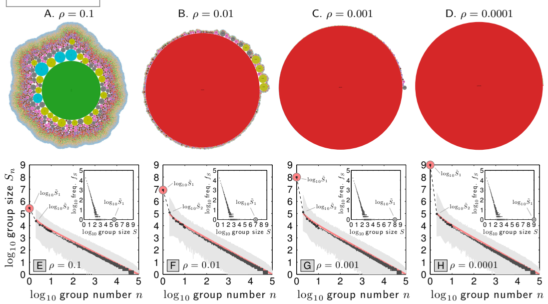
$$\gamma \simeq 2$$

- 'Wild' power-law size distribution of group sizes, bordering on 'infinite' mean.
- For $\rho \simeq 1$ (high innovation rate):

$$\gamma \simeq \infty$$

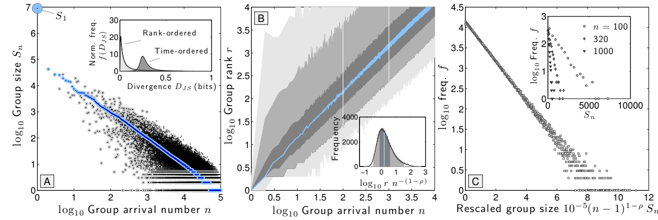
- All elephants have different flavors.
- Upshot: Tunable mechanism producing a family of universality classes.

"Simon's fundamental rich-get-richer model entails a dominant first-mover advantage"
Dodds et al.,
Physical Review E, **95**, 052301, 2017. [29]



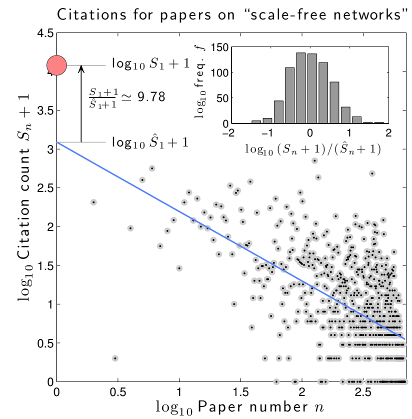
- See visualization at paper's online app-endices

Arrival variability:



- Any one simulation shows a high amount of disorder.
- Two orders of magnitude variation in possible rank.
- Rank ordering creates a smooth Zipf distribution.
- Size distribution for the nth arriving group show exponential decay.

Self-referential citation data:

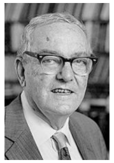


The Quickening—Mandelbrot v. Simon:

There Can Be Only One:

- Things there should be only one of: Theory, Highlander Films.
- Feel free to play Queen's It's a Kind of Magic in your head (funding remains tight).

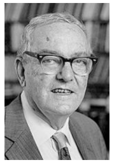
We were born to be Princes of the Universe

 vs. 

Mandelbrot vs. Simon:

- Mandelbrot (1953): “An Informational Theory of the Statistical Structure of Languages” [62]
- Simon (1955): “On a class of skew distribution functions” [92]
- Mandelbrot (1959): “A note on a class of skew distribution functions: analysis and critique of a paper by H.A. Simon” [63]
- Simon (1960): “Some further notes on a class of skew distribution functions” [93]

I have no rival, No man can be my equal

 vs. 

Mandelbrot vs. Simon:

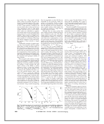
- Mandelbrot (1961): “Final note on a class of skew distribution functions: analysis and critique of a model due to H.A. Simon” [64]
- Simon (1961): “Reply to ‘final note’ by Benoit Mandelbrot” [95]
- Mandelbrot (1961): “Post scriptum to ‘final note”’ [65]
- Simon (1961): “Reply to Dr. Mandelbrot’s post scriptum” [94]

Scale-free networks

- Real networks with power-law degree distributions became known as scale-free networks.
- Scale-free refers specifically to the degree distribution having a power-law decay in its tail:

$$P_k \sim k^{-\gamma} \text{ for ‘large’ } k$$

- One of the seminal works in complex networks:



“Emergence of scaling in random networks”
Barabási and Albert,
Science, **286**, 509–511, 1999. [8]

Times cited: $\sim 43,853$ (as of May 19, 2023)

- Somewhat misleading nomenclature ...

“Organization of Growing Random Networks”
Krapivsky and Redner,
Phys. Rev. E, **63**, 066123, 2001. [57]

Fooling with the mechanism:

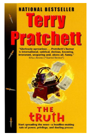
- Krapivsky & Redner [57] explored the general attachment kernel:

$$\textbf{Pr}(\text{attach to node } i) \propto A_k = k_i^{\nu}$$

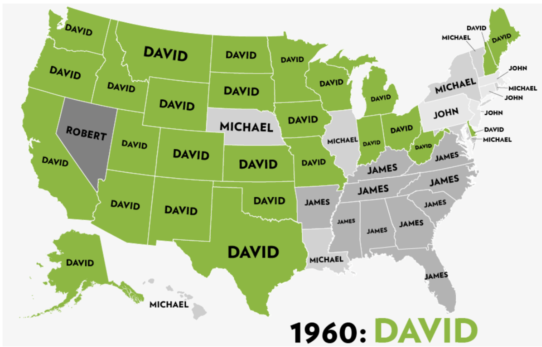
where A_k is the attachment kernel and $\nu > 0$.

- KR also looked at changing the details of the attachment kernel.

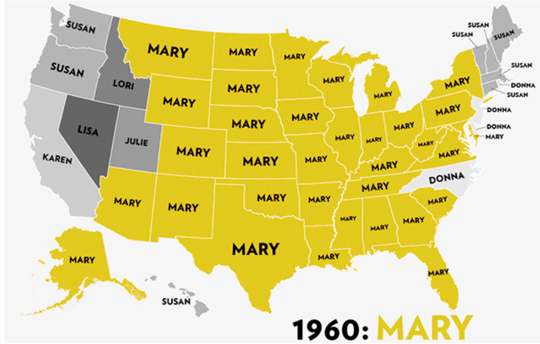
‘The rumor spread through the city like wildfire which had quite often spread through Ankh-Morpork since its citizens had learned the words “fire insurance”).’



“The Truth”
by Terry Pratchett (2000). [82]



From the Atlantic



From the Atlantic

Social Contagion

Some important models:

- Tipping models—Schelling (1971) [85, 86, 87]
  - Simulation on checker boards
  - Idea of thresholds
  - Polygon-themed online visualization. (Includes optional diversity-seeking proclivity.)
- Threshold models—Granovetter (1978) [47]
- Herding models—Bikhchandani, Hirschleifer, Welch (1992) [10, 11]
  - Social learning theory, Informational cascades,...

Social contagion models

Thresholds

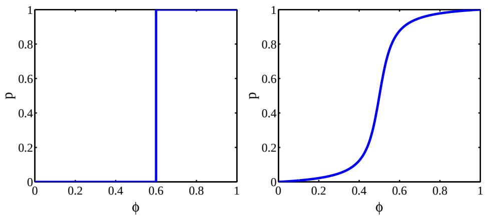
- Basic idea: individuals adopt a behavior when a certain fraction of others have adopted
- ‘Others’ may be everyone in a population, an individual’s close friends, any reference group.
- Response can be probabilistic or deterministic.
- Individual thresholds can vary
- Assumption: order of others’ adoption does not matter... (unrealistic).
- Assumption: level of influence per person is uniform (unrealistic).

Social Contagion

Some possible origins of thresholds:

- Inherent, evolution-devised inclination to coordinate, to conform, to imitate. [9]
- Lack of information: impute the worth of a good or behavior based on degree of adoption (social proof)
- Economics: Network effects or network externalities
  - Externalities = Effects on others not directly involved in a transaction
  - Examples: telephones, fax machine, TikTok, operating systems
  - An individual's utility increases with the adoption level among peers and the population in general

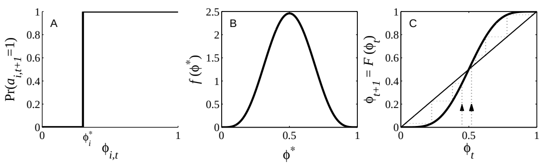
Threshold models—response functions



- Example threshold influence response functions: deterministic and stochastic
- ϕ = fraction of contacts 'on' (e.g., rioting)
- Two states: S and I.

Threshold models

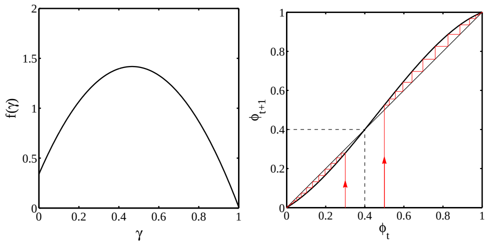
Action based on perceived behavior of others:



- Two states: S and I.
- ϕ = fraction of contacts 'on' (e.g., rioting)
- Discrete time update (strong assumption!)
- This is a Critical mass model

Threshold models

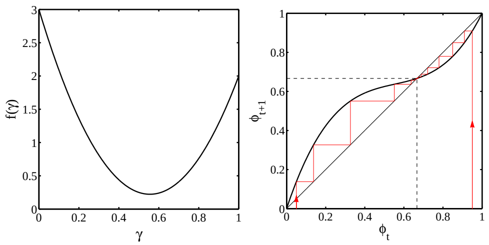
Another example of critical mass model:



- Fragility of fixed point at $\phi = 0$.
- Critical slope = 1.

