Two Lectures on Networks

Aaron Clauset
@aaronclauset
Associate Professor of Computer Science
University of Colorado Boulder
External Faculty, Santa Fe Institute

lecture 1: what are networks and how do we talk about them?
then (csss 2003)

now (prof. colorado)
what are networks?
what are networks?

• an approach
• a mathematical representation
• provide structure to complexity
• structure above
  individuals / components
• structure below
  system / population

system / population

individuals / components
these lectures

• build intuition
• expose key concepts
• highlight some big questions
• teach a little math
• provide some examples
• give pointers to further study
• prep for other CSSS lectures
• not a substitute for technical coursework

it's a big field now

About 5,890,000 results (0.07 sec)
Network Analysis and Modeling

Instructor: Aaron Clauset

This graduate-level course will examine modern techniques for analyzing and modeling the structure and dynamics of complex networks. The focus will be on statistical algorithms and methods, and both lectures and assignments will emphasize model interpretability and understanding the processes that generate real data. Applications will be drawn from computational biology and computational social science. No biological or social science training is required. (Note: this is not a scientific computing course, but there will be plenty of computing for science.)

Full lectures notes online (~150 pages in PDF)

http://santafe.edu/~aaronc/courses/5352/
Software

- R
- Python
- Matlab
- NetworkX [python]
- graph-tool [python, c++]
- GraphLab [python, c++]

Standalone editors

- UCI-Net
- NodeXL
- Gephi
- Pajek
- Network Workbench
- Cytoscape
- yEd graph editor
- Graphviz

Network data sets

- Colorado Index of Complex Networks

The Colorado Index of Complex Networks (ICON)

ICON is a comprehensive index of research-quality network data sets from all domains of network science, including social, web, information, biological, ecological, connectome, transportation, and technological networks.

Each network record in the index is annotated with and searchable or browsable by its graph properties, description, size, etc., and many records include links to multiple networks. The contents of ICON are curated by volunteer experts from Prof. Aaron Clauset’s research group at the University of Colorado Boulder.

Click on the NETWORKS tab above to get started.
1. defining a network
2. describing a network
3. null models for networks
4. inference and network models
**what is a vertex?**

\[ V \] distinct objects (vertices / nodes / actors)

**when are two vertices connected?**

\[ E \subseteq V \times V \]

pairwise relations (edges / links / ties)
<table>
<thead>
<tr>
<th>network</th>
<th>vertex</th>
<th>edge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet(1)</td>
<td>computer</td>
<td>IP network adjacency</td>
</tr>
<tr>
<td>Internet(2)</td>
<td>autonomous system (ISP)</td>
<td>BGP connection</td>
</tr>
<tr>
<td>software</td>
<td>function</td>
<td>function call</td>
</tr>
<tr>
<td>World Wide Web</td>
<td>web page</td>
<td>hyperlink</td>
</tr>
<tr>
<td>documents</td>
<td>article, patent, or legal case</td>
<td>citation</td>
</tr>
<tr>
<td>power grid transmission</td>
<td>generating or relay station</td>
<td>transmission line</td>
</tr>
<tr>
<td>rail system</td>
<td>rail station</td>
<td>railroad tracks</td>
</tr>
<tr>
<td>road network(1)</td>
<td>intersection</td>
<td>pavement</td>
</tr>
<tr>
<td>road network(2)</td>
<td>named road</td>
<td>intersection</td>
</tr>
<tr>
<td>airport network</td>
<td>airport</td>
<td>non-stop flight</td>
</tr>
<tr>
<td>friendship network</td>
<td>person</td>
<td>friendship</td>
</tr>
<tr>
<td>sexual network</td>
<td>person</td>
<td>intercourse</td>
</tr>
<tr>
<td>metabolic network</td>
<td>metabolite</td>
<td>metabolic reaction</td>
</tr>
<tr>
<td>protein-interaction network</td>
<td>protein</td>
<td>bonding</td>
</tr>
<tr>
<td>gene regulatory network</td>
<td>gene</td>
<td>regulatory effect</td>
</tr>
<tr>
<td>neuronal network</td>
<td>neuron</td>
<td>synapse</td>
</tr>
<tr>
<td>food web</td>
<td>species</td>
<td>predation or resource transfer</td>
</tr>
</tbody>
</table>
social networks

