Introduction to Networks

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

- an approach
- a mathematical representation
- provide structure to complexity
- *structure above* individuals / components
- *structure below* system / population
this lecture

- build intuition
- highlight a few concepts & questions
- provide some examples
- pointers to further study
- not a substitute for technical coursework

it's a big field now
Mark Newman
Professor of Physics
University of Michigan

External Faculty
Santa Fe Institute

http://www-personal.umich.edu/~mejn/
Network Analysis and Modeling

Instructor: Aaron Clauset or Daniel Larremore

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 networks, including social, web, information, biological, ecological, connectome, transportation, and technology networks.

Each network record in the index is annotated with and searchable or browsable by its graph parameters, 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.

Entries found: 609  Networks found: 4419
the two most fundamental questions in network science
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

dge: 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.

graphical representation:

**core metabolism**

[Diagram of core metabolism with annotations and nodes representing various biochemical processes and compounds.]

**grassland foodweb**

[Diagram of a grassland foodweb with interconnected nodes and edges illustrating the flow of energy and matter.]
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

<p>| | |</p>
<table>
<thead>
<tr>
<th></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
- Weighted node
- Self-loop

- undirected
- unweighted
- no self-loops
a less 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>0</td>
<td>0</td>
<td>0</td>
<td>{1, 1, 2}</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1/2</td>
<td>{2, 1}</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>{2, 1}</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>{1, 1, 2}</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

**adjacency list**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{ (5, 1), (5, 1), (5, 2) }</td>
</tr>
<tr>
<td>2</td>
<td>{ (1, 1), (2, 1/2), (3, 2), (3, 1), (4, 1) }</td>
</tr>
<tr>
<td>3</td>
<td>{ (2, 2), (2, 1), (4, 2), (5, 4), (6, 4) }</td>
</tr>
<tr>
<td>4</td>
<td>{ (2, 1), (3, 2) }</td>
</tr>
<tr>
<td>5</td>
<td>{ (1, 1), (1, 1), (1, 2), (3, 4) }</td>
</tr>
<tr>
<td>6</td>
<td>{ (3, 4), (6, 2) }</td>
</tr>
</tbody>
</table>
directed networks

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

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

no within-type edges

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 networks

bipartite network

no within-type edges

one-mode projections

one type only

authors & papers
actors & movies/scenes
musicians & albums
people & online groups
people & corporate boards

people & locations (checkins)
metabolites & reactions
genomes & substrings
words & documents
plants & pollinators
temporal networks

any network over time; comes in two flavors

1. discrete time (snapshots), edges \((i, j, t)\)
2. continuous time, edges \((i, j, t_s, \Delta t)\)

physical proximity over time
transportation connections over time
social interactions over time

**multiplex or multilayer networks**

multiple network "layers"

- each layer has same set of nodes $V$
- but different sets of edges $\{E_1, E_2, \ldots, E_\ell\}$

- different types of transportation within a city
- different types of social interactions (trust, socializing, co-located, etc.)
- interactions on different social media platforms

---

analyzing networks

what real networks look like…
analyzing networks

what real 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 does network shape dynamics?
• how can we tell?
analyzing networks

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 $\hat{\theta}$?
- 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?
6 major approaches

1. **exploratory data analysis**: count & compare all the things (degree distributions, centrality scores, community detection, etc.)
6 major approaches

1. **exploratory data analysis**: count & compare all the things (degree distributions, centrality scores, community detection, etc.)

2. **simple regressions**: convert network structure into node-level features, and do traditional explanatory modeling
analyzing networks

6 major approaches

1. **exploratory data analysis**: count & compare all the things (degree distributions, centrality scores, community detection, etc.)

2. **simple regressions**: convert network structure into node-level features, and do traditional explanatory modeling

3. **null models**: use some kind of random graph to identify non-random patterns as deviations from the null
analyzing networks

6 major approaches

1. **exploratory data analysis:** count & compare all the things (degree distributions, centrality scores, community detection, etc.)

2. **simple regressions:** convert network structure into node-level features, and do traditional explanatory modeling

3. **null models:** use some kind of random graph to identify non-random patterns as deviations from the null

4. **mechanisms / simulations:** explain structural or dynamical patterns as caused by specific process
analyzing networks

6 major approaches

1. **exploratory data analysis**: count & compare all the things (degree distributions, centrality scores, community detection, etc.)