Threshold models

Example of single stable state model:



Threshold models—Nutshell

Implications for collective action theory:

1. Collective uniformity $\not\Rightarrow$ individual uniformity
2. Small individual changes $\Rightarrow$ large global changes
3. The stories/dynamics of complex systems are conceptually inaccessible for individual-centric narratives.
4. System stories live in left null space of our stories—we can't even see them.
5. But we happily impose simplistic, individual-centric stories—we can't help ourselves.

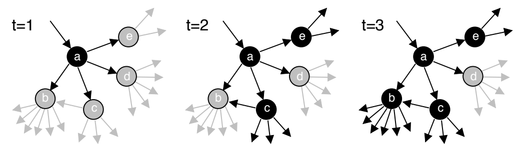
Many years after Granovetter and Soong's work:

- "A simple model of global cascades on random networks"
  D. J. Watts. Proc. Natl. Acad. Sci., 2002 [106]
  - Mean field model $\rightarrow$ network model
  - Individuals now have a limited view of the world

Also consider:

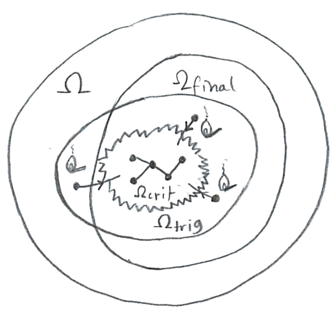
- "Seed size strongly affects cascades on random networks" [44]
  Gleeson and Cahalane, Phys. Rev. E, 2007.
- "Direct, phyiscally motivated derivation of the contagion condition for spreading processes on generalized random networks" [30] Dodds, Harris, and Payne, Phys. Rev. E, 2011
- "Influentials, Networks, and Public Opinion Formation" [108]
  Watts and Dodds, J. Cons. Res., 2007.

Threshold model on a network



- All nodes have threshold $\phi = 0.2$.

Example random network structure:



- Ω_{crit} = Ω_{vuln} = critical mass = global vulnerable component
- Ω_{trig} = triggering component
- Ω_{final} = potential extent of spread
- Ω = entire network

$\Omega_{\text{crit}} \subset \Omega_{\text{trig}}$; $\Omega_{\text{crit}} \subset \Omega_{\text{final}}$; and $\Omega_{\text{trig}}, \Omega_{\text{final}} \subset \Omega$.

Cascade condition

Back to following a link:

- A randomly chosen link, traversed in a random direction, leads to a degree k node with probability $\propto kP_k$.
- Follows from there being k ways to connect to a node with degree k.
- Normalization:

$$\sum_{k=0}^{\infty} kP_k = \langle k \rangle$$

- So

$$P(\text{linked node has degree } k) = \frac{kP_k}{\langle k \rangle}$$

Cascade condition

Next: Vulnerability of linked node

- Linked node is vulnerable with probability

$$\beta_k = \int_{\phi'_*=0}^{1/k} f(\phi'_*)\mathrm{d}\phi'_*$$

- If linked node is vulnerable, it produces $k-1$ new outgoing active links
- If linked node is not vulnerable, it produces no active links.

Cascade condition

Putting things together:

- Expected number of active edges produced by an active edge:

$$R = \left[\underbrace{\sum_{k=1}^{\infty} (k-1)\cdot \beta_k \cdot \frac{kP_k}{\langle k \rangle}}_{\text{success}} + \underbrace{0 \cdot (1-\beta_k)\cdot \frac{kP_k}{\langle k \rangle}}_{\text{failure}} \right]$$

$$= \sum_{k=1}^{\infty} (k-1)\cdot \beta_k \cdot \frac{kP_k}{\langle k \rangle}$$

Cascade condition

So... for random networks with fixed degree distributions, cacades take off when:

$$\sum_{k=1}^{\infty} (k-1)\cdot \beta_k \cdot \frac{kP_k}{\langle k \rangle} > 1.$$

- β_k = probability a degree k node is vulnerable.
- P_k = probability a node has degree k.

Cascade condition

Two special cases:

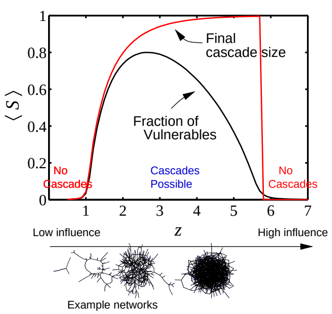
- (1) Simple disease-like spreading succeeds: $\beta_k = \beta$

$$\beta \cdot \sum_{k=1}^{\infty} (k-1)\cdot \frac{kP_k}{\langle k \rangle} > 1.$$

- (2) Giant component exists: $\beta = 1$

$$1 \cdot \sum_{k=1}^{\infty} (k-1)\cdot \frac{kP_k}{\langle k \rangle} > 1.$$

Cascades on random networks

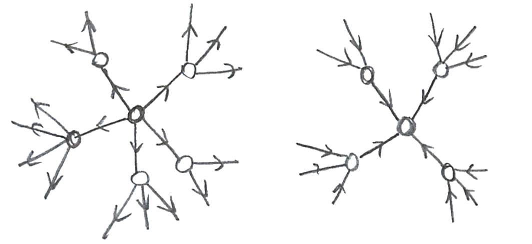


- Cascades occur only if size of max vulnerable cluster > 0.
- System may be 'robust-yet-fragile'.
- 'Ignorance' facilitates spreading.

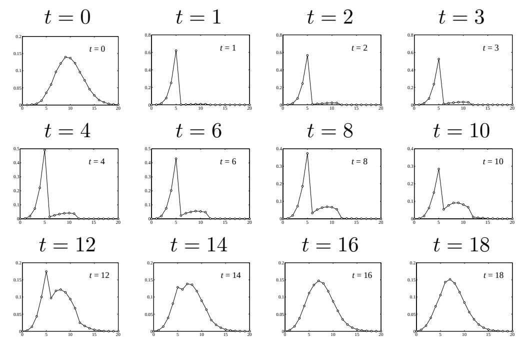
Expected size of spread

Pleasantness:

- Taking off from a single seed story is about expansion away from a node.
- Extent of spreading story is about contraction at a node.

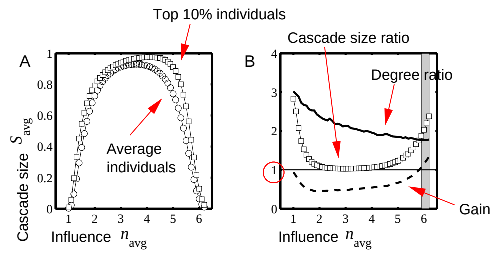


Early adopters—degree distributions



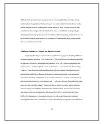
$P_{k,t}$ versus k

The multiplier effect:



- Fairly uniform levels of individual influence.
- Multiplier effect is mostly below 1.

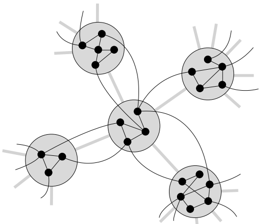
Extensions



"Threshold Models of Social Influence"
Watts and Dodds,
The Oxford Handbook of Analytical Sociology, **63**, 475–497, 2009. [109]

- Assumption of sparse interactions is good
- Degree distribution is (generally) key to a network's function
- Still, random networks don't represent all networks
- Major element missing: group structure

Group structure—Ramified random networks

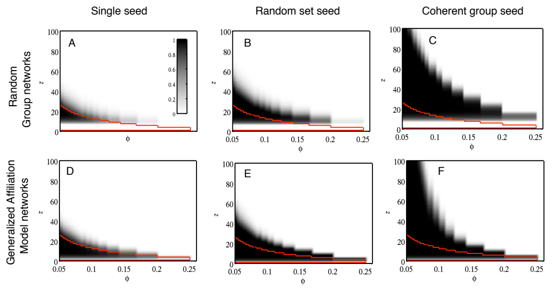


p = intergroup connection probability
q = intragroup connection probability.

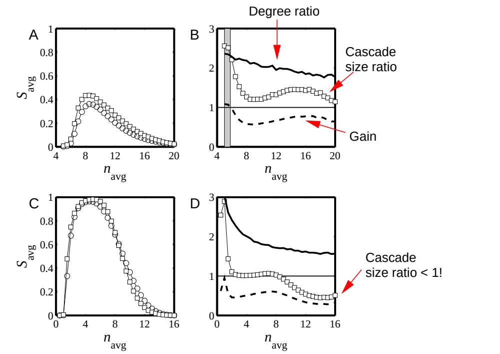
Generalized affiliation model networks with triadic closure

- Connect nodes with probability $\propto e^{-\alpha d}$ where
  α = homophily parameter
  and
  d = distance between nodes (height of lowest common ancestor)
- τ_1 = intergroup probability of friend-of-friend connection
- τ_2 = intragroup probability of friend-of-friend connection

Cascade windows for group-based networks

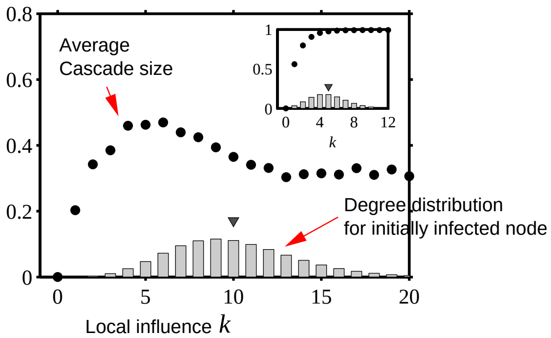


Multiplier effect for group-based networks:



- Multiplier almost always below 1.

