

Network Analysis & Modeling

lecture 0: what are networks and how do we talk about them?

Prof. Aaron Clauset
Computer Science & BioFrontiers Institute
@aaronclauset
aaron.clauset@colorado.edu

who are network scientists?

Physicists

Computer Scientists

Applied Mathematicians

Statisticians

Biologists

Ecologists

Sociologists

Political Scientists



it's a big community!

- different *traditions*
- different *tools*
- different *questions*

increasingly, not ONE community,
but MANY, only loosely interacting communities

who are network scientists?

Physicists

Computer Scientists

Applied Mathematicians

Statisticians

Biologists

Ecologists

Sociologists

Political Scientists

} phase transitions, universality
data / algorithm oriented, predictions
dynamical systems, diff. eq.
inference, consistency, covariates
experiments, causality, molecules
observation, experiments, species
individuals, differences, causality
rationality, influence, conflict

what are networks?

an approach.

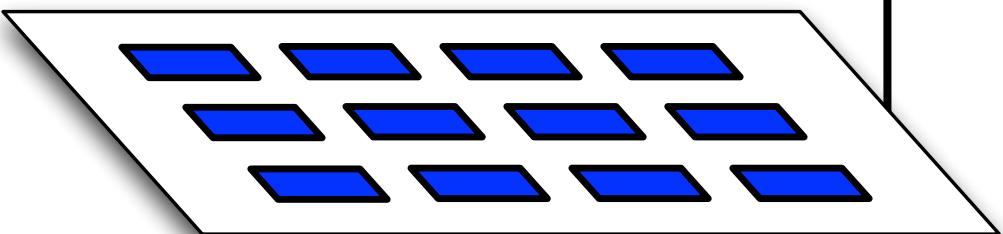
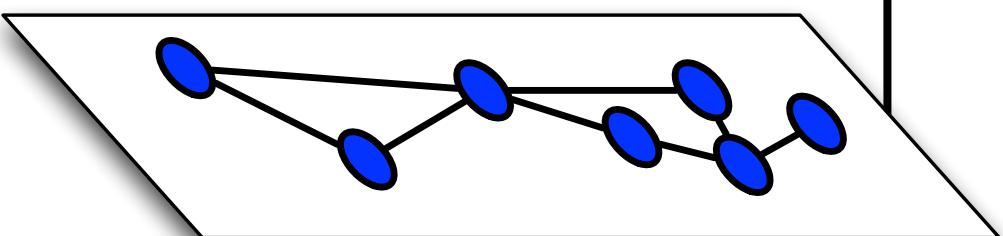
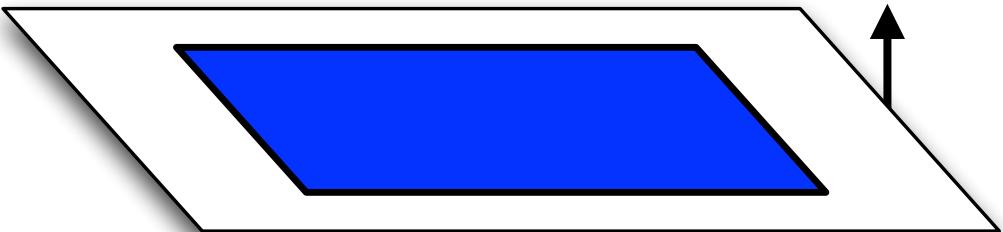
a mathematical representation
provide structure to complexity.

