Introduction to (Complex) Networks
v.0.2

Galen J. Wilkerson

wilkerson@mcc-berlin.net
Research Domain III: Sustainable Solutions,
Technical University of Berlin Department "Economics of Climate Change",
Mercator Institute for the Study of Global Commons and Climate Change

January 30, 2013

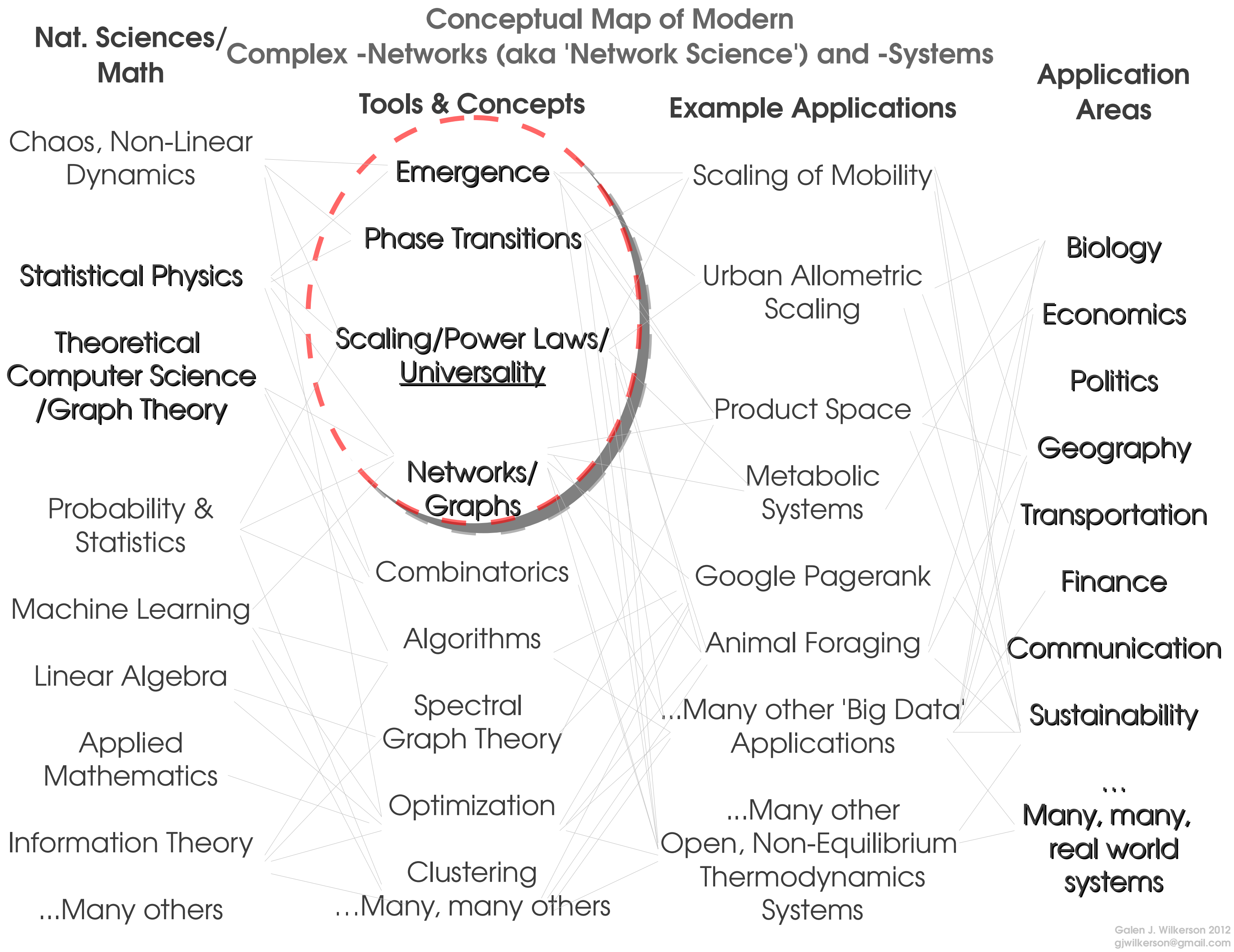


Figure 1: The field of *'Complex Networks'* almost always involves Networks and Another of the concepts in the red oval. '*Complex Systems*', also in the red oval, can include networks, but not necessarily. Arguably, they are the same, since all interactions in a system can be links in a network.

Appropriate and ironic 'Complex Networks' approximately defined - wide variety of big data and techniques to analyse → a *meta-science*.