Assortativity in group-based networks



- The most connected nodes aren't always the most 'influential.'
- Degree assortativity is the reason.

Social contagion

"Without followers, evil cannot spread." –Leonard Nimoy

Summary

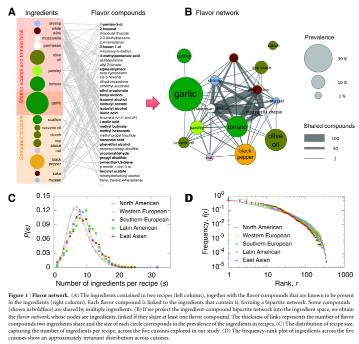
- 'Influential vulnerables' are key to spread.
- Early adopters are mostly vulnerables.
- Vulnerable nodes important but not necessary.
- Groups may greatly facilitate spread.
- Seems that cascade condition is a global one.
- Most extreme/unexpected cascades occur in highly connected networks
- 'Influentials' are posterior constructs.
- Many potential influentials exist.

Social contagion

Implications

- Focus on the influential vulnerables.
- Create entities that can be transmitted successfully through many individuals rather than broadcast from one 'influential.'
- Only simple ideas can spread by word-of-mouth.
  (Idea of opinion leaders spreads well...)
- Want enough individuals who will adopt and display.
- Displaying can be passive = free (yo-yo's, fashion), or active = harder to achieve (political messages; even so: buttons and hats).
- Entities can be novel or designed to combine with others, e.g. block another one.

"Flavor network and the principles of food pairing"
Ahn et al.,
Nature Scientific Reports, **1**, 196, 2011. [1]



“Flavor network and the principles of food pairing”
Ahn et al.,
Nature Scientific Reports, **1**, 196, 2011. [1]

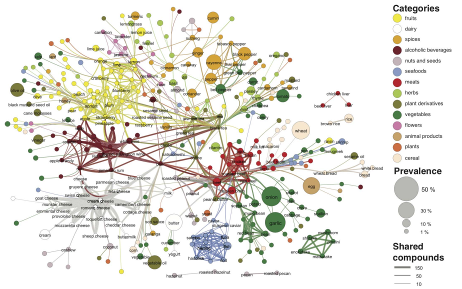


Figure 2 | The backbone of the flavor network. Each node denotes an ingredient, the node color indicates food category, and node size reflects the ingredient prevalence in recipes. Two ingredients are connected if they share a significant number of flavor compounds, link thickness representing the number of shared compounds between the two ingredients. Adjacent links are bundled to reduce the clutter. Note that the map shows only the statistically significant links, as identified by the algorithm of Refs.[28,29] for *p*-value 0.04. A drawing of the full network is too dense to be informative. We use, however, the full network in our subsequent measurements.

“Recipe recommendation using ingredient networks”
Teng, Lin, and Adamic,
Proceedings of the 3rd Annual ACM Web Science Conference, **1**, 298–307, 2012. [97]

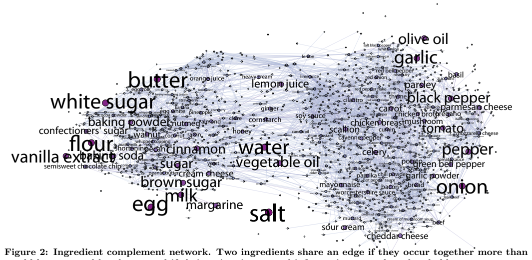
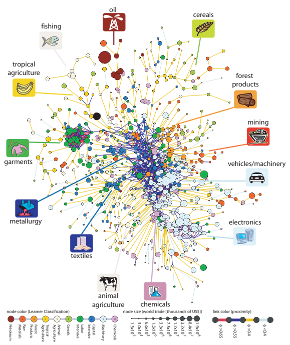


Figure 2: Ingredient complement network. Two ingredients share an edge if they occur together more than would be expected by chance and if their pointwise mutual information exceeds a threshold.

“The Product Space Conditions the Development of Nations”
Hidalgo et al.,
Science, **317**, 482–487, 2007. [52]



Networks and creativity:

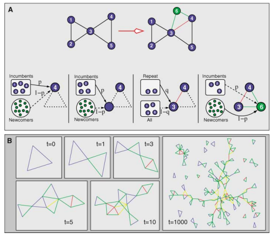
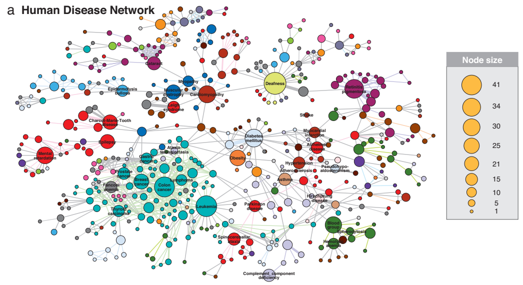


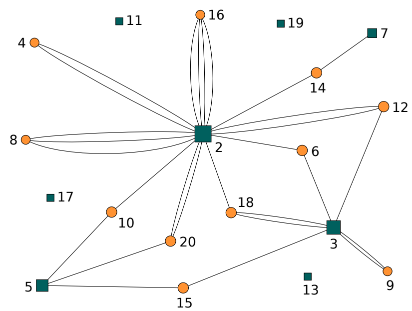
Fig. 2. Modeling the emergence of collaboration networks in creative enterprises. (**A**) Creation of a team with $m = 3$ agents. Consider, at time zero, a collaboration network comprising five agents, all incumbents (blue circles). Along with the incumbents, there is a large pool of newcomers (green circles) available to participate in new teams. Each agent in a team has a probability p of being drawn from the pool of incumbents and a probability $1 - p$ of being drawn from the pool of newcomers. For the second and subsequent agents selected from the incumbents' pool: (i) with probability q, the new agent is randomly selected from among the set of collaborators of a randomly selected incumbent already in the team; (ii) otherwise, he or she is selected at random among all incumbents in the network. For concreteness, let us assume that incumbent 4 is selected as the first agent in the new team (leftmost box). Let us also assume that the second agent is an incumbent, too (center-left box). In this example, the second agent is a past collaborator of agent 4, specifically agent 3 (center-right box). Lastly, the third agent is selected from the pool of newcomers; this agent becomes incumbent 6 (rightmost box). In these boxes and in the following panels and figures, blue lines indicate newcomer-newcomer collaborations, green lines indicate newcomer-incumbent collaborations, yellow lines indicate new incumbent-incumbent collaborations, and red lines indicate repeat collaborations. (**B**) Time evolution of the network of collaborations according to the model for $p = 0.5$, $q = 0.5$, and $m = 3$.

- Guimerà et al., Science 2005: [48] “Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance”
- Broadway musical industry
- Scientific collaboration in Social Psychology, Economics, Ecology, and Astronomy.

“The human disease network”
Goh et al.,
Proc. Natl. Acad. Sci., **104**, 8685–8690, 2007. [46]



“The complex architecture of primes and natural numbers”
García-Pérez, Serrano, and Boguñá,
https://arxiv.org/abs/1402.3612, 2014. [39]



Generatingfunctionology [115]

- Idea: Given a sequence $a_0, a_1, a_2, \ldots$, associate each element with a distinct function or other mathematical object.
- Well-chosen functions allow us to manipulate sequences and retrieve sequence elements.

Definition:

- The generating function (g.f.) for a sequence $\{a_n\}$ is

$$F(x) = \sum_{n=0}^{\infty} a_n x^n.$$

- Roughly: transforms a vector in R^∞ into a function defined on R^1.
- Related to Fourier, Laplace, Mellin, …

Simple examples:

Rolling dice and flipping coins:

- $p_k^{(⚀)} = \mathbf{Pr}(\text{throwing a } k) = 1/6$ where $k = 1, 2, \ldots, 6.$

$$F^{(⚀)}(x) = \sum_{k=1}^{6} p_k^{(⚀)} x^k = \frac{1}{6}(x + x^2 + x^3 + x^4 + x^5 + x^6).$$

- $p_0^{(\text{coin})} = \mathbf{Pr}(\text{head}) = 1/2,\ p_1^{(\text{coin})} = \mathbf{Pr}(\text{tail}) = 1/2.$

$$F^{(\text{coin})}(x) = p_0^{(\text{coin})} x^0 + p_1^{(\text{coin})} x^1 = \frac{1}{2}(1 + x).$$

- A generating function for a probability distribution is called a Probability Generating Function (p.g.f.).
- We'll come back to these simple examples as we derive various delicious properties of generating functions.