2. **simple regressions**: convert network structure into node-level features, and do traditional explanatory modeling

3. **null models**: use some kind of random graph to identify non-random patterns as deviations from the null

4. **mechanisms / simulations**: explain structural or dynamical patterns as caused by specific process

5. **predictive models**: fit parametric model of network structure & use it to predict missing or future data (edges, labels, etc.)
analyzing networks

6 major approaches

1. **exploratory data analysis**: count & compare all the things (degree distributions, centrality scores, community detection, etc.)

2. **simple regressions**: convert network structure into node-level features, and do traditional explanatory modeling

3. **null models**: use some kind of random graph to identify non-random patterns as deviations from the null

4. **mechanisms / simulations**: explain structural or dynamical patterns as caused by specific process

5. **predictive models**: fit parametric model of network structure & use it to predict missing or future data (edges, labels, etc.)

6. **network experiments**: manipulate structure and measure node-level or graph-level behavior as function of changes
analyzing networks

• degrees & distributions
• network position & centrality scores
• some applications
degree distributions
degree distributions

degree:
number of connections $k$

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

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: number of connections \(k\)

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

political blogs

Adamic & Glance. WWW Workshop on the Weblogging Ecosystem (2005)
degree distributions

political blogs

Adamic & Glance, WWW Workshop on the Weblogging Ecosystem (2005)
degree distributions

Adamic & Glance, WWW Workshop on the Weblogging Ecosystem (2005)
degree distributions

is this a power law?

political blogs
fun facts 1:

- 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 identifying them in data (and not confusing them for other things) requires statistics
fun facts 2:

- **degree distribution** is the first-order description of network structure.

- **degree heterogeneity** alone drives interesting phenomena ("friendship paradox", spreading dynamics, etc.)

- **degree heterogeneity** alone explains many other network patterns (various centralities, disassortativity, etc.)

- the **configuration model** is how to tell: random graph model with specified degree sequence.

---

Fosdick et al., SIAM Review 60(2), 315-355 (2018)
network position

position
network position

position = centrality:
structural vs. dynamical importance

harmonic centrality

closeness centrality

betweenness centrality

degree centrality

eigenvector centrality

PageRank

Katz centrality

many many more…

structural importance = cheap estimate of dynamical importance (aka "influence")

Borgatti, Social Networks 27, 55–71 (2005)
position = centrality:
structural vs. dynamical importance

centrality = unsupervised node ranking

\[ f : G \rightarrow \vec{v} \]

there are an infinite number of choices of \( f \)!
most are correlated
choose \( f \) that is most meaningful for downstream analysis
network position

position = centrality:
harmonic, closeness centrality

importance = being in “center” of the network

\[ c_i = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}} \]

distance: \( d_{ij} = \begin{cases} l_{ij} & \text{if } j \text{ reachable from } i \\ \infty & \text{otherwise} \end{cases} \)

Borgatti, Social Networks 27, 55–71 (2005)
network position

position = centrality:
PageRank, Katz, eigenvector centrality

importance = sum of importances of nodes that point at you*

\[ I_i = \sum_{j \rightarrow i} \frac{I_j}{k_j} \]

or, the right eigenvector of

\[ Ax = \lambda x \]

*modulo several technical details
Robust Action and the Rise of the Medici, 1400–1434

John F. Padgett and Christopher K. Ansell

University of Chicago
nodes: Florence families
edges: inter-family marriages

which family is most central?
network position

nodes: Florence families
edges: inter-family marriages

which family is most central? Medici.

\[
C_{\text{Medici}} = 6 \left( \frac{1}{1} \right) + 5 \left( \frac{1}{2} \right) + 3 \left( \frac{1}{3} \right)
\]

\[
= 9.5
\]
network position

actually, it’s complicated…

[1] Marriage edges were only one type of inter-family interaction; hence, centrality on them alone is a simplification, and the deeper questions are about dynamics (how did the full network assemble over time?) and about function (was network position causally related to Medici dominance?).
let’s apply these network concepts
Systematic inequality and hierarchy in faculty hiring networks

Aaron Clauset,1,2,3* Samuel Arbesman,4 Daniel B. Larremore5,6

faculty market is a network

- vertices are PhD-granting universities
- consumers ← producers
- \( u \) hires from \( u \), add an edge \( u \rightarrow v \)
faculty market is a network

- vertices are PhD-granting universities
- consumers $\leftrightarrow$ producers
- $v$ hires from $u$, add an edge $u \rightarrow v$

[1] actual exchanges of 267 faculty among 10 elite CS departments, from our 2011 data, without self-hires
collecting the data

collecting the data

collecting the data

collecting the data

collecting the data

collecting the data

complete, hand-curated data for 19,000 tenure-track faculty across 461 departments in