Networks, basic terminology

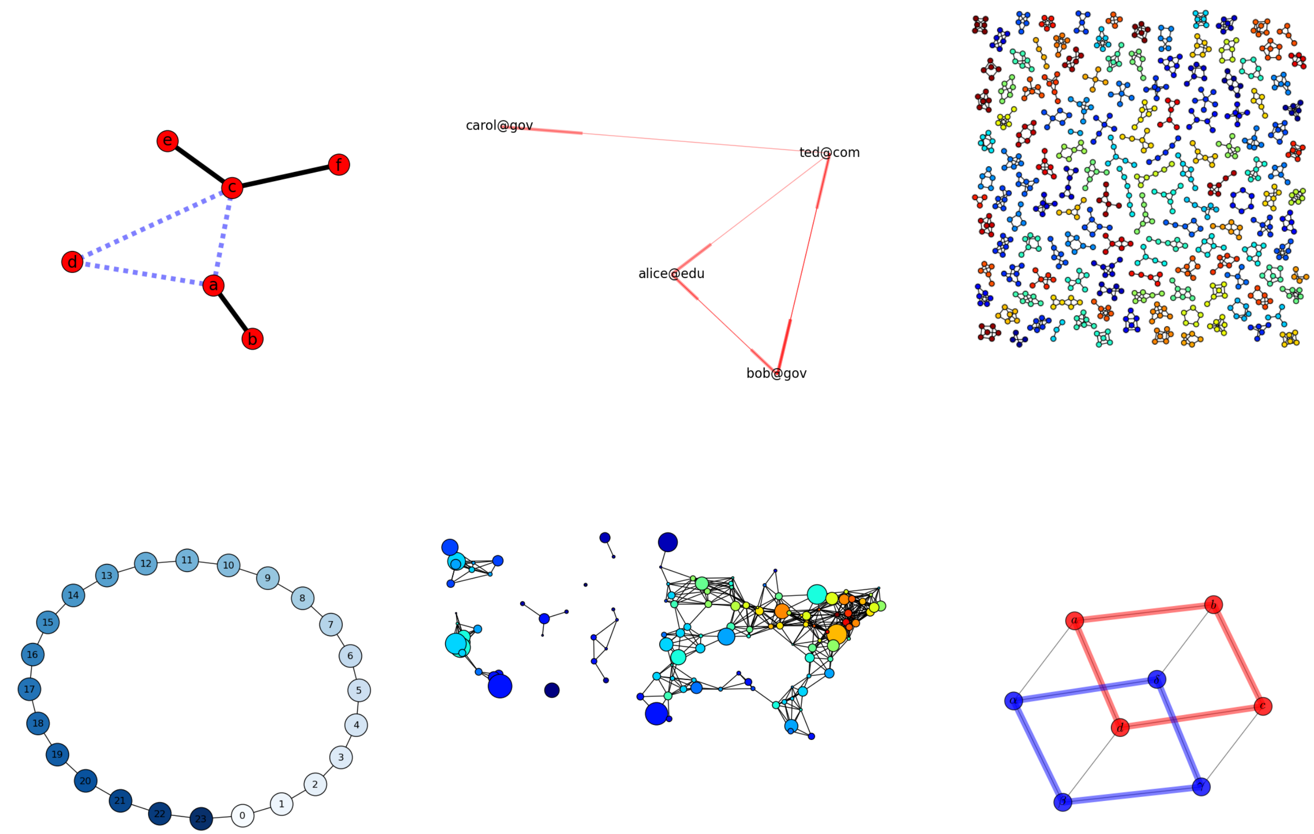


Figure 2: Some ways to represent information in networks. Top-center is a directed network of email; Bottom-center is a spatial network.

- constructed from <u>*any*</u> kind of objects and <u>*any*</u> relationships.
- set of nodes: $V = \{v1, v2, v3, ...\}$, $|V| = N$
- set of links (node pairs): $E = \{(v1, v3), (v2, v5), ...\}$, number of edges $= |E|$.
- properties, weights, labels (of nodes, links)
- directed/undirected, spatial/non-spatial, multiple edges
- Why V and E? (explained below). Why '*complex*'?

Representations (& ways of storing in memory)

adjacency matrix

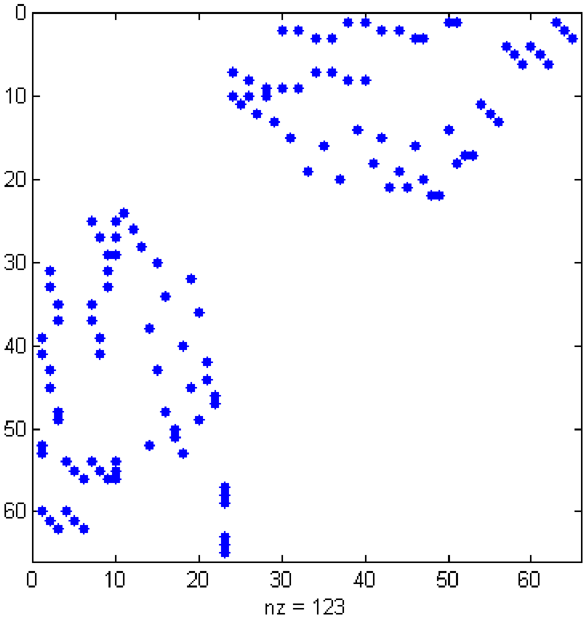


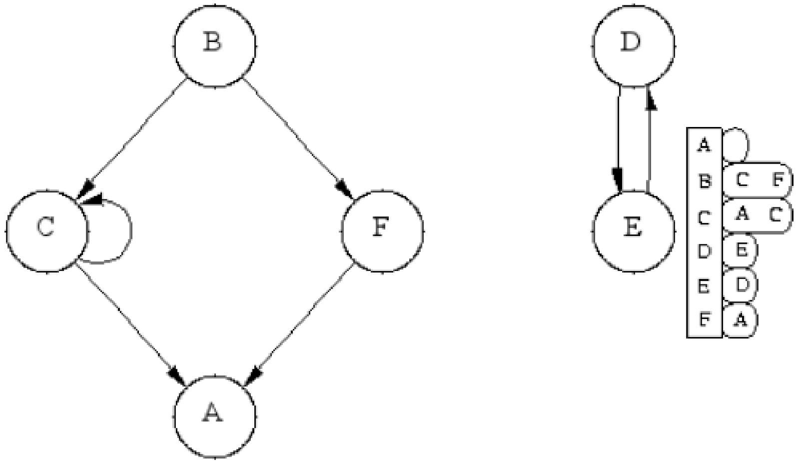
Figure 3: Graphical representation of an undirected adjacency matrix (Matlab)

... it really is a matrix! (storing link values/weights, more later) [1].
E.g.:

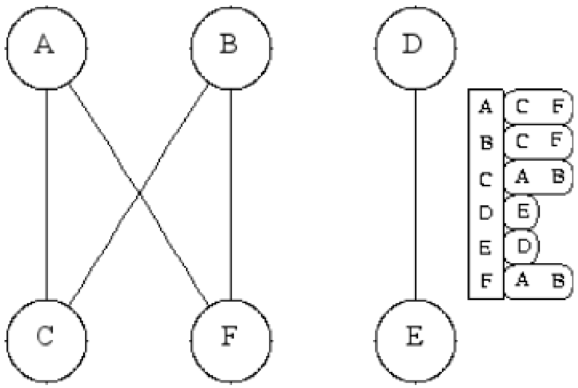
$$\mathbf{M_{ij}} = \begin{bmatrix} 0 & 5 & \cdots \\ 3 & 0 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

, where $i, j \in \{users\}$, and $e_{i \rightarrow j}$ indicates user i has sent an email to user j.
(From the matrix alone, is this directed or undirected?)

adjacency list (efficient, basically a sparse matrix storage)



(a) Directed graph; its adjacency list



(b) Undirected graph, adjacency list

Figure 4: (4a) A directed graph and its adjacency list; (4b) An undirected graph. What are the adjacency matrix representations for these graphs? (Source: boost.org)

Topology and visualisations...