Useful pieces for probability distributions:

- Normalization:

$$F(1) = 1$$

- First moment:

$$\langle k \rangle = F'(1)$$

- Higher moments:

$$\langle k^n \rangle = \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^n F(x) \bigg|_{x=1}$$

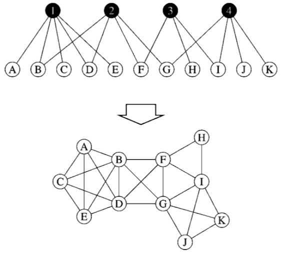
- kth element of sequence (general):

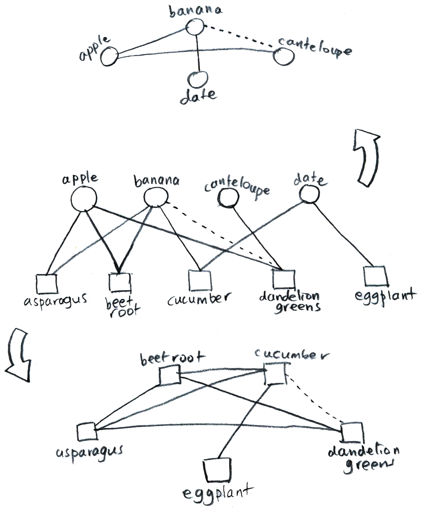
$$P_k = \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}x^k} F(x) \bigg|_{x=0}$$

Random bipartite networks:

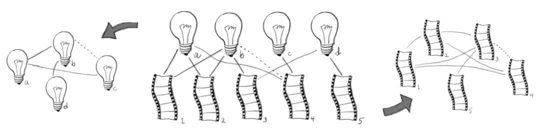
We'll follow this rather well cited paper:

"Random graphs with arbitrary degree distributions and their applications"
Newman, Strogatz, and Watts,
Phys. Rev. E, **64**, 026118, 2001. [80]





Example of a bipartite affiliation network and the induced networks:

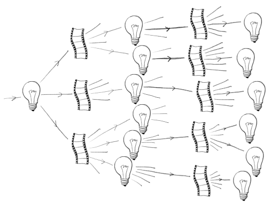


- Center: A small story-trope bipartite graph. [28]
- Induced trope network and the induced story network are on the left and right.
- The dashed edge in the bipartite affiliation network indicates an edge added to the system, resulting in the dashed edges being added to the two induced networks.

Basic story:

- An example of two inter-affiliated types:
  - 🎞 = stories,
  - 💡 = tropes.
- Stories contain tropes, tropes are in stories.
- Consider a story-trope system with $N_{\text{🎞}}$ = # stories and $N_{\text{💡}}$ = # tropes.
- $m_{\text{🎞},\text{💡}}$ = number of edges between 🎞 and 💡.
- Let's have some underlying distributions for numbers of affiliations: $P_k^{(\text{🎞})}$ (a story has k tropes) and $P_k^{(\text{💡})}$ (a trope is in k stories).
- Average number of affiliations: $\langle k\rangle_{\text{🎞}}$ and $\langle k\rangle_{\text{💡}}$.
  - $\langle k\rangle_{\text{🎞}}$ = average number of tropes per story.
  - $\langle k\rangle_{\text{💡}}$ = average number of stories containing a given trope.
- Must have balance: $N_{\text{🎞}} \cdot \langle k\rangle_{\text{🎞}} = m_{\text{🎞},\text{💡}} = N_{\text{💡}} \cdot \langle k\rangle_{\text{💡}}$.

Spreading through bipartite networks:



- View as bouncing back and forth between the two connected populations. [28]
- Actual spread may be within only one population (ideas between between people) or through both (failures in physical and communication networks).
- The gain ratio for simple contagion on a bipartite random network = product of two gain ratios.

Usual helpers for understanding network's structure:

- Randomly select an edge connecting a 🎞 to a 💡.
- Probability the 🎞 contains k other tropes:

$$R_k^{(\text{🎞})} = \frac{(k+1)P_{k+1}^{(\text{🎞})}}{\sum_{j=0}^{N_{\text{🎞}}}(j+1)P_{j+1}^{(\text{🎞})}} = \frac{(k+1)P_{k+1}^{(\text{🎞})}}{\langle k\rangle_{\text{🎞}}}.$$

- Probability the 💡 is in k other stories:

$$R_k^{(\text{💡})} = \frac{(k+1)P_{k+1}^{(\text{💡})}}{\sum_{j=0}^{N_{\text{💡}}}(j+1)P_{j+1}^{(\text{💡})}} = \frac{(k+1)P_{k+1}^{(\text{💡})}}{\langle k\rangle_{\text{💡}}}.$$

Networks of 🎞 and 💡 within bipartite structure:

- $P_{\text{ind},k}^{(\text{🎞})}$ = probability a random 🎞 is connected to k stories by sharing at least one 💡.
- $P_{\text{ind},k}^{(\text{💡})}$ = probability a random 💡 is connected to k tropes by co-occurring in at least one 🎞.
- $R_{\text{ind},k}^{(\text{💡}-\text{🎞})}$ = probability a random edge leads to a 🎞 which is connected to k other stories by sharing at least one 💡.
- $R_{\text{ind},k}^{(\text{🎞}-\text{💡})}$ = probability a random edge leads to a 💡 which is connected to k other tropes by co-occurring in at least one 🎞.
- Goal: find these distributions □.
- Another goal: find the induced distribution of component sizes and a test for the presence or absence of a giant component.
- Unrelated goal: be 10% happier/weep less.

Unstoppable spreading: Is this thing connected?

- Always about the edges: when following a random edge toward a 🎞, what's the expected number of new edges leading to other stories via tropes?
- We want to determine $\langle k\rangle_{R,\text{🎞},\text{ind}} = F'_{R_{\text{ind}}^{(\text{💡}-\text{🎞})}}(1)$ (and $F'_{R_{\text{ind}}^{(\text{🎞}-\text{💡})}}(1)$ for the trope side of things).
- We compute with joy:

$$\langle k\rangle_{R,\text{🎞},\text{ind}} = \frac{\mathrm{d}}{\mathrm{d}x} F_{R_{\text{ind},k}^{(\text{💡}-\text{🎞})}}(x)\Big|_{x=1} = \frac{\mathrm{d}}{\mathrm{d}x} F_{R^{(\text{🎞})}}\left(F_{R^{(\text{💡})}}(x)\right)\Big|_{x=1}$$

$$= F'_{R^{(\text{💡})}}(1)F'_{R^{(\text{🎞})}}\left(F_{R^{(\text{💡})}}(1)\right) = F'_{R^{(\text{💡})}}(1)F'_{R^{(\text{🎞})}}(1) = \frac{F''_{P^{(\text{💡})}}(1)}{F'_{P^{(\text{💡})}}(1)}\frac{F''_{P^{(\text{🎞})}}(1)}{F'_{P^{(\text{🎞})}}(1)}$$

- Note symmetry.
- $happiness++;

- In terms of the underlying distributions:

$$\langle k\rangle_{R,\text{🎞},\text{ind}} = \frac{\langle k(k-1)\rangle_{\text{🎞}}}{\langle k\rangle_{\text{🎞}}}\frac{\langle k(k-1)\rangle_{\text{💡}}}{\langle k\rangle_{\text{💡}}}$$

- We have a giant component in both induced networks when

$$\langle k\rangle_{R,\text{🎞},\text{ind}} \equiv \langle k\rangle_{R,\text{💡},\text{ind}} > 1$$

.

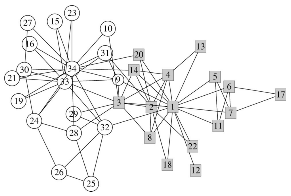
- See this as the product of two gain ratios. #excellent #physics
- We can mess with this condition to make it mathematically pleasant and pleasantly inscrutable:

$$\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} kk'(kk'-k-k')P_k^{(\text{🎞})}P_{k'}^{(\text{💡})} = 0.$$

Nutshell

- Generating functions allow us to strangely calculate features of random networks.
- They're a bit scary and magical.
- Generating functions can be used to study contagion.
- But: For essential results like possibility and probability of global spread, more direct, physics-bearing calculations are possible.
- Good real thing: Bipartite affiliation structures.
- Groups, groups, groups, ...

Structure detection

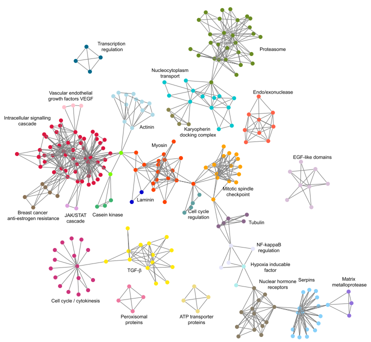


▲ Zachary's karate club [119, 79]

- The issue: how do we elucidate the internal structure of large networks across many scales?

- Possible substructures: hierarchies, cliques, rings, ...
- Plus: All combinations of substructures.
- Much focus on hierarchies (pyramids) ......

"Community detection in graphs"
Santo Fortunato,
Physics Reports, **486**, 75–174, 2010. [38]

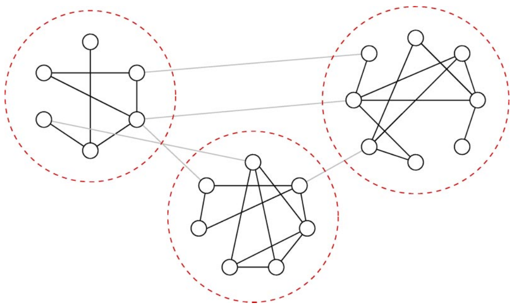


Hierarchy by division

Top down:

- Idea: Identify global structure first and recursively uncover more detailed structure.
- Basic objective: find dominant components that have significantly more links within than without, as compared to randomized version.
- We'll first work through "Finding and evaluating community structure in networks" by Newman and Girvan (PRE, 2004). [79]
- See also
  1. "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality" by Newman (PRE, 2001). [75, 78]
  2. "Community structure in social and biological networks" by Girvan and Newman (PNAS, 2002). [42]

Hierarchy by division



- Idea: Edges that connect communities have higher betweenness than edges within communities.