structure that exists *above*
individuals / components

or: structure that exists *below*
system / population



system / population



individuals / components

tools and resources

Software

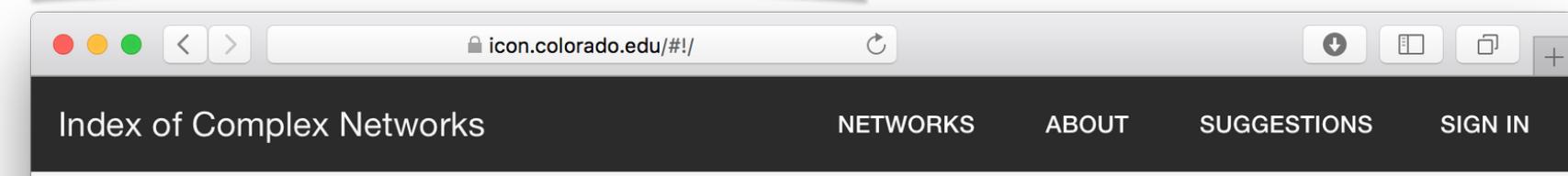
[R](#)
[Python](#)
[Matlab](#)
[NetworkX \[python\]](#)
[graph-tool \[python, c++\]](#)
[GraphLab \[python, c++\]](#)

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.

Entries found: 609 Networks found: 4419



learning goals

1. develop a **network intuition** for reasoning about network phenomena
2. understand **network representations, basic terminology, and concepts.**
3. learn principles and methods for **describing and clustering network data**
4. learn to **predict missing network information**
5. understand how to **conduct and interpret** numerical network experiments,
to **explore and test hypotheses** about networks
6. analyze and model **real-world network data**, using math and computation

course format

- course meets in-person in ECEE 283 + over Zoom
- lectures 2 times a week, some guest lectures and some class discussions
- biweekly problem sets (6 total)
- class project: proposal, presentation, final report
- all content via class Canvas (lecture notes, recordings, problem sets, submissions)
- **see syllabus** for all course policies

course schedule

week by week

- ◆ ● 1. fundamentals of networks
- ◆ ● ★ 2. representations and summary statistics
- ◆ ● 3. simple random graphs
- ◆ ★ 4. better random graphs
- ★ 5. predicting missing node attributes
- ★ ■ 6. predicting missing links
- ◆ ● 7. community structure and mixing patterns
- ★ ■ 8. community structure models
- ◆ ● 9. spreading processes and cascades
- ◆ ● ■ 10. spreading processes with structure (epidemics)
- ◆ ★ 11. data incompleteness and sampling
- ★ ■ 12. ranking in networks
- ◆ ● 13. ethics and networks
- 14. student project presentations

- ◆ building intuition
- basic concepts, tools
- ★ practical tools
- advanced tools

lessons learned

what's difficult

1. students need to know many different things:

some probability

some mathematics

some statistics

some machine learning

some programming

Erdos-Renyi, configuration, calculations

physics-style calculations, phase transitions

basic data analysis, correlations, distributions

prediction, likelihoods, features, estimation algorithms

data wrangling, coding up measures and algorithms

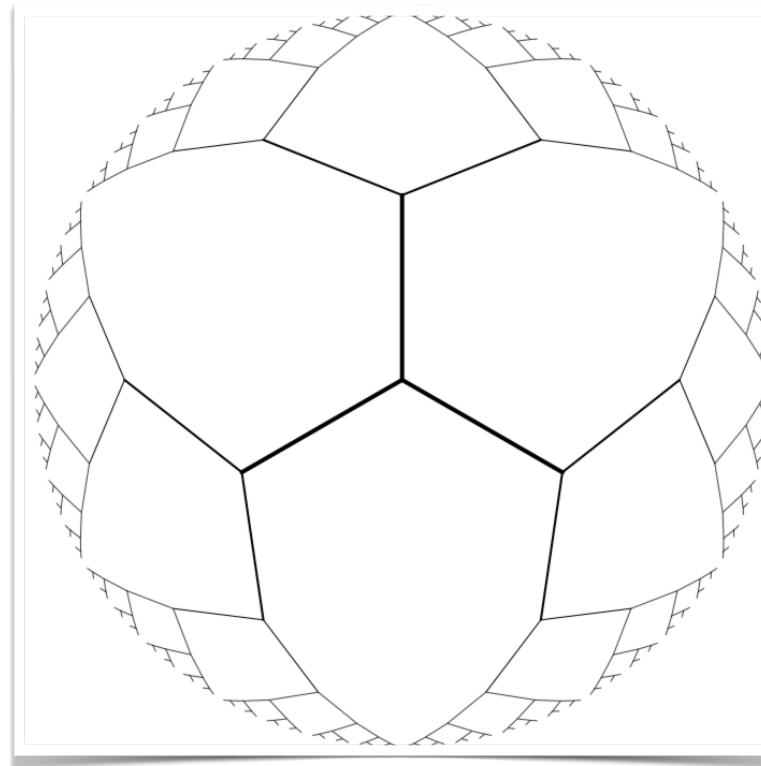
2. can't teach all of these things to all types of students!