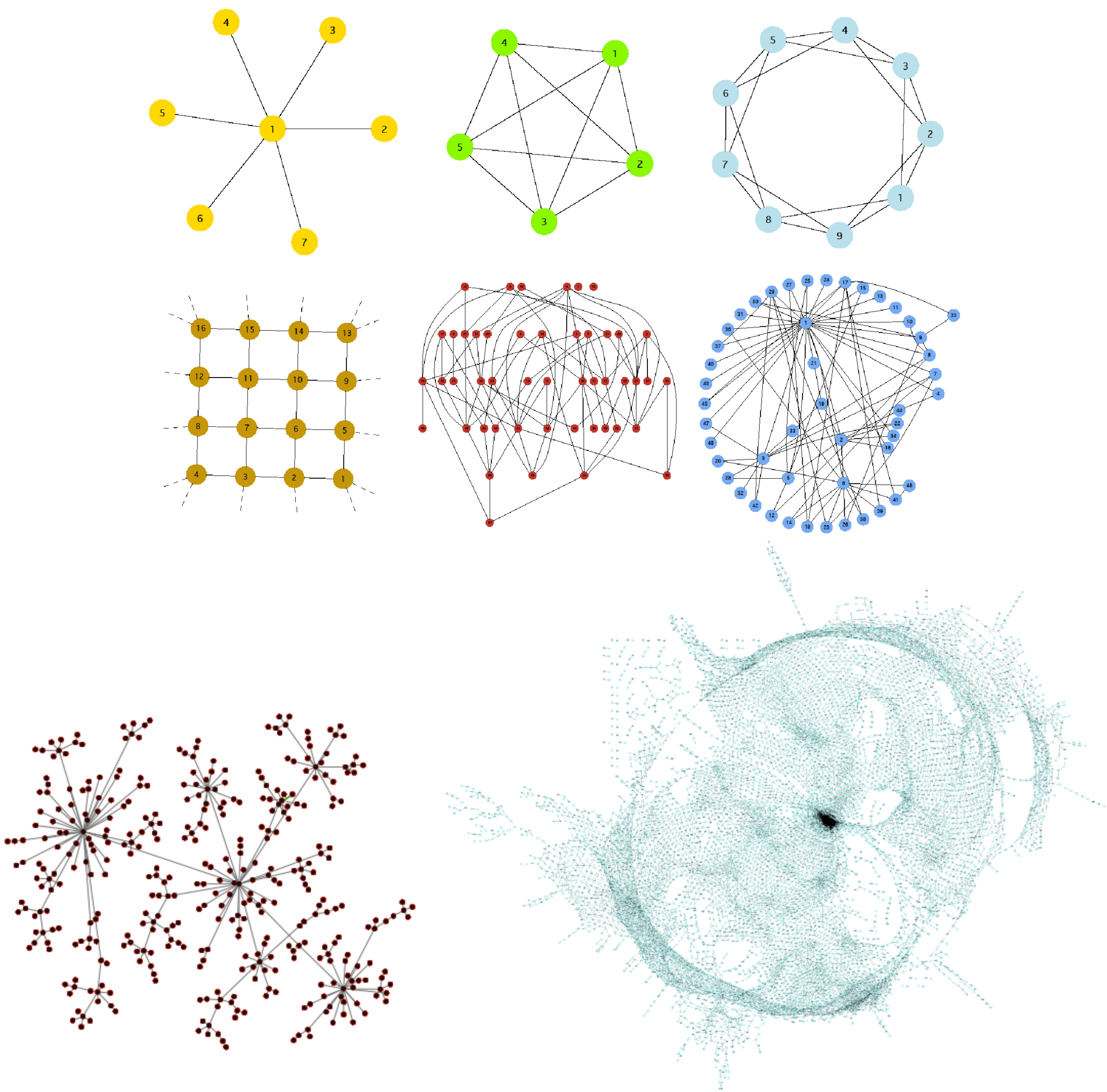


Figure 5: Some examples of network topology and layouts, star, ring, lattice, hierarchical, energy-minimizing, unwieldy, (huge) and perhaps unclear. (show *salient* features of interest) Note difference between *topology* and choice of visualisation. (some figures courtesy Alex Rademacher; others networkx)