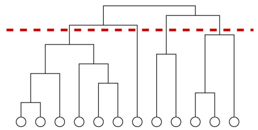
Hierarchy by division

One class of structure-detection algorithms:

1. Compute edge betweenness for whole network.
2. Remove edge with highest betweenness.
3. Recompute edge betweenness
4. Repeat steps 2 and 3 until all edges are removed.

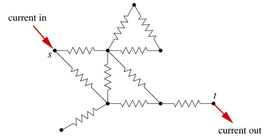
5 Record when components appear as a function of # edges removed.

6 Generate dendogram revealing hierarchical structure.



Red line indicates appearance of four (4) components at a certain level.

Betweenness for electrons:



- Unit resistors on each edge.
- For every pair of nodes s (source) and t (sink), set up unit currents in at s and out at t.
- Measure absolute current along each edge ℓ, $|I_{\ell,st}|$.
- Sum $|I_{\ell,st}|$ over all pairs of nodes to obtain electronic betweenness for edge ℓ.
- (Equivalent to random walk betweenness.)
- Contributing electronic betweenness for edge between nodes i and j:

$$B^{\text{elec}}_{ij,st} = a_{ij}|V_{i,st} - V_{j,st}|.$$

Electronic betweenness

- Define some arbitrary voltage reference.
- Kirchhoff's laws: current flowing out of node i must balance:

$$\sum_{j=1}^{N} \frac{1}{R_{ij}}(V_j - V_i) = \delta_{is} - \delta_{it}.$$

- Between connected nodes, $R_{ij} = 1 = a_{ij} = 1/a_{ij}$.
- Between unconnected nodes, $R_{ij} = \infty = 1/a_{ij}$.
- We can therefore write:

$$\sum_{j=1}^{N} a_{ij}(V_i - V_j) = \delta_{is} - \delta_{it}.$$

- Some gentle jiggery-pokery on the left hand side:
  $\sum_j a_{ij}(V_i - V_j) = V_i \sum_j a_{ij} - \sum_j a_{ij}V_j$
  $= V_i k_i - \sum_j a_{ij}V_j = \sum_j [k_i \delta_{ij} V_j - a_{ij} V_j]$
  $= [(\mathbf{K} - \mathbf{A})\vec{V}]_i$

Electronic betweenness

- Write right hand side as $[I^{\text{ext}}]_{i,st} = \delta_{is} - \delta_{it}$, where I^{ext}_{st} holds external source and sink currents.
- Matrixingly then:

$$(\mathbf{K} - \mathbf{A})\vec{V} = I^{\text{ext}}_{st}.$$

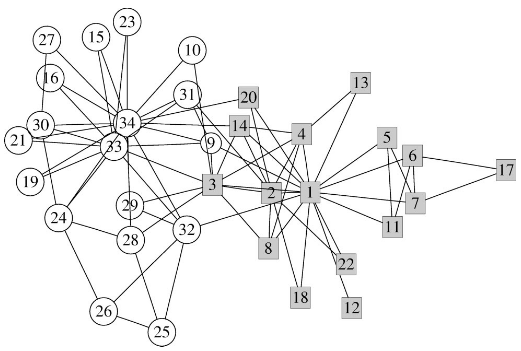
- $\mathbf{L} = \mathbf{K} - \mathbf{A}$ is a beast of some utility—known as the Laplacian.
- Solve for voltage vector $\vec{V}$ by **LU** decomposition (Gaussian elimination).
- Do not compute an inverse!
- Note: voltage offset is arbitrary so no unique solution.
- Presuming network has one component, null space of $\mathbf{K} - \mathbf{A}$ is one dimensional.
- In fact, $\mathcal{N}(\mathbf{K} - \mathbf{A}) = \{c\vec{1}, c \in R\}$ since $(\mathbf{K} - \mathbf{A})\vec{1} = \vec{0}$.

Alternate betweenness measures:

Random walk betweenness:

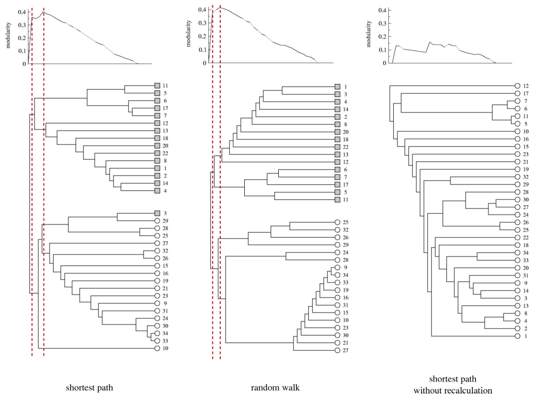
- Asking too much: Need full knowledge of network to travel along shortest paths.
- One of many alternatives: consider all random walks between pairs of nodes i and j.
- Walks starts at node i, traverses the network randomly, ending as soon as it reaches j.
- Record the number of times an edge is followed by a walk.
- Consider all pairs of nodes.
- Random walk betweenness of an edge = absolute difference in probability a random walk travels one way versus the other along the edge.
- Equivalent to electronic betweenness (see also diffusion).

Hierarchy by division



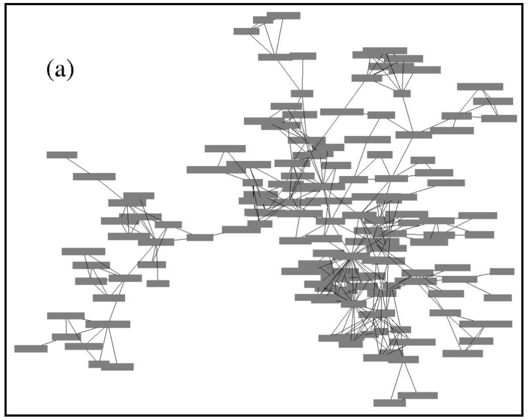
- Factions in Zachary's karate club network.[119]

Hierarchy by division

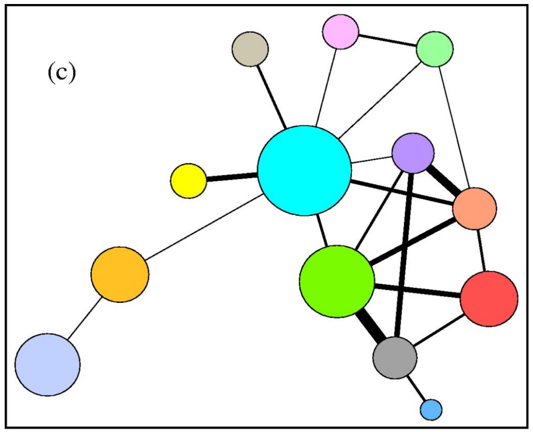


- Third column shows what happens if we don't recompute betweenness after each edge removal.

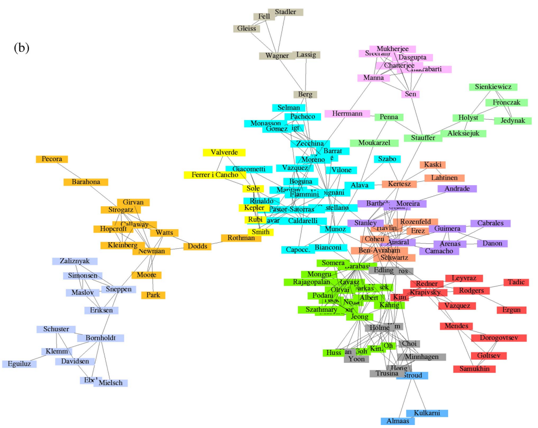
Scientists working on networks (2004)



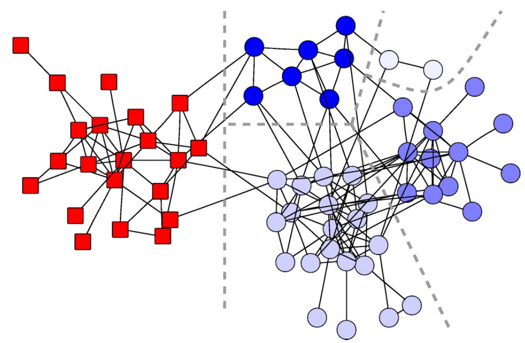
Scientists working on networks (2004)



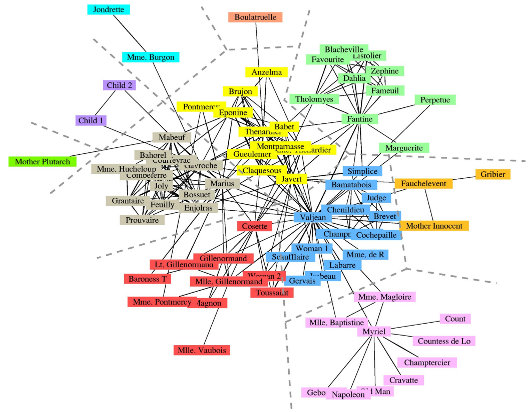
Scientists working on networks (2004)



Dolphins!