- Computer Science (205 depts)
- Business (112)
- History (144)

roughly 5000 hours of manual data collection

[2] all data from public sources, mainly faculty CVs and homepages
[3] data collected by a team of 12 students over 3 years, using a random 20% re-collection protocol, with script-based post-processing to detect errors & inconsistencies, which were then corrected by hand
collecting the data

complete, hand-curated data for 19,000 tenure-track faculty across 461 departments in

- Computer Science (205 depts)
- Business (112)
- History (144)

roughly 5000 hours of manual data collection

[2] all data from public sources, mainly faculty CVs and homepages
[3] data collected by a team of 12 students over 3 years, using a random 20% re-collection protocol, with script-based post-processing to detect errors & inconsistencies, which were then corrected by hand
## what’s in the data

<table>
<thead>
<tr>
<th></th>
<th>Computer Science</th>
<th>Business</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>institutions</td>
<td>205</td>
<td>112</td>
<td>144</td>
</tr>
<tr>
<td>tenure-track faculty</td>
<td>5032</td>
<td>9336</td>
<td>4556</td>
</tr>
<tr>
<td>mean size</td>
<td>25</td>
<td>83</td>
<td>32</td>
</tr>
</tbody>
</table>

\[ \sum = 18,924 \]
## what’s in the data

<table>
<thead>
<tr>
<th></th>
<th>Computer Science</th>
<th>Business</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>institutions</td>
<td>205</td>
<td>112</td>
<td>144</td>
</tr>
<tr>
<td>tenure-track faculty</td>
<td>5032</td>
<td>9336</td>
<td>4556</td>
</tr>
<tr>
<td>mean size</td>
<td>25</td>
<td>83</td>
<td>32</td>
</tr>
</tbody>
</table>

- **Full Professors**
  - **Computer Science**: 2400 (48%)
  - **Business**: 4294 (46%)
  - **History**: 2097 (46%)

- **Associate Prof.**
  - **Computer Science**: 1772 (35%)
  - **Business**: 2521 (27%)
  - **History**: 1611 (35%)

- **Assistant Prof.**
  - **Computer Science**: 860 (17%)
  - **Business**: 2521 (27%)
  - **History**: 848 (19%)

- **Female**
  - **Computer Science**: 15%
  - **Business**: 22%
  - **History**: 36%

\[ \sum = 18,924 \]
### what’s in the data

<table>
<thead>
<tr>
<th></th>
<th>Computer Science</th>
<th>Business</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>institutions</td>
<td>205</td>
<td>112</td>
<td>144</td>
</tr>
<tr>
<td>tenure-track faculty</td>
<td>5032</td>
<td>9336</td>
<td>4556</td>
</tr>
<tr>
<td>mean size</td>
<td>25</td>
<td>83</td>
<td>32</td>
</tr>
<tr>
<td>Full Professors</td>
<td>2400 (48%)</td>
<td>4294 (46%)</td>
<td>2097 (46%)</td>
</tr>
<tr>
<td>Associate Prof.</td>
<td>1772 (35%)</td>
<td>2521 (27%)</td>
<td>1611 (35%)</td>
</tr>
<tr>
<td>Assistant Prof.</td>
<td>860 (17%)</td>
<td>2521 (27%)</td>
<td>848 (19%)</td>
</tr>
<tr>
<td>female</td>
<td>15%</td>
<td>22%</td>
<td>36%</td>
</tr>
<tr>
<td>PhDs in-sample</td>
<td>87%</td>
<td>84%</td>
<td>89%</td>
</tr>
</tbody>
</table>

∑ = 18,924

Nearly closed hiring systems
explore the data yourself:
http://danlarrremore.com/faculty/
huge inequalities in faculty production
huge inequalities in faculty production

Gini coefficients (out-degree)
- 0.69, 0.62, 0.72

50% of faculty from
- 18, 16, 8 universities

net producers $k_{\text{out}}/k_{\text{in}} > 1$
- 24%, 36%, 18%

1-10 producers vs.
- 11-20 : 1.6, 2.1, 3.0x more
- 21-30 : 3.1, 2.3, 5.6x more

[2] U.S. Income Gini coefficient = 0.45
a prestige hierarchy

- difficult to talk about inequalities in academia without talking about rankings (centralities!)
- let’s extract a data-driven ranking from the network
a prestige hierarchy
a prestige hierarchy

- select permutation (a ranking) $\pi$ that minimizes the number of "rank violations" : edges $(u, v)$ where $\pi_v < \pi_u$
- higher-ranked nodes have greater "placement power"
- equivalent to minimum feedback arc set problem (NP-hard)