Bipartite networks

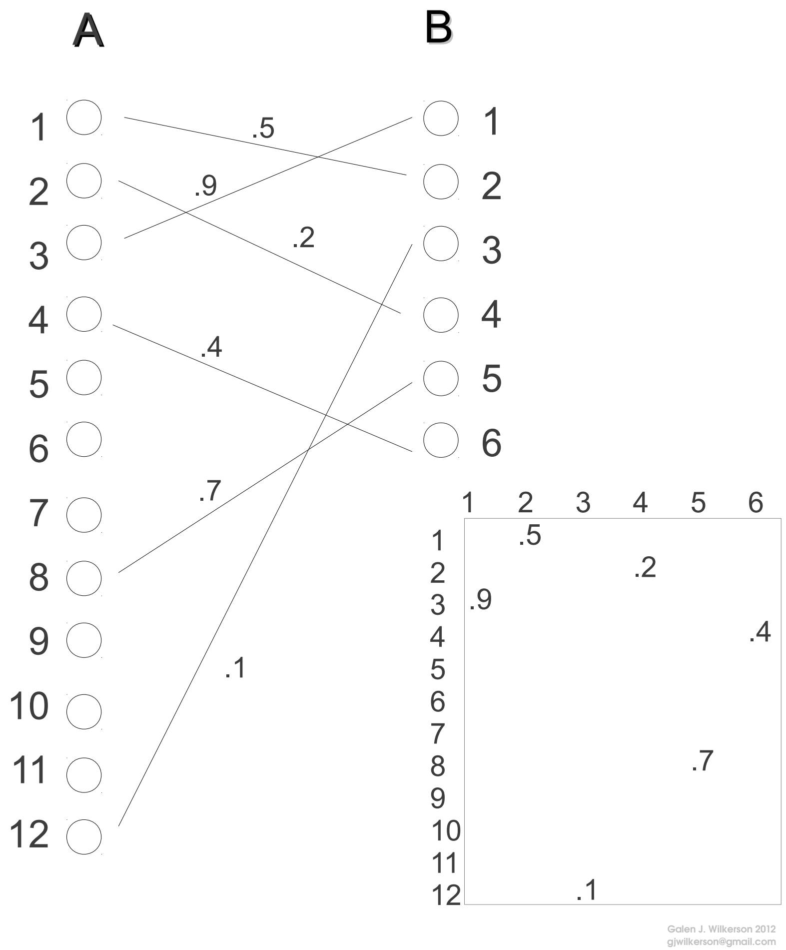
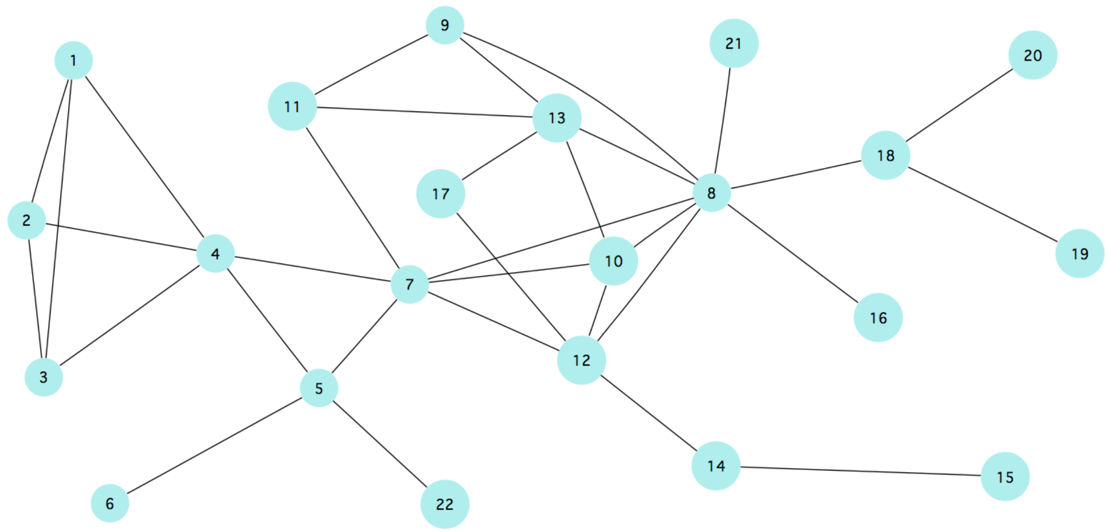
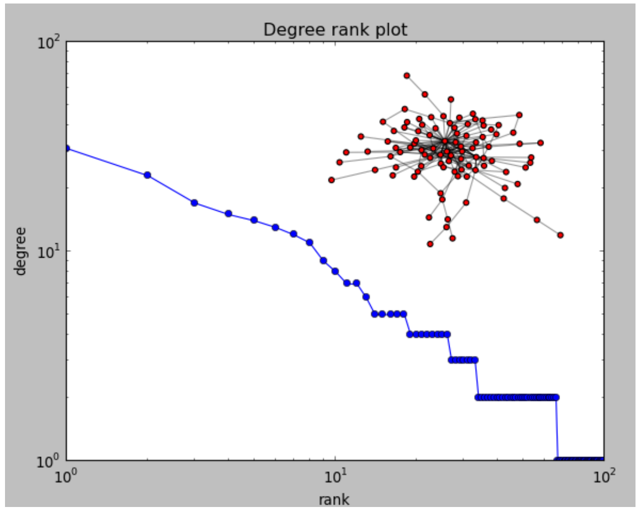


Figure 6: A bipartite network and its adjacency matrix. (Note how sparse: $\frac{6}{12*6} = \frac{1}{12} = 0.08\overline{33}$ of the possible 72 edges are occupied.)

Connects two separate sets of nodes.
$a \in \{a1, a2, ..., aN\}$
$b \in \{b1, b2, ..., bM\}$
$E = \{(a3, b1), (a5, b7), ...\}$
All connections are strictly from one set to the other.
Adjacency-matrix, -list representations still valid.
Note, can build non-bipartite network out of bipartite. For example: connect countries to countries using M_{cp}.



(a) Demo network



(b) Degree distribution is a common feature to measure

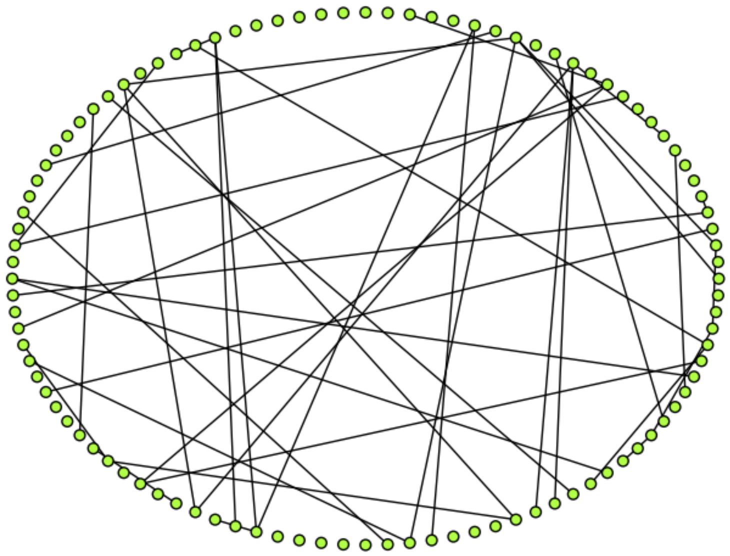
Figure 7: (7a) Notice some features of the above-left network: Clustering, betweenness, assortativity, degree distribution, diameter; (7b) The degree distribution is a common measure taken on real-world networks. (courtesy Alex Rademacher; source networkx)