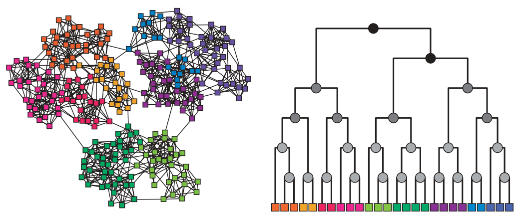
Les Miserables



- More network analyses for Les Miserables here and here.

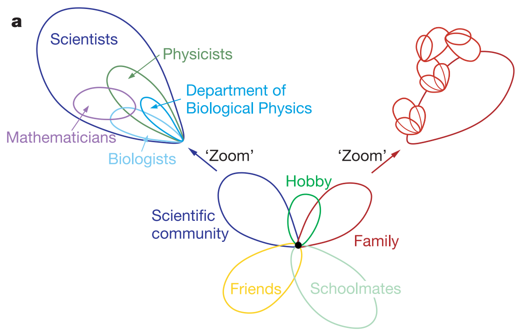
Hierarchies and missing links

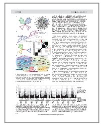
Clauset *et al.*, Nature (2008)[25]



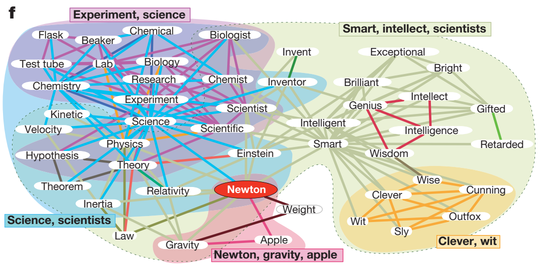
- Idea: Shades indicate probability that nodes in left and right subtrees of dendogram are connected.
- Handle: Hierarchical random graph models.
- Plan: Infer consensus dendogram for a given real network.
- Obtain probability that links are missing (big problem...).

"Uncovering the overlapping community structure of complex networks in nature and society"
Palla et al.,
Nature, **435**, 814–818, 2005.[81]



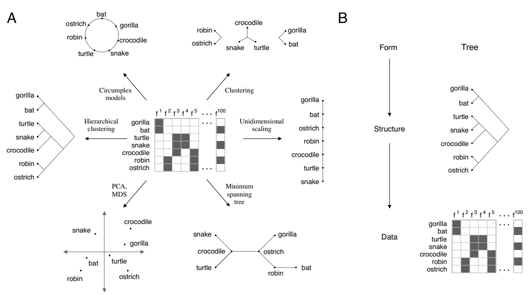


"Link communities reveal multiscale complexity in networks"
Ahn, Bagrow, and Lehmann,
Nature, **466**, 761–764, 2010.[2]

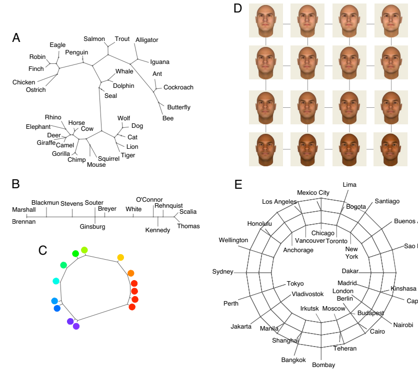


General structure detection

- "The discovery of structural form"
  Kemp and Tenenbaum, PNAS (2008)[54]



Example learned structures:



- Biological features; Supreme Court votes; perceived color differences; face differences; & distances between cities.

Nutshell:

Overview Key Points:

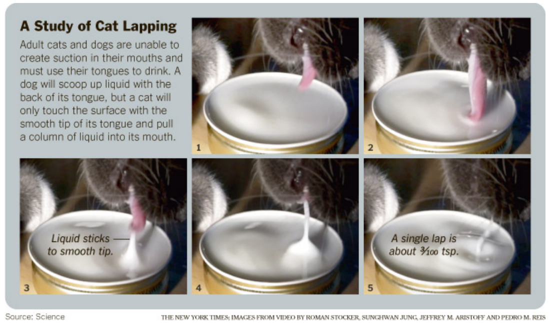
- The field of complex networks came into existence in the late 1990s.
- Explosion of papers and interest since 1998/99.
- Hardened up much thinking about complex systems.
- Specific focus on networks that are large-scale, sparse, natural or people-made, evolving and dynamic, and (crucially) measurable.
- Three main (blurred) categories:
  1. Physical (e.g., river networks),
  2. Interactional (e.g., social networks),
  3. Abstract (e.g., thesauri).
- To solve network problems: "Follow the edges."

More Allegations:

- The map is not the territory.
- Sometimes the map is not the territory because the territory does not exist.
- "But it might one day!" yelled Captain Survivor Bias IV while holding up two pineapples to gauge the distance between waves.
- And the mapper is never the map.
- (Scientific truths shouldn't be named after individuals.)

Rather silly but great example of real science:

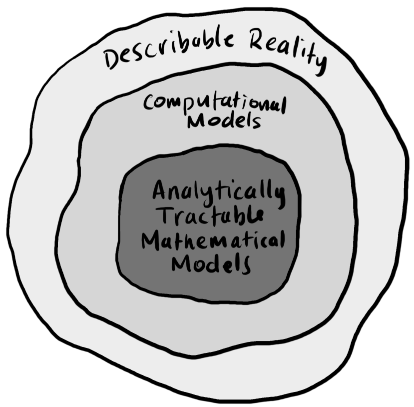
"How Cats Lap: Water Uptake by *Felis catus*"
Reis et al., *Science*, 2010.

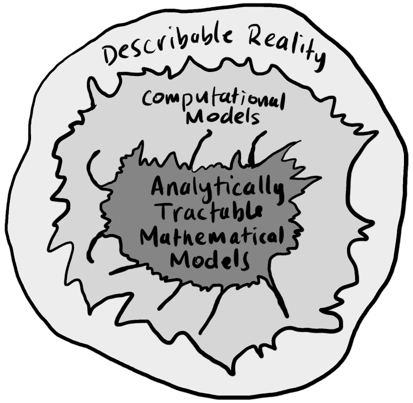


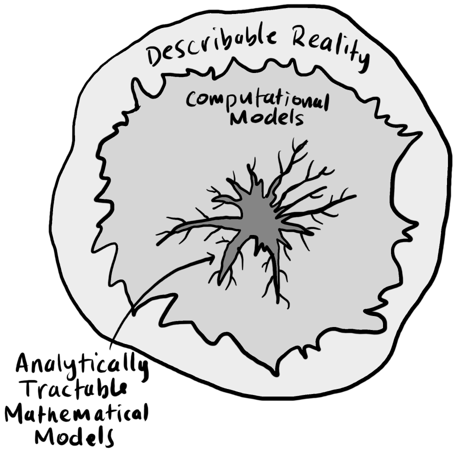
Amusing interview here

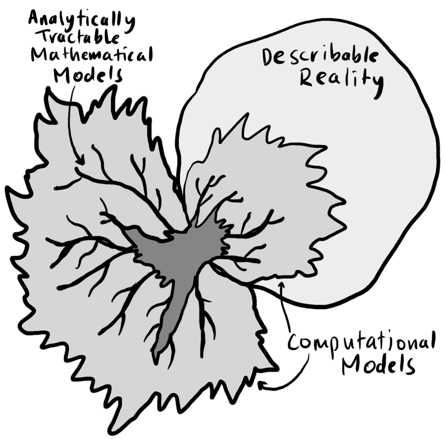
Warnings:

- Networks aren't everything.
- Famous models of networks aren't everything in networks.
- Mathematical tractability $\neq$ meaningfulness or viable existence in reality
- Even when networks are core to a system, the best level of analysis may involve some scale of grouping/averaging.
- Groups, groups, groups.
- And pyramids ($\sim$ hierarchies)









Basic Science $\simeq$ Describe + Explain:

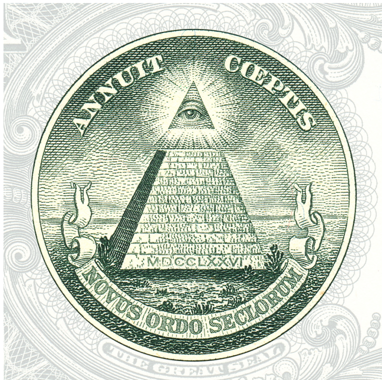
Lord Kelvin (possibly):

- "To measure is to know."
- "If you cannot measure it, you cannot improve it."

Bonus:

- "X-rays will prove to be a hoax."
- "There is nothing new to be discovered in physics now, All that remains is more and more precise measurement."
- "Beards will always be cool."

The Pyramid knows what you did.



Mass surveillance by story.

The absolute basics:

Modern basic science in three steps:

1. Find interesting/meaningful/important phenomena, optionally involving spectacular amounts of data.
2. Describe what you see.
3. Explain it.