- vast amounts of advanced material in each of these directions
- students have little experience / intuition of what makes good science

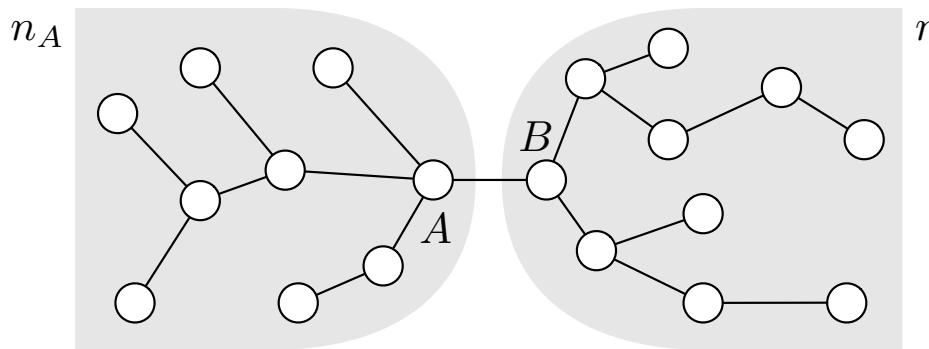
lessons learned

what works well

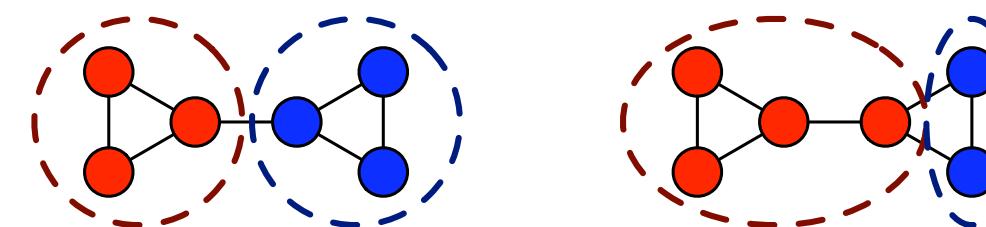
1. simple mathematical problems—build intuition & practice with concepts



