Properties of Complex Networks

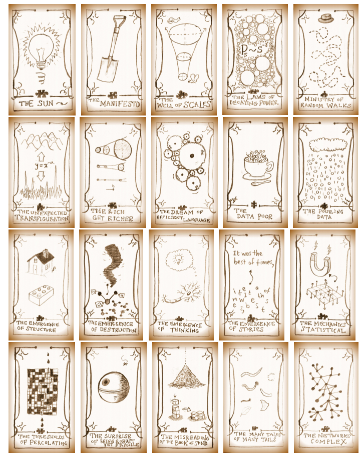
Principles of Complex Systems | @pocsvox
CSYS/MATH 300, Fall, 2016 | #FallPoCS2016

Prof. Peter Dodds | @peterdodds

Dept. of Mathematics & Statistics | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont

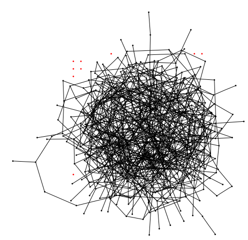
These slides are brought to you by:

Outline



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

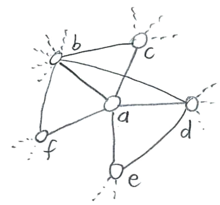
- P_k is the probability that a randomly selected node has degree k.
- k = node degree = number of connections.
- ex 1: Erdős-Rényi random networks have Poisson degree distributions:
  Insert question from assignment 7

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

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

Local socialness:

4. Clustering:



- Your friends tend to know each other.
- Two measures (explained on following slides):
  1. Watts & Strogatz [8]

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

  2. Newman [6]

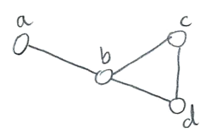
$$C_2 = \frac{3 \times \text{\#triangles}}{\text{\#triples}}$$

Properties

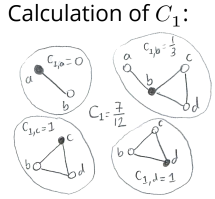
Note:

- 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.

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

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

  where k_i is node i’s degree, and N_i is the set of i’s neighbors.
- Averaging over all nodes, we have:
  $C_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j_1 j_2 \in N_i} a_{j_1 j_2}}{k_i(k_i-1)/2} = \left\langle \frac{\sum_{j_1 j_2 \in N_i} a_{j_1 j_2}}{k_i(k_i-1)/2} \right\rangle_i$

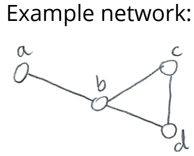
Properties

2. Assortativity/3. Homophily:

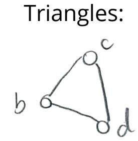
- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network: [5] similar degree nodes connecting to each other.
  Often social: company directors, coauthors, actors.
- Disassortative network: high degree nodes connecting to low degree nodes.
  Often techological or biological: Internet, WWW, protein interactions, neural networks, food webs.

Triples and triangles

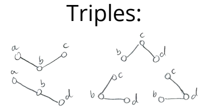
Example network:



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.
- $$\#\text{triples} = \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{\ell=1}^{N}\left[A^2\right]_{i\ell} - \text{Tr}A^2\right)$$
- $$\#\text{triangles} = \frac{1}{6}\text{Tr}A^3$$

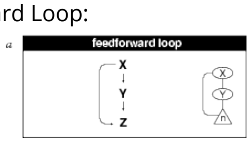
Properties

- For sparse networks, C_1 tends to discount highly connected nodes.
- C_2 is a useful and often preferred variant
- In general, $C_1 \neq C_2$.
- C_1 is a global average of a local ratio.
- C_2 is a ratio of two global quantities.

Properties

5. motifs:

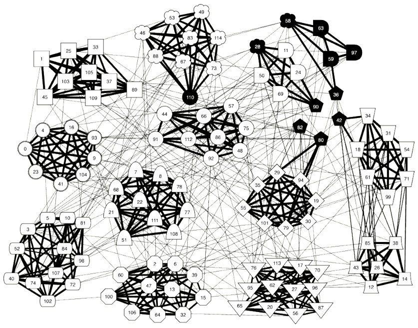
- small, recurring functional subnetworks
- e.g., Feed Forward Loop:



Shen-Orr, Uri Alon, *et al.* [7]

Properties

6. modularity and structure/community detection:



Clauset *et al.*, 2006 [2]: 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 [4]
- "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 [3])

Properties

Interconnected networks and robustness (two for one deal):

"Catastrophic cascade of failures in interdependent networks" [1]. Buldyrev et al., Nature 2010.

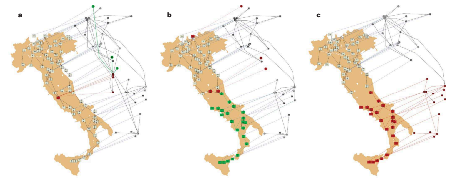


Figure 1 | Modelling a blackout in Italy. Illustration of an iterative process of a cascade of failures using real-world data from a power network (located on the map of Italy) and an Internet network (shifted above the map) that were implicated in an electrical blackout that occurred in Italy in September 2003[20]. The networks are drawn using the real geographical locations and every Internet server is connected to the geographically nearest power station. **a**, One power station is removed (red node on map) from the power network and as a result the Internet nodes depending on it are removed from the Internet network (red nodes above the map). The nodes that will be disconnected from the giant cluster (a cluster that spans the entire network) at the next step are marked in green. **b**, Additional nodes that were disconnected from the Internet communication network giant component are removed (red nodes above map). As a result the power stations depending on them are removed from the power network (red nodes on map). Again, the nodes that will be disconnected from the giant cluster at the next step are marked in green. **c**, Additional nodes that were disconnected from the giant component of the power network are removed (red nodes on map) as well as the nodes in the Internet network that depend on them (red nodes above map).

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 man-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).

scale-free-networks,

References I

[1] 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

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

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

References II

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

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

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

[7] 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 III

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