If you succeed at 1–3:

4. Create.
5. Share.

Always:

6. Be good people.

References I

[1] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási.
Flavor network and the principles of food pairing.
Nature Scientific Reports, 1:196, 2011. pdf

[2] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann.
Link communities reveal multiscale complexity in networks.
Nature, 466(7307):761–764, 2010. pdf

[3] R. Albert and A.-L. Barabási.
Statistical mechanics of complex networks.
Rev. Mod. Phys., 74:47–97, 2002. pdf

[4] P. W. Anderson.
More is different.
Science, 177(4047):393–396, 1972. pdf

References II

[5] F. Auerbach.
Das gesetz der bevölkerungskonzentration.
Petermanns Geogr. Mitteilungen, 59:73–76, 1913.

[6] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo.
Topology of the fittest transportation network.
Phys. Rev. Lett., 84:4745–4748, 2000. pdf

[7] J. R. Banavar, A. Maritan, and A. Rinaldo.
Size and form in efficient transportation networks.
Nature, 399:130–132, 1999. pdf

[8] A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286:509–511, 1999. pdf

References III

[9] A. Bentley, M. Earls, and M. J. O'Brien.
I'll Have What She's Having: Mapping Social Behavior.
MIT Press, Cambridge, MA, 2011.

[10] S. Bikhchandani, D. Hirshleifer, and I. Welch.
A theory of fads, fashion, custom, and cultural change as informational cascades.
J. Polit. Econ., 100:992–1026, 1992.

[11] S. Bikhchandani, D. Hirshleifer, and I. Welch.
Learning from the behavior of others: Conformity, fads, and informational cascades.
J. Econ. Perspect., 12(3):151–170, 1998. pdf

[12] P. M. Blau and J. E. Schwartz.
Crosscutting Social Circles.
Academic Press, Orlando, FL, 1984.

References IV

[13] K. L. Blaxter, editor.
Energy Metabolism; Proceedings of the 3rd symposium held at Troon, Scotland, May 1964.
Academic Press, New York, 1965.

[14] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang.
Complex networks: Structure and dynamics.
Physics Reports, 424:175–308, 2006. pdf

[15] M. Boguñá and M. Ángeles Serrano.
Generalized percolation in random directed networks.
Phys. Rev. E, 72:016106, 2005. pdf

References V

[16] S. Bohn and M. O. Magnasco.
Structure, scaling, and phase transition in the optimal transport network.
Phys. Rev. Lett., 98:088702, 2007. pdf

[17] J. Bollen, B. Gonçalves, I. van de Leemput, and G. Ruan.
The happiness paradox: Your friends are happier than you.
EPJ Data Science, 6:4, 2017. pdf

[18] J. Bollen, H. Van de Sompel, A. Hagberg, L. Bettencourt, R. Chute, M. A. Rodriguez, and B. Lyudmila.
Clickstream data yields high-resolution maps of science.
PLoS ONE, 4:e4803, 2009. pdf

References VI

[19] S. Bornholdt and H. G. Schuster, editors.
Handbook of Graphs and Networks.
Wiley-VCH, Berlin, 2003.

[20] R. L. Breiger.
The duality of persons and groups.
Social Forces, 53(2):181–190, 1974. pdf

[21] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin.
Catastrophic cascade of failures in interdependent networks.
Nature, 464:1025–1028, 2010. pdf

[22] J. M. Carlson and J. Doyle.
Highly optimized tolerance: A mechanism for power laws in designed systems.
Phys. Rev. E, 60(2):1412–1427, 1999. pdf

References VII

[23] J. M. Carlson and J. Doyle.
Highly Optimized Tolerance: Robustness and design in complex systems.
Phys. Rev. Lett., 84(11):2529–2532, 2000. pdf

[24] A. Clauset, C. Moore, and M. E. J. Newman.
Structural inference of hierarchies in networks, 2006. pdf

[25] A. Clauset, C. Moore, and M. E. J. Newman.
Hierarchical structure and the prediction of missing links in networks.
Nature, 453:98–101, 2008. pdf

[26] D. J. de Solla Price.
Networks of scientific papers.
Science, 149:510–515, 1965. pdf

References VIII

[27] P. S. Dodds.
Optimal form of branching supply and collection networks.
Phys. Rev. Lett., 104(4):048702, 2010. pdf

[28] P. S. Dodds.
A simple person's approach to understanding the contagion condition for spreading processes on generalized random networks.
In S. Lehmann and Y.-Y. Ahn, editors, Spreading Dynamics in Social Systems. 2017. pdf

[29] P. S. Dodds, D. R. Dewhurst, F. F. Hazlehurst, C. M. Van Oort, L. Mitchell, A. J. Reagan, J. R. Williams, and C. M. Danforth.
Simon's fundamental rich-get-richer model entails a dominant first-mover advantage.
Physical Review E, 95:052301, 2017. pdf

References IX

[30] P. S. Dodds, K. D. Harris, and J. L. Payne.
Direct, phyisically motivated derivation of the contagion condition for spreading processes on generalized random networks.
Phys. Rev. E, 83:056122, 2011. pdf

[31] P. S. Dodds and D. H. Rothman.
Unified view of scaling laws for river networks.
Physical Review E, 59(5):4865–4877, 1999. pdf

[32] P. S. Dodds and D. H. Rothman.
Scaling, universality, and geomorphology.
Annu. Rev. Earth Planet. Sci., 28:571–610, 2000. pdf

References X

[33] P. S. Dodds and D. H. Rothman.
Geometry of river networks. I. Scaling, fluctuations, and deviations.
Physical Review E, 63(1):016115, 2001. pdf

[34] S. N. Dorogovtsev and J. F. F. Mendes.
Evolution of Networks.
Oxford University Press, Oxford, UK, 2003.

[35] Y.-H. Eom and H.-H. Jo.
Generalized friendship paradox in complex networks: The case of scientific collaboration.
Nature Scientific Reports, 4:4603, 2014. pdf

[36] J.-B. Estoup.
Gammes sténographiques: méthode et exercices pour l'acquisition de la vitesse.
Institut Sténographique, 1916.

References XI

[37] S. L. Feld.
Why your friends have more friends than you do.
Am. J. of Sociol., 96:1464–1477, 1991. pdf

[38] S. Fortunato.
Community detection in graphs.
Physics Reports, 486:75–174, 2010. pdf

[39] L. P. García-Pérez, M. A. Serrano, and M. Boguñá.
The complex architecture of primes and natural numbers, 2014.
https://arxiv.org/abs/1402.3612. pdf

[40] M. T. Gastner and M. E. J. Newman.
Optimal design of spatial distribution networks.
Phys. Rev. E, 74:016117, 2006. pdf

References XII

[41] M. T. Gastner and M. E. J. Newman.
Shape and efficiency in spatial distribution networks.
J. Stat. Mech.: Theor. & Exp., 1:P01015, 2006. pdf

[42] M. Girvan and M. E. J. Newman.
Community structure in social and biological networks.
Proc. Natl. Acad. Sci., 99:7821–7826, 2002. pdf

[43] M. Gladwell.
The Tipping Point.
Little, Brown and Company, New York, 2000.

References XIII

[44] J. P. Gleeson and D. J. Cahalane.
Seed size strongly affects cascades on random networks.
Phys. Rev. E, 75:056103, 2007. pdf

[45] W. S. Glock.
The development of drainage systems: A synoptic view.
The Geographical Review, 21:475–482, 1931. pdf

[46] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási.
The human disease network.
Proc. Natl. Acad. Sci., 104:8685–8690, 2007. pdf

References XIV

[47] M. Granovetter.
Threshold models of collective behavior.
Am. J. Sociol., 83(6):1420–1443, 1978. pdf

[48] R. Guimerà, B. Uzzi, J. Spiro, and L. A. N. Amaral.
Team assembly mechanisms determine collaboration network structure and team performance.
Science, 308:697–702, 2005. pdf

[49] S. M. Gusein-Zade.
Bunge's problem in central place theory and its generalizations.
Geogr. Anal., 14:246–252, 1982. pdf

References XV

[50] J. T. Hack.
Studies of longitudinal stream profiles in Virginia and Maryland.
United States Geological Survey Professional Paper, 294-B:45–97, 1957. pdf

[51] A. Halevy, P. Norvig, and F. Pereira.
The unreasonable effectiveness of data.
IEEE Intelligent Systems, 24:8–12, 2009. pdf

[52] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausman.
The product space conditions the development of nations.
Science, 317:482–487, 2007. pdf

References XVI

[53] R. E. Horton.
Erosional development of streams and their drainage basins; hydrophysical approach to quatitative morphology.
Bulletin of the Geological Society of America, 56(3):275–370, 1945. pdf

[54] C. Kemp and J. B. Tenenbaum.
The discovery of structural form.
Proc. Natl. Acad. Sci., 105:10687–10692, 2008. pdf

[55] J. W. Kirchner.
Statistical inevitability of Horton's laws and the apparent randomness of stream channel networks.
Geology, 21:591–594, 1993. pdf

References XVII

[56] J. M. Kleinberg.
Authoritative sources in a hyperlinked environment.
Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. pdf

[57] P. L. Krapivsky and S. Redner.
Organization of growing random networks.
Phys. Rev. E, 63:066123, 2001. pdf

[58] M. Kretzschmar and M. Morris.
Measures of concurrency in networks and the spread of infectious disease.
Math. Biosci., 133:165–95, 1996. pdf

References XVIII

[59] P. La Barbera and R. Rosso.
Reply.
Water Resources Research, 26(9):2245–2248, 1990. pdf

[60] L. B. Leopold.
A View of the River.
Harvard University Press, Cambridge, MA, 1994.