Some basic measures, intuitions

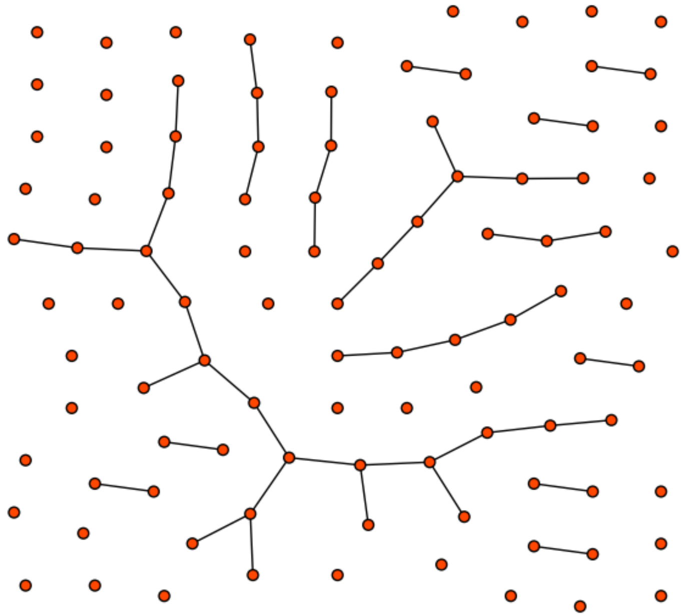
- degree distribution: p(number of neighbors)
- clustering coefficient: how many triangles (out of total possible) ?
- size, distribution, number of connected components: 'islands' in the network
- assortativity: same-degree nodes are neighbors?
- betweenness centrality: how many paths must pass through nodes? edges?
- all-pairs shortest path/diameter: average shortest distance between all pairs?
- others! Pagerank: total 'watershed' flow through each node ≈ eigen-values, -vectors of adjacency matrix 'random surfers on www network' ≈ principle components of network.
- many, many more measures and processes - nestedness, planarity, fan-out, Hamiltonian Circuit, number of 2-colorings, hyper-plane embedding, minimum spanning tree, min-cut, etc. etc.

Three significant graph/network types

1: Erdös-Renyi graphs (aka 'random' graphs)



(a) Ring layout



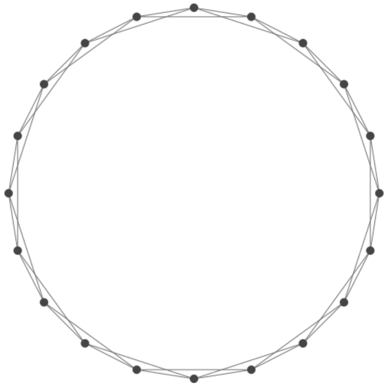
(b) Force-directed layout

Figure 8: Two random networks, both having 100 nodes and $p = .01$ (networkx and graphviz)

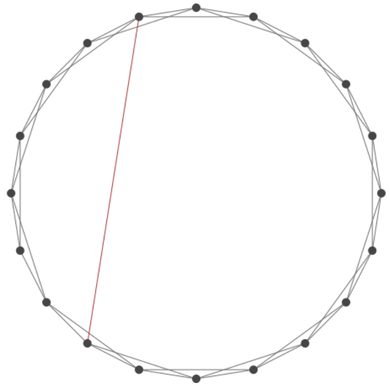
.

- Create random graphs: connect each pair of nodes with fixed probability p
- Wait, what are 'graphs'? (graph theory vs. (complex) 'network science')
  - theoretical objects, study many graphs not found in real world,
  - or phenomena that are not always relevant to 'system properties' of real networks (e.g. coloring).
- algorithms, traversals (BFS, DFS), NP Completeness. Possibly, but not always the best approach for real networks.
- terminology 'graphs', 'vertices' (V), 'edges'(E) - used interchangeably with 'networks', 'nodes', 'links'.

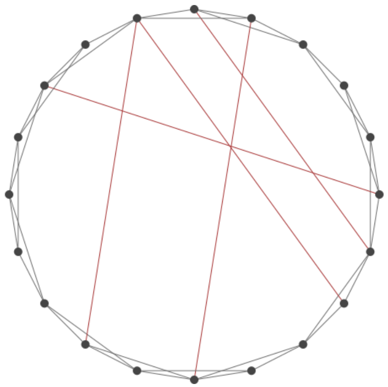
2: Small-world networks



(a) No rewirings



(b) 1 rewiring

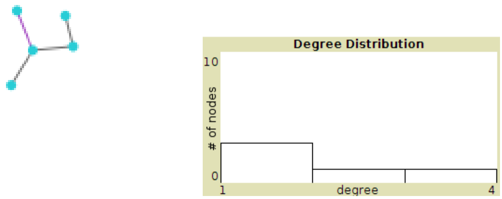


(c) 5 rewirings

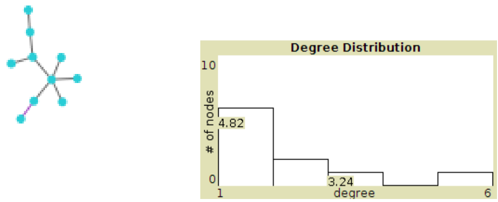
Figure 9: Small-world network ($N = 20$), (9a) with no re-wirings, having average path length = 2.895, ; (9b) with 1 re-wiring, average path length drops considerably, to = 2.7; (9c) with only 5 re-wirings, average path length drops to 2.379. Does it keep decreasing or converge to a positive value?

Create regular networks, re-wire each pair with probability p.

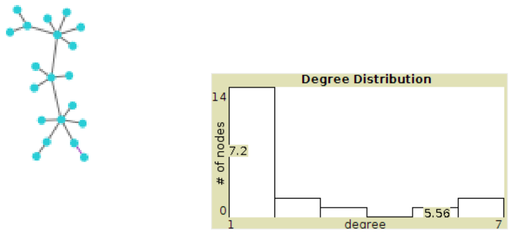
3: Scale-free networks



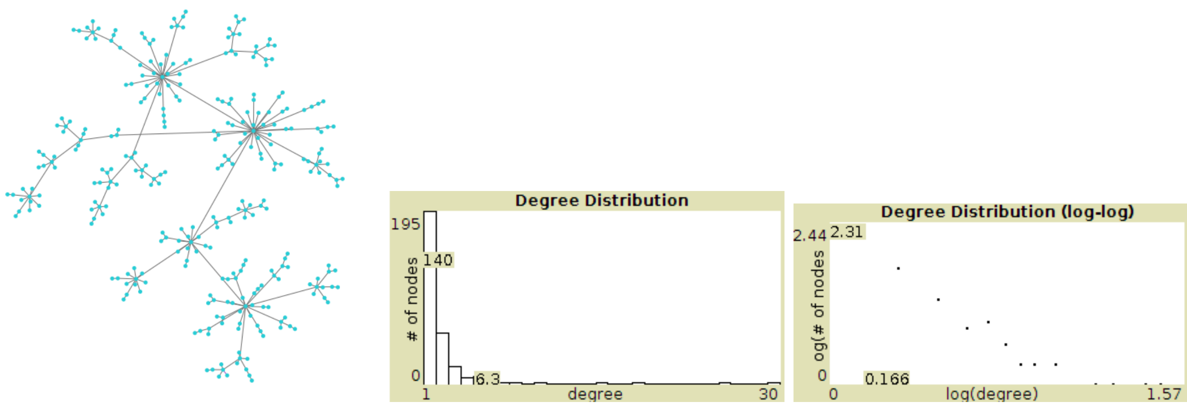
(a) 5 nodes



(b) 10 nodes



(c) 20 nodes



(d) 302 nodes

Key Concepts: 1: Preferential Attachment

Preferential attachment: At a party, new people keep coming in room. Probability of new people being friends with 'old' people in room is proportional to number of friends 'old' people have. Also known as *Yule Process* [6, 2]. Structure and *growth process* of scale-free networks tied together.