**vertex:** a person

**edge:** friendship, collaborations, sexual contacts, communication, authority, exchange, etc.
information networks

**vertex:** books, blogs, webpages, etc.

**edge:** citations, hyperlinks, recommendations, similarity, etc.
communication networks

**vertex:** network router, ISP, email address, mobile phone number, etc.

**edge:** exchange of information
transportation networks

**vertex:** city, airport, junction, railway station, river confluence, etc.

**edge:** physical transportation of material
biological networks

vertex: species, metabolic, protein, gene, neuron, etc.

edge: predation, chemical reaction, binding, regulation, activation, etc.
representing networks
a simple network

- undirected
- unweighted
- no self-loops
a simple network

adjacency matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

adjacency list

<table>
<thead>
<tr>
<th>A</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>→ {2, 5}</td>
</tr>
<tr>
<td>2</td>
<td>→ {1, 3, 4}</td>
</tr>
<tr>
<td>3</td>
<td>→ {2, 4, 5, 6}</td>
</tr>
<tr>
<td>4</td>
<td>→ {2, 3}</td>
</tr>
<tr>
<td>5</td>
<td>→ {1, 3}</td>
</tr>
<tr>
<td>6</td>
<td>→ {3}</td>
</tr>
</tbody>
</table>

undirected
unweighted
no self-loops
a less simple network

- Directed edge
- Weighted edge
- Multi-edge
- Self-loop
- Weighted node

- undirected
- unweighted
- no self-loops
a less simple network

adjacency matrix

\[
\begin{array}{ccccccc}
A & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
1 & 0 & 0 & 0 & 0 & \{1, 1, 2\} & 0 \\
2 & 1 & \frac{1}{2} & \{2, 1\} & 1 & 0 & 0 \\
3 & 0 & \{2, 1\} & 0 & 2 & 4 & 4 \\
4 & 0 & 1 & 2 & 0 & 0 & 0 \\
5 & \{1, 1, 2\} & 0 & 4 & 0 & 0 & 0 \\
6 & 0 & 0 & 4 & 0 & 0 & 2 \\
\end{array}
\]

adjacency list

\[
A \\
\hline
1 \rightarrow \{(5, 1), (5, 1), (5, 2)\} \\
2 \rightarrow \{(1, 1), (2, \frac{1}{2}), (3, 2), (3, 1), (4, 1)\} \\
3 \rightarrow \{(2, 2), (2, 1), (4, 2), (5, 4), (6, 4)\} \\
4 \rightarrow \{(2, 1), (3, 2)\} \\
5 \rightarrow \{(1, 1), (1, 1), (1, 2), (3, 4)\} \\
6 \rightarrow \{(3, 4), (6, 2)\}
\]
directed networks

\[ A_{ij} \neq A_{ji} \]

citation networks
directed acyclic graph
citation networks
foodwebs*
edemicological
others?
directed graph
WWW
friendship?
flows of goods, information
economic exchange
dominance
neuronal
transcription
time travelers
bipartite networks

no within-type edges

authors & papers  people & locations (checkins)
actors & movies/scenes  metabolites & reactions
musicians & albums  genes & substrings
people & online groups  words & documents
people & corporate boards  plants & pollinators
bipartite networks

- authors & papers
- actors & movies/scenes
- musicians & albums
- people & online groups
- people & corporate boards
- people & locations (checkins)
- metabolites & reactions
- genes & substrings
- words & documents
- plants & pollinators

bipartite network

no within-type edges

one-mode projections

one type only
temporal networks

discrete time (snapshots), edges \((i, j, t)\)
continuous time, edges \((i, j, t_s, \Delta t)\)
describing networks

what networks look like
describing networks

what networks look like

questions:
• how are the edges organized?
• how do vertices differ?
• does network location matter?
• are there underlying patterns?

what we want to know

• what processes shape these networks?
• how can we tell?
describing networks

a first step: describe its features
describing networks

A first step: describe its features

\[ f : G \rightarrow \{ x_1, \ldots, x_k \} \]

- degree distributions
- short-loop density (triangles, etc.)
- shortest paths (diameter, etc.)
- vertex positions
- correlations between these
describing networks

a first step: describe its features

\[ f : \text{object} \rightarrow \{x_1, \ldots, x_k\} \]
describing networks

A first step: describe its features

\[ f : \text{object} \rightarrow \{x_1, \ldots, x_k\} \]

- Physical dimensions
- Material density, composition
- Radius of gyration
- Correlations between these

Helpful for exploration, but not what we want…
What we want: understand its structure

\[ f : \text{object} \rightarrow \{ \theta_1, \ldots, \theta_k \} \]

• What are the fundamental parts?
• How are these parts organized?
• Where are the degrees of freedom \( \vec{\Theta} \)?
• How can we define an abstract class?
• Structure — dynamics — function?