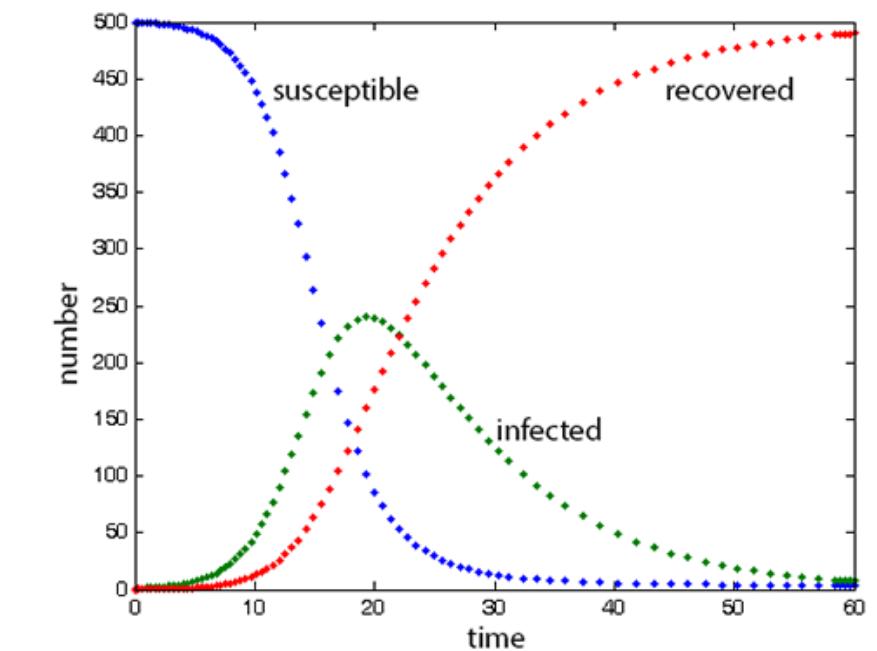
calculate summary statistics



derive mathematical relations



clustering highly-structured networks

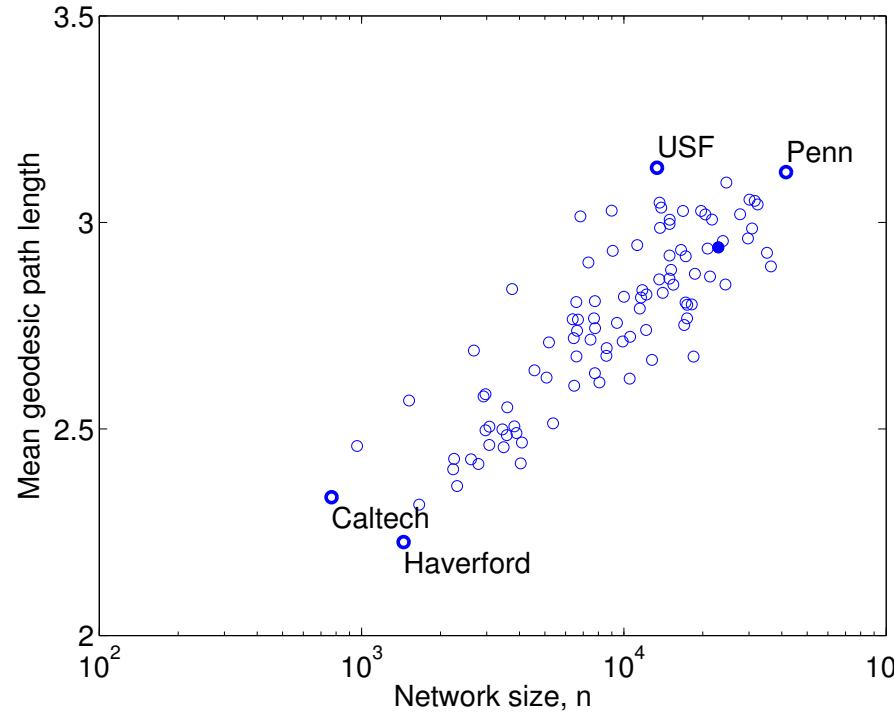


spreading process on networks

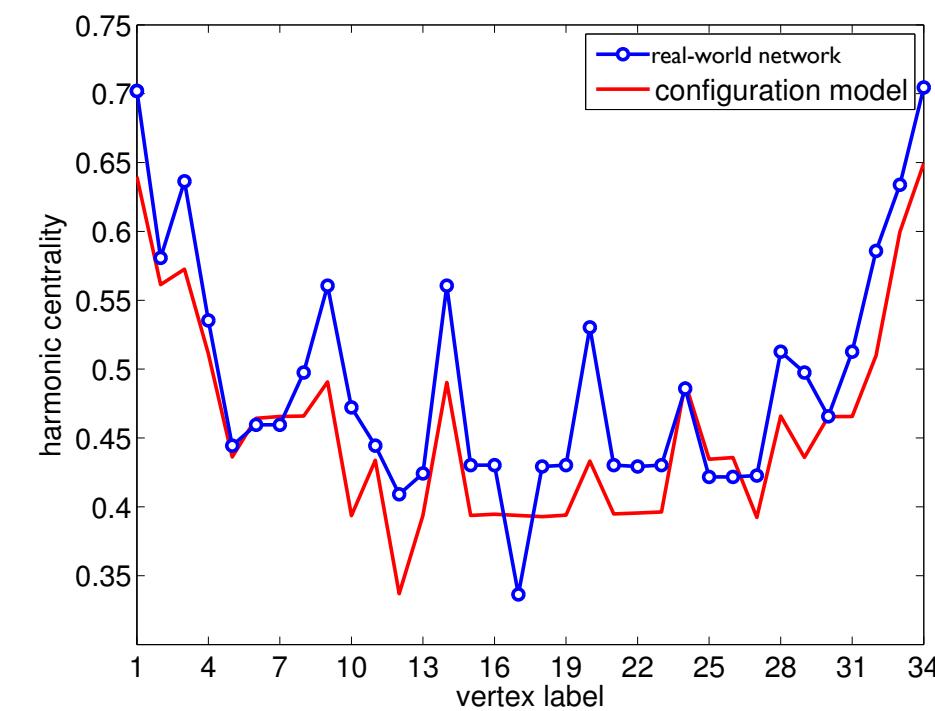