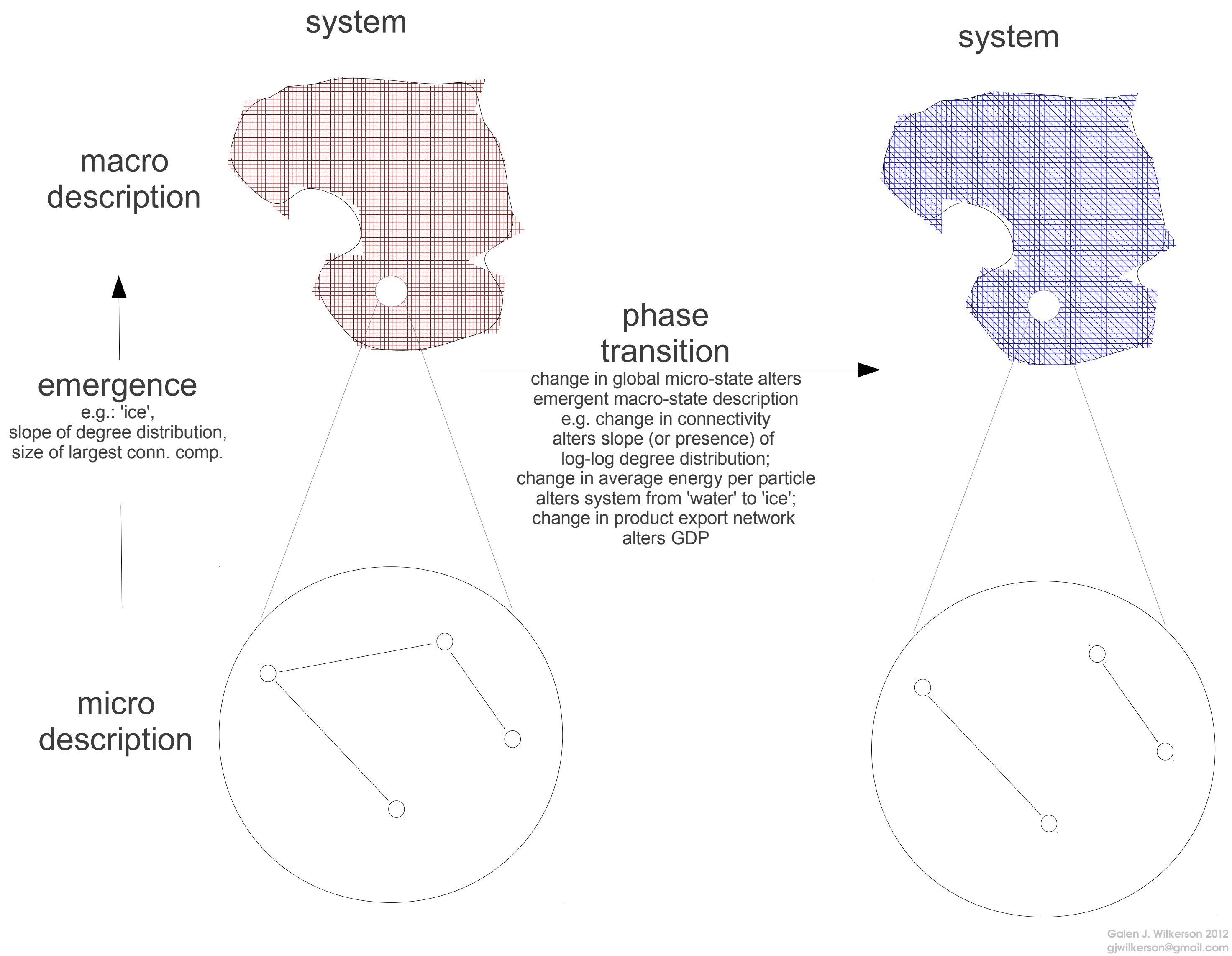


Figure 10: *Emergence* describes micro to macro relationships; and *phase transitions* describe changes in overall macro-states due to changes in micro-states.