[1] these "MVR"s have a deep history in social theory for extracting dominance or prestige hierarchies from data, especially in animal behavior
[2] MFAS: find the set of arcs of minimum cardinality whose removal converts a directed graph $G$ into a directed acyclic graph
[3] there are many equivalent MVRs for our network. we sample these using a zero-temp MCMC, and average across them to obtain $\langle \pi \rangle$
simple zero-temperature MCMC sampler:

- given an ordering $\pi$ with $\psi(\pi, A)$ rank violations on network $A$

- repeat *ad infinitum*: choose a pair $(u, v)$, swap their ranks $\pi_u \leftrightarrow \pi_v$ to obtain $\pi'$, compute $\psi(\pi', A)$, accept change if $\psi(\pi', A) \geq \psi(\pi, A)$

- for instance:
simple zero-temperature MCMC sampler:

- given an ordering $\pi$ with $\psi(\pi, A)$ rank violations on network $A$

- repeat ad infinitum: choose a pair $(u, v)$, swap their ranks $\pi_u \leftrightarrow \pi_v$ to obtain $\pi'$, compute $\psi(\pi', A)$, accept change if $\psi(\pi', A) \geq \psi(\pi, A)$

- for instance:
simple zero-temperature MCMC sampler:

- given an ordering $\pi$ with $\psi(\pi, A)$ rank violations on network $A$

- repeat *ad infinitum*: choose a pair $(u, v)$, swap their ranks $\pi_u \leftrightarrow \pi_v$ to obtain $\pi'$, compute $\psi(\pi', A)$, accept change if $\psi(\pi', A) \geq \psi(\pi, A)$

- for instance:
a prestige hierarchy

• what do these prestige hierarchies look like?
• what do they tell us about the structure of faculty hiring?
• what predicts placement?
prestige rankings correlate with USNews and NRC

- here, prestige $\pi$ quantifies *placement power*
- uncertainty increases as prestige decreases
- similar results, but different orderings for Business and History

[1] uncertainties derived from the set of MVRs
degrees alone do not explain all of the hierarchy

- use configuration model with observed degree sequence
- extract MVRs for random graphs & measure fraction of unviolated edges
- compare that null distribution to empirical fraction
- the gap is prestige effect beyond faculty production alone
most placements are down the hierarchy
most placements are down the hierarchy

- down : 88%, 86%, 91%
- up : 12%, 14%, 9%
- $\langle \Delta \pi \rangle = 47, 27, 42$ steps down
- CS: top 15% of departments produce 68% of their own faculty and hire 7% from outside top 25% of departments

what predicts placement?

• compare 10 node-level features ("importances"):  
  prestige  
  US News rank  
  NRC rank  
  out-degree  
  in-degree  
  out/in degree  
  eigenvector centrality  
  harmonic centrality  
  closeness centrality  
  random
what predicts placement?

- prestige best single predictor in all 3 fields
- order of other features varies by field
- AUCs all below 0.67 = plenty of room for improvement
prestige correlates with network position

- core and periphery
prestige correlates with network position

- core and periphery
Prestige correlates with network position

- Core and periphery: homeland and colonies
- Prestige is influence, via doctoral placement, over research agendas, research communities, and departmental norms across the discipline.
inequality and prestige hierarchies

- prestige is *influence*, encoded in faculty hiring network
- faculty flow out of network core, into periphery ("the colonies")
- small fraction stay inside core
- only ~10% of hires flow "upstream"

future work

- how to measure cultural influence of core departments?
- what is different about "upstream" hires?
- what role for other inequalities: gender, ethnicity/race, SES, neighborhood effects, productivity, etc.?
conclusions and outlook
conclusions and outlook

networks are cool

[ obviously, right? ]
networks are cool
[ obviously, right? ]

powerful window into structure of complex systems
[ structure + dynamics = function ]

network methods for exploiting rich data
[ connectivity + node annotations + edge weights + temporal information | link or label prediction | etc. ]

abundance of interesting science applications
[ dynamics of social influence | emergence of hierarchy | online social network assembly | etc. ]

but be careful: network methods have fundamental limits:
  • networks are themselves a model of underlying system
  • centralities and community detection typically unsupervised
  • some supervision & auxiliary data = better inferences / predictions
  • formulate your mechanism in terms of nodes, edges, and attributes
  • have fun!
Network Analysis and Modeling

Instructor: Aaron Clauset or Daniel Larremore

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/