lessons learned

what works well

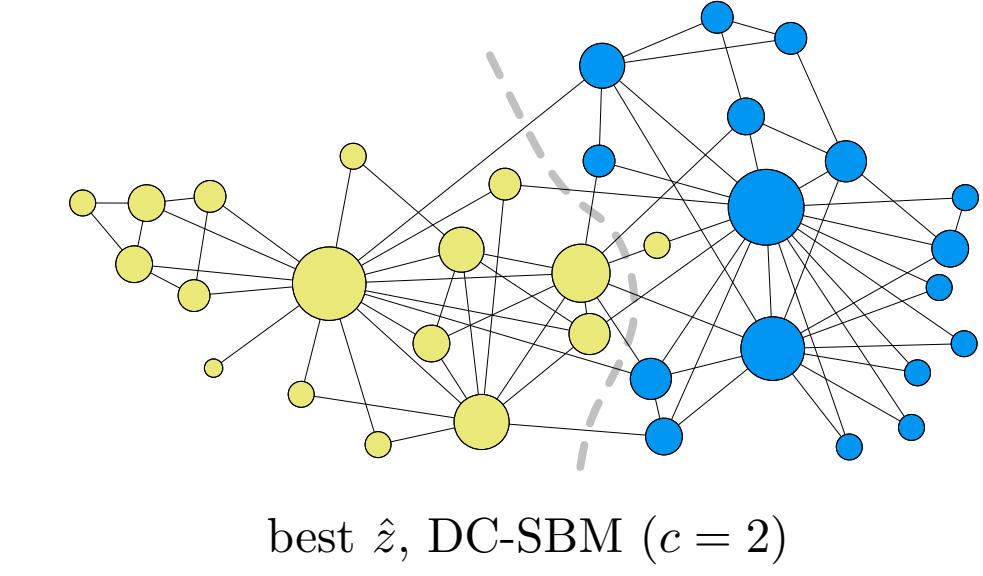
2. analyze *real* networks—test understanding & practice with *implementing* methods



what patterns really occur?



how much does randomness explain?
(when is a pattern *interesting*?)