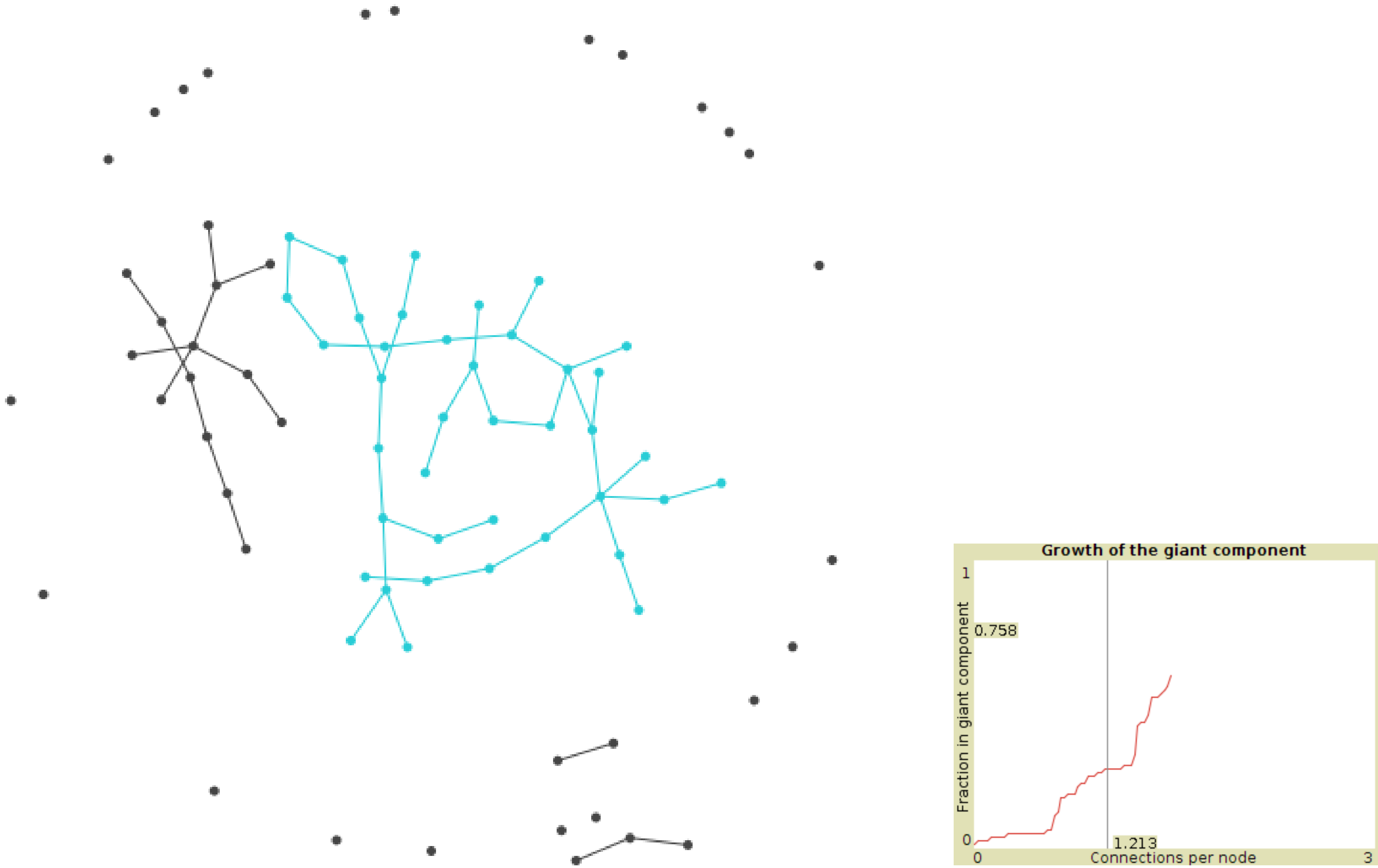


Figure 11: A *phase transition* in giant components occurs in random networks as p is increased. The 'percolation threshold' ≈ 'critical point' is $< k >= 1$, i.e. the $p_c = \frac{1}{N}$. (think about $< k >$, N, and p. see vertical line, above-right) (netlogo)

During phase transition, system passes through a state with no particular dominating state: e.g. correlation length in materials or fractals.

Scale-free networks are *fractal* (self-similar on different scales), exponent gives the relationships between different scales.

Universality comes because many systems have this relation between hubs (high-degree nodes) and low-degree nodes.

4: The Big Deal: Universality; Die Lorelei of Power Law Distributions, or ”Look, another power law!!”

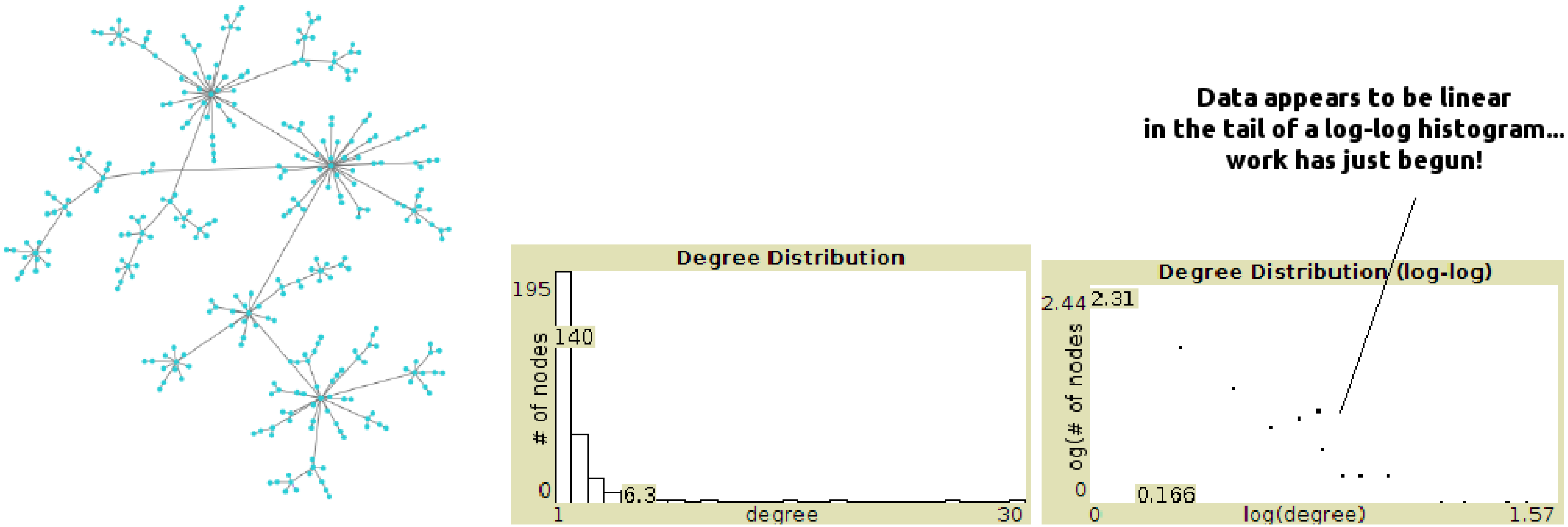


Figure 12: The degree distribution of this network looks like it is fit by a power law, but is it?

General form $p(x) = Cx^{-\alpha}$
Why the line? $\log p(x) = \log(Cx^{-\alpha}) = \log(C) - \alpha \log(x)$

Power laws seem to occur often in nature, as a result of certain kinds of processes [14]. This is very attractive to those looking for fundamental laws of nature. However, one must be *careful*.

Many distributions can look linear! Remember what log does (orders of magnitude).

- Look at the tail. (what does the tail mean in a network?) Are there at least 3 orders of magnitude on x-axis in the ’linear’ tail?
- Only use mathematically-rigorous techniques for fitting and comparison to other fits [6, 19]:
  - Find xmin, α using Kolmogorov-Smirnov Statistic.
  - Find goodness-of-fit (p-value)
  - Compare to other distributions using likelihood ratio test.
- What is the exponent? (the slope of the log-log line) often called α, β, or γ) *Does your value make any sense?*
- Investigate mechanisms for generation [14].

Thank you! [1]

Questions?