What does local-level structure look like?
What does large-scale structure look like?
How does structure constrain function?
describing networks

degree
degree:
number of connections $k$

$$k_i = \sum_j A_{ij}$$
describing networks

**degree:**
number of connections \( k \)

\[
k_i = \sum_j A_{ij}
\]

number of edges

\[
m = \frac{1}{2} \sum_{i=1}^n k_i = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n A_{ij} = \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n A_{ji}
\]
describing networks

number of edges

\[ m = \frac{1}{2} \sum_{i=1}^{n} k_i = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} = \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} A_{ji} \]

mean degree

\[ \langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{2m}{n} \]

**degree:**

number of connections \( k \)

\[ k_i = \sum_{j} A_{ij} \]
describing networks

degree: number of connections $k$

$$k_i = \sum_j A_{ij}$$

degree sequence \{1, 2, 2, 2, 3, 4\}

degree distribution $Pr(k) = \left[ \left( 1, \frac{1}{6} \right), \left( 2, \frac{3}{6} \right), \left( 3, \frac{1}{6} \right), \left( 4, \frac{1}{6} \right) \right]$
degree distributions

Zachary karate club*

degree distributions

Zachary karate club*

degree distributions

political blogs*

degree distributions

political blogs*

degree distributions

political blogs*

degree "wealth"

what fraction of total wealth $W$ is owned by richest fraction $P$

Lorenz curve
degree distributions

degree "wealth"

what fraction of total wealth $W$ is owned by richest fraction $P$

$$Pr(k) \propto e^{-\lambda k}$$

exponential distribution

Forbes 500 list

Lorenz curve
degree distributions

degree "wealth"

what fraction of total wealth $W$ is owned by richest fraction $P$

$$\Pr(k) \propto \frac{1}{k} e^{-\left(\frac{\ln k - \mu}{\sigma \sqrt{2}}\right)^2}$$

log-normal distribution

Lorenz curve

what fraction of total wealth $W$ is owned by richest fraction $P$
degree distributions

degree "wealth"

what fraction of total wealth $W$ is owned by richest fraction $P$

$$\Pr(k) \propto k^{-\alpha}$$

power-law distribution

80/20 rule

Lorenz curve
degree distributions

is this a power law?

power-law curves

Richest fraction of the population $P$

Fraction of total wealth $W$

$\alpha = 3.5$

$\alpha = 2.7$

$\alpha = 2.4$

$\alpha = 2.2$

$\alpha = 2.1$

political blogs*

fun facts:

- nearly all real networks exhibit a heavy-tailed degree distribution

- very few networks exhibit perfect power-law degree distributions

- some distributions exhibit power-law tails

- power laws are cool! but knowing one from garbage requires statistics

* Clauset et al. SIAM Review 51, 661–703 (2009)
fun facts:

- nearly all real networks exhibit a **heavy-tailed degree distribution**
- **very few** networks exhibit perfect power-law degree distributions
- **some** distributions exhibit power-law tails
- power laws are cool! but knowing one from garbage **requires statistics**

---

degree: number of connections $k$

$$k_i = \sum_j A_{ij}$$

when does node degree matter?
network degrees

spreading processes on networks

biological (diseases)
• SIS and SIR models

social (information)
• SIS, SIR models
• threshold models
network degrees

Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks

Peter S. Bearman  James Moody  Katherine Stovel
Columbia University  Ohio State University University of Washington

• relationship network in “Jefferson High”
• this subgraph is 52% of school
• who are most important disease spreaders?
The Dynamics of Viral Marketing