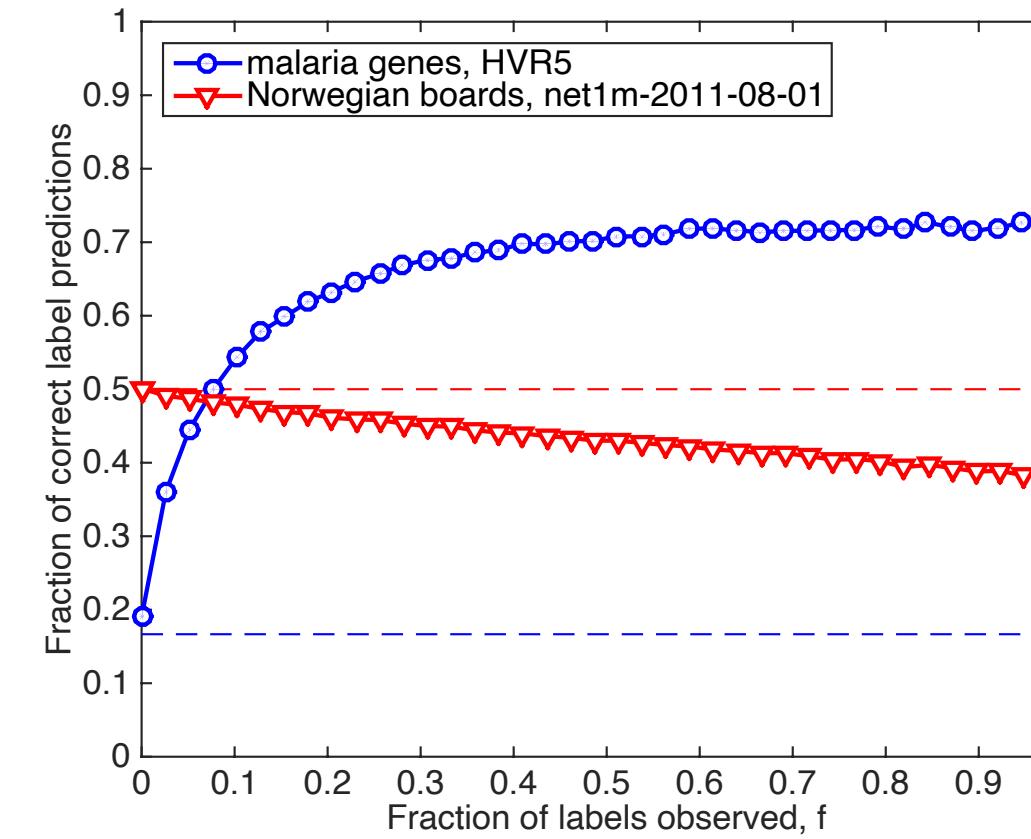


stochastic block models

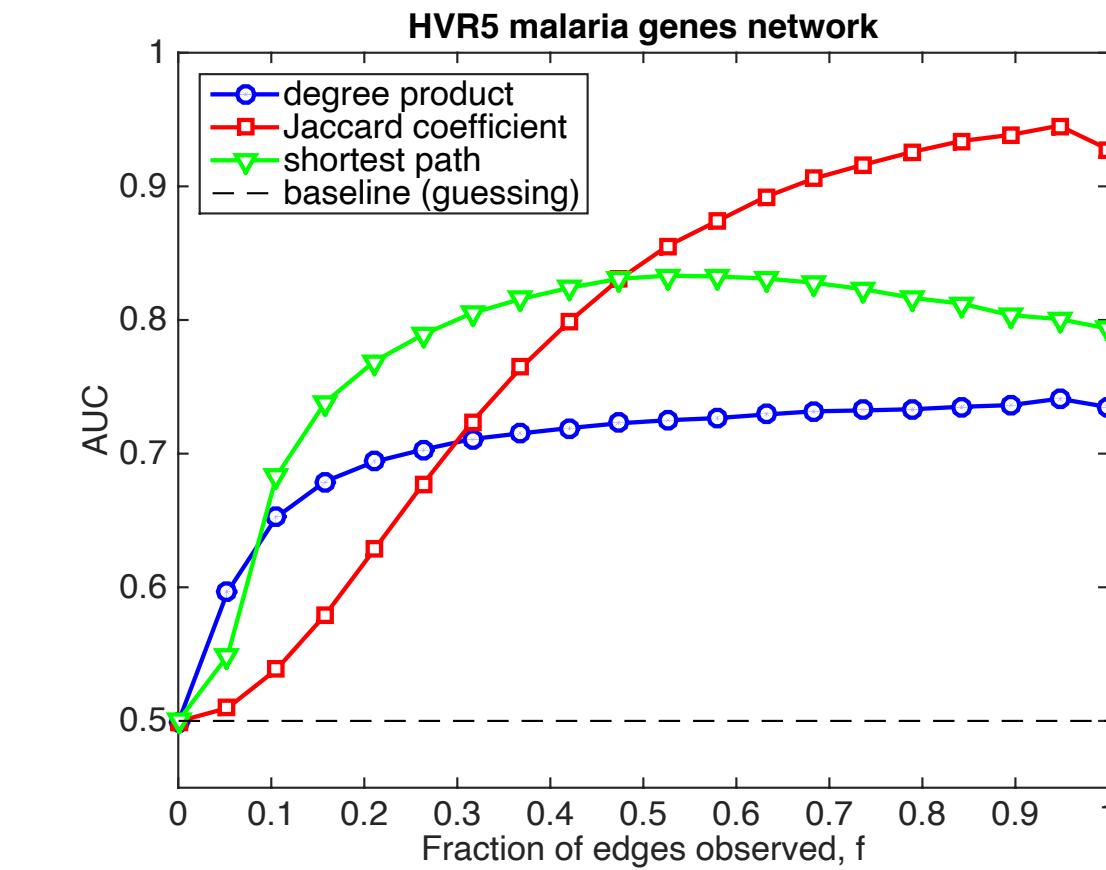
lessons learned

what works well