Galen J. Wilkerson
Visiting Scientist
wilkerson@mcc-berlin.net

[1] Thanks to Alex Rademacher, MCC for his contributions.

Appendices

Some tools

Matlab

(slowish, but very easy to program, useful for linear algebra on adjacency matrices) various libraries in Mathworks pages

- Code by Aaron Clauset, Yogesh Virkar for KS-statistic and Likelihood Ratio for fitting power laws.
- sparse matrices
- statistical toolbox
- bioinformatics toolbox

Python

(slow but very easy to program)

- networkx (easy, high-level)
- igraph

R

slowish, high-level statistics

- Code by Aaron Clauset, Yogesh Virkar for KS-statistic and Likelihood Ratio for fitting power laws.

C/C++

(very fast, some more programming overhead)

- Boost
- iGraph
- STL

Quick Literature Review

General References on Complex Systems/Networks

- A recent text and older introductory article on Networks by Marc Newman [12, 13].
- Online book/notes by Laszlo Barabasi. barabasilab.com
- Lay-person's book, also by Barabasi [3].
- Melanie Mitchell's introductory text on Complex Systems [8].

References on Key Concepts in Statistical Physics

- An elegant derivation of the Boltzmann [11] from basic mathematics.
- Phase transitions and critical phenomena [17].
- Non-equilibrium thermodynamics. [15, 16]

General References on Erdös-Renyi (Random) Graphs, Graph Theory

- Many Introduction to Graph Theory texts (Amazon, library)
- Early paper on Random Graphs [7].

Watts-Strogatz (Small World) Networks

Works by Duncan Watts and Steve Strogatz outline the mechanisms for this phenomenon affecting network diameter Watts, Strogatz [20]

Barabasi-Albert (Scale-Free) Networks

A very highly-cited paper that discusses key issues of scale-free networks and power law relationships Barabasi, Albert, 1999 [2].

Complex Systems of Cities, Geographic Networks, Sustainability, etc.

- A provocative work by Bettencourt et al. on allometric scaling of phenomena in cities, including *innovation, GDP, and energy consumption* [5].
- An article about allometric scaling of countries [22].
- A statistical physics argument about the distribution of wealth, energy, and income [21].
- A nice recent review of spatial networks by Marc Barthelemy [4].
- Some discussion of fundamental power law and scale-free phenomena by Michael Mitzenmacher, Mark Newman, and Eugene Stanley [9, 6, 18].
- A fascinating article that develops a *scaling-based sustainability index* [10].

- Language

  Our presentation from Barabasi's course:

  2-grams (word-pairs occurring juxtaposed in Google Books)

  http://www.prezi.com/wuhpytgxc721/exploring-the-english-language-network-with-google-n-grams/

Some Useful, Local, and/or Major Research Groups

- Hidalgo lab
- Santa Fe Institute - also interested in sustainability, cities.
- Mark Newman lab
- Barabasi Lab
- Steve Strogatz
- Duncan Watts
- Max Planck Dresden
- ETHZ - Helbing (FuturICT - sustainability)
- L'Institut des Systèmes Complexes – Paris Île-de-France (ISC-PIF)
- Lyon IXXI
- Sole lab, Barcelona
- IFISC, Majorca (interest in cities)
- NECSI

References

[1] A Banerjee, *The spectrum of the graph Laplacian as a tool for analyzing structure and evolution of networks*, (2008).

[2] AL Barabási and R Albert, *Emergence of scaling in random networks*, science (1999).

[3] A.L. Barabási and J. Frangos, *Linked: The new science of networks science of networks*, Basic Books, 2002.

[4] Marc Barthélemy, *Spatial networks*, Physics Reports **499** (2011), no. 1-3, 1–101.

[5] LMA Bettencourt and J Lobo, *Growth, innovation, scaling, and the pace of life in cities*, Proceedings of the ... (2007).

[6] A Clauset, CR Shalizi, and MEJ Newman, *Power-law distributions in empirical data*, SIAM review (2009).

[7] P Erdos and A Renyi, *On random graphs I.*, Publ. Math. Debrecen (1959).

[8] M. Mitchell, *Complexity: a guided tour*, Oxford University Press, USA, 2009.

[9] Michael Mitzenmacher, *A Brief History of Generative Models for Power Law and Lognormal Distributions*, Internet Mathematics **1** (2004), no. 2, 226–251.

[10] Rachata Muneepeerakul and Murad R. Qubbaj, *The effect of scaling and connection on the sustainability of a socio-economic resource system*, Ecological Economics **77** (2012), 123–128.

[11] L.K. Nash, *Elements of statistical thermodynamics*, Dover publications, 2006.

[12] M. Newman, *Networks: an introduction*, Oxford University Press, Inc., 2010.

[13] M. E. J. Newman, *The Structure and Function of Complex Networks*, SIAM Review **45** (2003), no. 2, 167–256.

[14] M E J Newman, *Power laws, Pareto distributions and Zipf's law*, Contemporary Physics **46** (2005), no. 5, 323–351.

[15] I Prigogine, *Introduction to thermodynamics of irreversible processes*, New York: Interscience, 1967, 3rd ed. (1967).

[16] P. Glansdorff Prigogine and I., *Thermodynamic Theory of Structure Stability and Fluctuations*, John Wiley & Sons, New York, NY, USA, 1971.

[17] H.E. Stanley, *Introduction to phase transitions and critical phenomena*, Introduction to Phase Transitions and Critical Phenomena, by H Eugene Stanley, pp. 336. Foreword by H Eugene Stanley. Oxford University Press, Jul 1987. ISBN-10: 0195053168. ISBN-13: 9780195053166 **1** (1987).

[18] HE Stanley, *Scaling, universality, and renormalization: Three pillars of modern critical phenomena*, Reviews of modern physics **71** (1999), no. 2, S358–S366.

[19] Yogesh Virkar and Aaron Clauset, *Power-law distributions in binned empirical data*, arXiv preprint arXiv:1208.3524 (2012).

[20] D J Watts and S H Strogatz, *Collective dynamics of 'small-world' networks.*, Nature **393** (1998), no. 6684, 440–2.

[21] VM Yakovenko, *Statistical mechanics approach to econophysics*, Encyclopedia of Complexity and System ... (2007), 24.

[22] Jiang Zhang and Tongkui Yu, *Allometric scaling of countries*, Physica A: Statistical Mechanics and its Applications (2010), 1–24.