Overview of Complex Networks
Complex Networks, CSYS/MATH 303, Spring, 2010

Prof. Peter Dodds
Department of Mathematics & Statistics
Center for Complex Systems
Vermont Advanced Computing Center
University of Vermont

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.

Class Admin

- Office hours:
  - Tuesday 1:00 pm–2:30 pm (Farrell Hall)
  - Appointments by email.
- Course outline
- Projects

Outline

- Class admin
- Basic definitions
- Popularity
- Examples of Complex Networks
- Properties of Complex Networks
- Modelling Complex Networks
- Nutshell
- References

Exciting details regarding these slides:

- Three versions (all in pdf):
  1. Presentation,
  2. Flat Presentation,
  3. Handout (2x2).
- Presentation versions are navigable and hyperlinks are clickable.
- Web links look like this (⊞).
- References in slides link to full citation at end.[1]
- Citations contain links to papers in pdf (if available).
- Brought to you by a concoction of \LaTeX, Beamer, and perl.
Basic definitions

Complex System—Some ingredients:

- Distributed system of many interrelated parts
- No centralized control
- Nonlinear relationships
- Existence of feedback loops
- Complex systems are open (out of equilibrium)
- Presence of Memory
- Modular (nested)/multiscale structure
- Opaque boundaries
- Emergence—'More is Different' \[1\]
- Many phenomena can be complex: social, technical, informational, geophysical, meteorological, fluidic, ...

Net•work  ['net,werk]
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 computer stations.
   • [ 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.
From Keith Briggs's excellent etymological investigation:

- **Opus reticulatum:**
- A Latin origin?

Net and Work are venerable old words:
- ‘Net’ first used to mean spider web (King Ælfréd, 888).
- ‘Work’ appears to have long meant purposeful action.
- ‘Network’ = something built based on the idea of natural, flexible lattice or web.
- c.f., ironwork, stonework, fretwork.

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

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 last decade 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, …)
Popularity (according to ISI)

“Collective dynamics of ‘small-world’ networks” [21]

- Watts and Strogatz
- \( \approx 4100 \) citations (as of January 18, 2010)
- Over 1100 citations in 2008 alone.

“Emergence of scaling in random networks” [2]

- Barabási and Albert
  Science, 1999
- \( \approx 4400 \) citations (as of January 18, 2010)
- Over 1100 citations in 2008 alone.

Popularity according to books:

- The Tipping Point: How Little Things can make a Big Difference—Malcolm Gladwell [9]
  Nexus: Small Worlds and the Groundbreaking Science of Networks—Mark Buchanan

Popularity according to 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 [19]

Numerous others:

- Complex Social Networks—F. Vega-Redondo [18]
- 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
- Evolution of Networks—S. N. Dorogovtsev and J. F. F. Mendes [8]
More observations

- But surely networks aren’t new...
- Graph theory is 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.
- Real networks occupy a tiny, low entropy part of all network space and require specific attention.
- A worthy goal: establish mechanistic explanations.
- What kinds of dynamics lead to these real networks? “If this is upsetting, maybe string theory is for you..."

More observations

- Web-scale data sets can be overly exciting.

Witness:

- The End of Theory: The Data Deluge Makes the Scientific Theory Obsolete (Anderson, Wired) [1]
- “The Unreasonable Effectiveness of Data,” Halevy et al. [10]
- c.f. Wigner’s “The Unreasonable Effectiveness of Mathematics in the Natural Sciences” [22]

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,...

Basic definitions

Links = Connections between nodes
- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstract (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.
Basic definitions

Node degree = Number of links per node

- Notation: Node $i$'s degree $= k_i$.
- $k_i = 0, 1, 2, \ldots$
- Notation: the average degree of a network $= \langle k \rangle$ (and sometimes as $z$)
- For undirected networks, connection between number of edges $m$ and average degree:
  $$\langle k \rangle = \frac{2m}{N}$$
- For directed networks,
  $$\langle k_{\text{out}} \rangle = \langle k_{\text{in}} \rangle = \frac{m}{N}$$
- Defn: $\mathcal{N}_i$ = the set of $i$'s $k_i$ neighbors

Basic definitions

Adjacency matrix:

- We represent a graph or network by a matrix $A$ with link weight $a_{ij}$ for nodes $i$ and $j$ in entry $(i,j)$.
- e.g.,
  $$A = \begin{bmatrix}
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}$$
- (n.b., for numerical work, we always use sparse matrices.)

Examples

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
- Road networks
- Power grids
- Distribution (branching) versus redistribution (cyclical)
Examples

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

Examples

Interaction networks: social networks
- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com, facebook.com

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

Examples

Relational networks
- Consumer purchases
  (Wal-Mart: \( \approx 1 \) petabyte \( = 10^{15} \) bytes)
- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging: del.icio.us, flickr

common tags: cloud | list

community daily dictionary education encyclopedia english free imported info information internet knowledge learning news reference research resources search tools useful web web2.0 wiki wikipedia
Clickworthy Science:

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.