JURE LESKOVEC  LADA A. ADAMIC  BERNARDO A. HUBERMAN

• amazon.com viral marketing
• viral trace for “Oh my Goddess!” community
• very high degrees!
• most attempts to “influence” fail
network degrees

$R_0 = 0.923 \ldots$

cascade
epidemic
branching process
spreading process

$R_0 = \text{net reproductive rate}$
$= \text{average degree } \langle k \rangle$

caveat:
ignores network structure, dynamics, etc.
Network degrees

$R_0 < 1$
“sub-critical”
small outbreaks

$R_0 = 1$
“critical”
outbreaks of all sizes

$R_0 > 1$
“super-critical”
global epidemics
<table>
<thead>
<tr>
<th>Disease</th>
<th>R0</th>
<th>Vaccination minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measles</td>
<td>5-18</td>
<td>90-95%</td>
</tr>
<tr>
<td>Chicken pox</td>
<td>7-12</td>
<td>85-90%</td>
</tr>
<tr>
<td>Polio</td>
<td>5-7</td>
<td>82-87%</td>
</tr>
<tr>
<td>Smallpox</td>
<td>1.5-20+</td>
<td>70-80%</td>
</tr>
<tr>
<td>H1N1 influenza</td>
<td>1.0-3.0</td>
<td></td>
</tr>
</tbody>
</table>

Data from Lauren Ancel Meyers (UT Austin)
<table>
<thead>
<tr>
<th>Disease</th>
<th>$R_0$</th>
<th>Vaccination Minimum (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measles</td>
<td>5-18</td>
<td>90-95%</td>
</tr>
<tr>
<td>Chicken Pox</td>
<td>7-12</td>
<td>85-90%</td>
</tr>
<tr>
<td>Polio</td>
<td>5-7</td>
<td>82-87%</td>
</tr>
<tr>
<td>Smallpox</td>
<td>1.5-20+</td>
<td>70-80%</td>
</tr>
<tr>
<td>H1N1 Influenza</td>
<td>1.0-3.0</td>
<td></td>
</tr>
</tbody>
</table>

data from Lauren Ancel Meyers (UT Austin)
network degrees

bigger cascades

• smaller overlap among neighbors
• more expander-like
  [more like a random graph]
• higher transmission probability
• lower activation threshold

smaller cascades

• larger overlap among neighbors
• more triangles
• smaller "communities"
• more spatial-like organization
• lower transmission probability
• higher activation threshold

how could we halt the spread?
- break network into disconnected pieces
network degrees

Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong & Albert-László Barabási

2000

homogeneous in degree

heterogeneous in degree

two networks
network degrees

two networks
strategy: delete vertices
1. uniformly at random (“failure”)
2. in order of degree (“attack”)

network degrees
network degrees

what promotes spreading?

- high-degree vertices*
- centrally-located vertices

homogeneous in degree

heterogeneous in degree
network degrees

strategy: delete vertices

3. build “fire breaks”

patient 0

vaccinated = deleted
(“fire break”
**network degrees**

- **vaccination strategies**
  - the “front line” (hospitals)
  - high degree nodes
  - the vulnerable (old/young)

**Diagram:**
- **patient 0**
- **effective buffer**
The vaccination "game"

- by Salathe group (EPFL)
network degrees

but, in social networks...
network degrees

Influentials, Networks, and Public Opinion Formation

DUNCAN J. WATTS
PETER SHERIDAN DODDS*

2007

- classic information marketing
- message saturation
- degree is most important

broadcast influence
network degrees

Influentials, Networks, and Public Opinion Formation

DUNCAN J. WATTS
PETER SHERIDAN DODDS*

2007

• “network” (decentralized) marketing
• high-degree = “opinion leader”
• high-degree alone = irrelevant
• a cascade requires a legion of susceptibles (a system-level property)
“The only thing worse than being talked about is not being talked about.”

- "influence" not really about the influencer
- as much about the susceptibles
network degrees

degrees:
• *first-order* description of network structure
• direct implications for spreading processes
• cascades require both susceptible population *and* spreaders

open questions:
• impact of degrees on other dynamics
• feedback from dynamics to degree
  [adaptive behaviors like self-quarantine, evangelism]
• when does degree *not* matter
end of lecture 1

lecture 2: describing network structure

lecture 3: null models for networks

lecture 4: inference and network models