[61] A. J. Lotka.
The frequency distribution of scientific productivity.
Journal of the Washington Academy of Science, 16:317–323, 1926.

References XIX

[62] B. B. Mandelbrot.
An informational theory of the statistical structure of languages.
In W. Jackson, editor, Communication Theory, pages 486–502. Butterworth, Woburn, MA, 1953. pdf

[63] B. B. Mandelbrot.
A note on a class of skew distribution function. Analysis and critique of a paper by H. A. Simon.
Information and Control, 2:90–99, 1959.

[64] B. B. Mandelbrot.
Final note on a class of skew distribution functions: analysis and critique of a model due to H. A. Simon.
Information and Control, 4:198–216, 1961.

References XX

[65] B. B. Mandelbrot.
Post scriptum to 'final note'.
Information and Control, 4:300–304, 1961.

[66] K. A. McCulloh, J. S. Sperry, and F. R. Adler.
Water transport in plants obeys Murray's law.
Nature, 421:939–942, 2003. pdf

[67] K. A. McCulloh, J. S. Sperry, and F. R. Adler.
Murray's law and the hydraulic vs mechanical functioning of wood.
Functional Ecology, 18:931–938, 2004. pdf

[68] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon.
On the uniform generation of random graphs with prescribed degree sequences, 2003. pdf

References XXI

[69] D. R. Montgomery and W. E. Dietrich.
Channel initiation and the problem of landscape scale.
Science, 255:826–30, 1992. pdf

[70] R. Munroe.
Thing Explainer: Complicated Stuff in Simple Words.
Houghton Mifflin Harcourt, 2015.

[71] C. D. Murray.
The physiological principle of minimum work applied to the angle of branching of arteries.
J. Gen. Physiol., 9(9):835–841, 1926. pdf

References XXII

[72] C. D. Murray.
The physiological principle of minimum work. I. The vascular system and the cost of blood volume.
Proc. Natl. Acad. Sci., 12:207–214, 1926. pdf

[73] C. D. Murray.
A relationship between circumference and weight in trees and its bearing on branching angles.
J. Gen. Physiol., 10:725–729, 1927. pdf

[74] M. Newman.
Assortative mixing in networks.
Phys. Rev. Lett., 89:208701, 2002. pdf

References XXIII

[75] M. E. J. Newman.
Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality.
Phys. Rev. E, 64(1):016132, 2001. pdf

[76] M. E. J. Newman.
Ego-centered networks and the ripple effect,.
Social Networks, 25:83–95, 2003. pdf

[77] M. E. J. Newman.
The structure and function of complex networks.
SIAM Rev., 45(2):167–256, 2003. pdf

References XXIV

[78] M. E. J. Newman.
Erratum: Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality [Phys. Rev. E 64, 016132 (2001)].
Phys. Rev. E, 73:039906(E), 2006. pdf

[79] M. E. J. Newman and M. Girvan.
Finding and evaluating community structure in networks.
Phys. Rev. E, 69(2):026113, 2004. pdf

[80] M. E. J. Newman, S. H. Strogatz, and D. J. Watts.
Random graphs with arbitrary degree distributions and their applications.
Phys. Rev. E, 64:026118, 2001. pdf

References XXV

[81] G. Palla, I. Derényi, I. Farkas, and T. Vicsek.
Uncovering the overlapping community structure of complex networks in nature and society.
Nature, 435(7043):814–818, 2005. pdf

[82] T. Pratchett.
The Truth.
HarperCollins, 2000.

[83] D. D. S. Price.
A general theory of bibliometric and other cumulative advantage processes.
Journal of the American Society for Information Science, pages 292–306, 1976. pdf

References XXVI

[84] I. Rodríguez-Iturbe and A. Rinaldo.
Fractal River Basins: Chance and Self-Organization.
Cambridge University Press, Cambrigde, UK, 1997.

[85] T. C. Schelling.
Dynamic models of segregation.
J. Math. Sociol., 1:143–186, 1971. pdf

[86] T. C. Schelling.
Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities.
J. Conflict Resolut., 17:381–428, 1973. pdf

References XXVII

[87] T. C. Schelling.
Micromotives and Macrobehavior.
Norton, New York, 1978.

[88] S. A. Schumm.
Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey.
Bulletin of the Geological Society of America, 67:597–646, 1956. pdf

[89] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon.
Network motifs in the transcriptional regulation network of *Escherichia coli*.
Nature Genetics, 31:64–68, 2002. pdf

References XXVIII

[90] T. F. Sherman.
On connecting large vessels to small. The meaning of Murray's law.
The Journal of general physiology, 78(4):431–453, 1981. pdf

[91] G. Simmel.
The number of members as determining the sociological form of the group. I.
American Journal of Sociology, 8:1–46, 1902.

[92] H. A. Simon.
On a class of skew distribution functions.
Biometrika, 42:425–440, 1955. pdf

References XXIX

[93] H. A. Simon.
Some further notes on a class of skew distribution functions.
Information and Control, 3:80–88, 1960.

[94] H. A. Simon.
Reply to Dr. Mandelbrot's post scriptum.
Information and Control, 4:305–308, 1961.

[95] H. A. Simon.
Reply to 'final note' by Benoît Mandelbrot.
Information and Control, 4:217–223, 1961.

[96] D. Sornette.
Critical Phenomena in Natural Sciences.
Springer-Verlag, Berlin, 1st edition, 2003.

References XXX

[97] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic.
Recipe recommendation using ingredient networks.
In Proceedings of the 3rd Annual ACM Web Science Conference, WebSci '12, pages 298–307, New York, NY, USA, 2012. ACM. pdf

[98] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki.
Rules for biologically inspired adaptive network design.
Science, 327(5964):439–442, 2010. pdf

[99] D. W. Thompson.
On Growth and Form.
Cambridge University Pres, Great Britain, 2nd edition, 1952.

References XXXI

[100] D. W. Thompson.
On Growth and Form — Abridged Edition.
Cambridge University Press, Great Britain, 1961.

[101] E. Tokunaga.
The composition of drainage network in Toyohira River Basin and the valuation of Horton's first law.
Geophysical Bulletin of Hokkaido University, 15:1–19, 1966. pdf

[102] E. Tokunaga.
Consideration on the composition of drainage networks and their evolution.
Geographical Reports of Tokyo Metropolitan University, 13:G1–27, 1978. pdf

References XXXII

[103] E. Tokunaga.
Ordering of divide segments and law of divide segment numbers.
Transactions of the Japanese Geomorphological Union, 5(2):71–77, 1984.

[104] J. Um, S.-W. Son, S.-I. Lee, H. Jeong, and B. J. Kim.
Scaling laws between population and facility densities.
Proc. Natl. Acad. Sci., 106:14236–14240, 2009. pdf

[105] F. Vega-Redondo.
Complex Social Networks.
Cambridge University Press, 2007.

References XXXIII

[106] D. J. Watts.
A simple model of global cascades on random networks.
Proc. Natl. Acad. Sci., 99(9):5766–5771, 2002. pdf

[107] D. J. Watts.
Six Degrees.
Norton, New York, 2003.

[108] D. J. Watts and P. S. Dodds.
Influentials, networks, and public opinion formation.
Journal of Consumer Research, 34:441–458, 2007. pdf

References XXXIV

[109] D. J. Watts and P. S. Dodds.
Threshold models of social influence.
In P. Hedström and P. Bearman, editors, The Oxford Handbook of Analytical Sociology, chapter 20, pages 475–497. Oxford University Press, Oxford, UK, 2009. pdf

[110] D. J. Watts, P. S. Dodds, and M. E. J. Newman.
Identity and search in social networks.
Science, 296:1302–1305, 2002. pdf

[111] D. J. Watts, R. Muhamad, D. Medina, and P. S. Dodds.
Multiscale, resurgent epidemics in a hierarchcial metapopulation model.
Proc. Natl. Acad. Sci., 102(32):11157–11162, 2005. pdf

References XXXV

[112] D. J. Watts and S. J. Strogatz.
Collective dynamics of 'small-world' networks.
Nature, 393:440–442, 1998. pdf

[113] G. B. West, J. H. Brown, and B. J. Enquist.
A general model for the origin of allometric scaling laws in biology.
Science, 276:122–126, 1997. pdf

[114] E. Wigner.
The unreasonable effectivenss of mathematics in the natural sciences.
Communications on Pure and Applied Mathematics, 13:1–14, 1960. pdf

[115] H. S. Wilf.
Generatingfunctionology.
A K Peters, Natick, MA, 3rd edition, 2006. pdf

References XXXVI

[116] Q. Xia.
Optimal paths related to transport problems.
Communications in Contemporary Mathematics, 5:251–279, 2003. pdf

[117] Q. Xia.
The formation of a tree leaf.
ESAIM: Control, Optimisation and Calculus of Variations, 13:359–377, 2007. pdf

[118] G. U. Yule.
A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S.
Phil. Trans. B, 213:21–87, 1925. pdf

References XXXVII

[119] W. W. Zachary.
An information flow model for conflict and fission in small groups.
J. Anthropol. Res., 33:452–473, 1977.

[120] G. K. Zipf.
Human Behaviour and the Principle of Least-Effort.
Addison-Wesley, Cambridge, MA, 1949.