Properties

Some key aspects of real complex networks:

- degree distribution
- assortativity
- homophily
- clustering
- motifs
- modularity

- concurrency
- hierarchical scaling
- network distances
- centrality
- efficiency
- robustness

- + Coevolution of network structure and processes on networks.

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:

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

- Distribution is Poisson
Properties

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

Properties

2. assortativity/3. homophily:
   - Social networks: Homophily (をしている) = birds of a feather
   - e.g., degree is standard property for sorting: measure degree-degree correlations.
   - Assortative network: similar degree nodes connecting to each other.
     Often social: company directors, coauthors, actors.
   - Disassortative network: high degree nodes connecting to low degree nodes.
     Often technological or biological: Internet, WWW, protein interactions, neural networks, food webs.

Clustering

4. clustering:
   - Your friends tend to know each other.
   - Two measures:
     1. Watts & Strogatz\textsuperscript{[21]}\[ C_1 = \left( \frac{\sum_{i,j\in N_i} a_{ij}}{k_i(k_i-1)/2} \right)_i \]
     2. Newman\textsuperscript{[14]} \[ C_2 = \frac{3 \times \# \text{triangles}}{\# \text{triples}} \]
Properties

First clustering measure:

▶ $C_1$ is the average fraction of pairs of neighbors who are connected.

▶ Fraction of pairs of neighbors who are connected is

$$\frac{\sum_{j,k \in \mathcal{N}_i} a_{j,k}}{k_i(k_i - 1)/2}$$

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

$$C_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j,k \in \mathcal{N}_i} a_{j,k}}{k_i(k_i - 1)/2} = \left( \frac{\sum_{j,k \in \mathcal{N}_i} a_{j,k}}{k_i(k_i - 1)/2} \right)_i$$

Properties

Triples and triangles

▶ 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**

▶ Social Network Analysis (SNA): fraction of **transitive triples**.

▶ The ‘3’ appears because for each triangle, we have 3 closed triples.

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:

![Feed Forward Loop](image)

Shen-Orr, Uri Alon, *et al.* [16]
Properties

6. modularity and structure/community detection:

Clauset et al., 2006 [7]: NCAA football

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 [12]

8. Horton-Strahler ratios:

- Metrics for branching networks:
  - Method for ordering streams hierarchically
  - Number: \( R_n = N_ω/N_{ω+1} \)
  - Segment length: \( R_l = \langle l_{ω+1} \rangle/\langle l_ω \rangle \)
  - Area/Volume: \( R_a = \langle a_{ω+1} \rangle/\langle a_ω \rangle \)

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

Models

Some important models:
1. generalized random networks (touched on in 300)
2. scale-free networks (covered in 300)
3. small-world networks (covered in 300)
4. statistical generative models ($p^*$)
5. generalized affiliation networks (partly covered in 300)

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 $\ell$’s betweenness = fraction of shortest paths that travel along $\ell$.
   - ex 4: Recursive centrality: Hubs and Authorities (Jon Kleinberg\[11\])
Models

2. ‘scale-free networks’:

- Introduced by Barabasi and Albert \[^{[2]}\]
- Generative model
- Preferential attachment model with growth:
  \(P\) [attachment to node \(i\)] \(\propto k_i^\alpha\).
- Produces \(P_k \sim k^{-\gamma}\) when \(\alpha = 1\).
- Trickiness: other models generate skewed degree distributions.

\(\gamma = 2.5\)
\(\langle k \rangle = 1.8\)
\(N = 150\)

Models

3. small-world networks

- Introduced by Watts and Strogatz \[^{[21]}\]

Two scales:

- local regularity (an individual's friends know each other)
- global randomness (shortcuts).

- Shortcuts allow disease to jump
- Number of infectives increases exponentially in time
- Facilitates synchronization

Models

5. generalized affiliation networks

Bipartite affiliation networks: boards and directors, movies and actors.
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).

Nutshell:

Overview Key Points (cont.):

- Obvious connections with the vast extant field of graph theory.
- But focus on dynamics is more of a physics/stat-mech/comp-sci flavor.
- Two main areas of focus:
  1. Description: Characterizing very large networks
  2. Explanation: Micro story ⇒ Macro features
- Some essential structural aspects are understood: degree distribution, clustering, assortativity, group structure, overall structure,...
- Still much work to be done, especially with respect to dynamics... exciting!

References I

More is different.
Science, 177(4047):393–396, August 1972. pdf

Emergence of scaling in random networks.

Crosscutting Social Circles.

Clickstream data yields high-resolution maps of science.
References II


References III


References IV


References V