3. simple prediction tasks—test intuition & run numerical experiments



label prediction via homophily

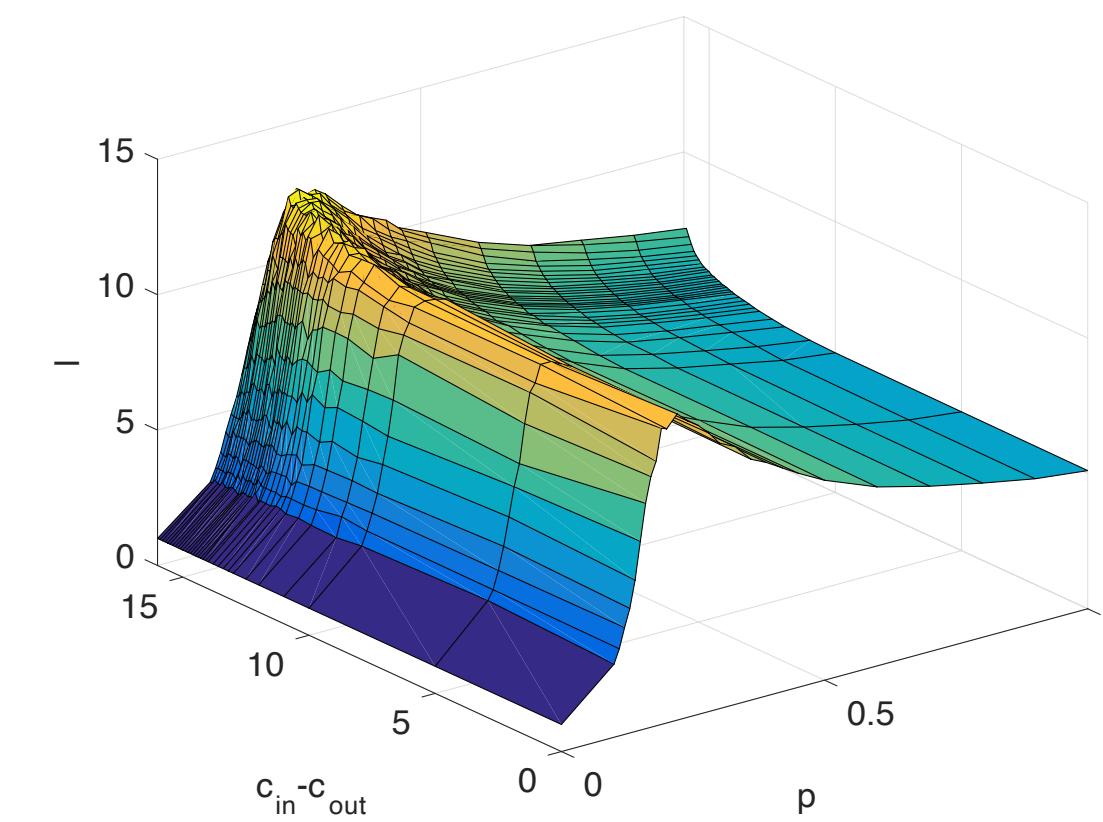


link prediction via heuristic

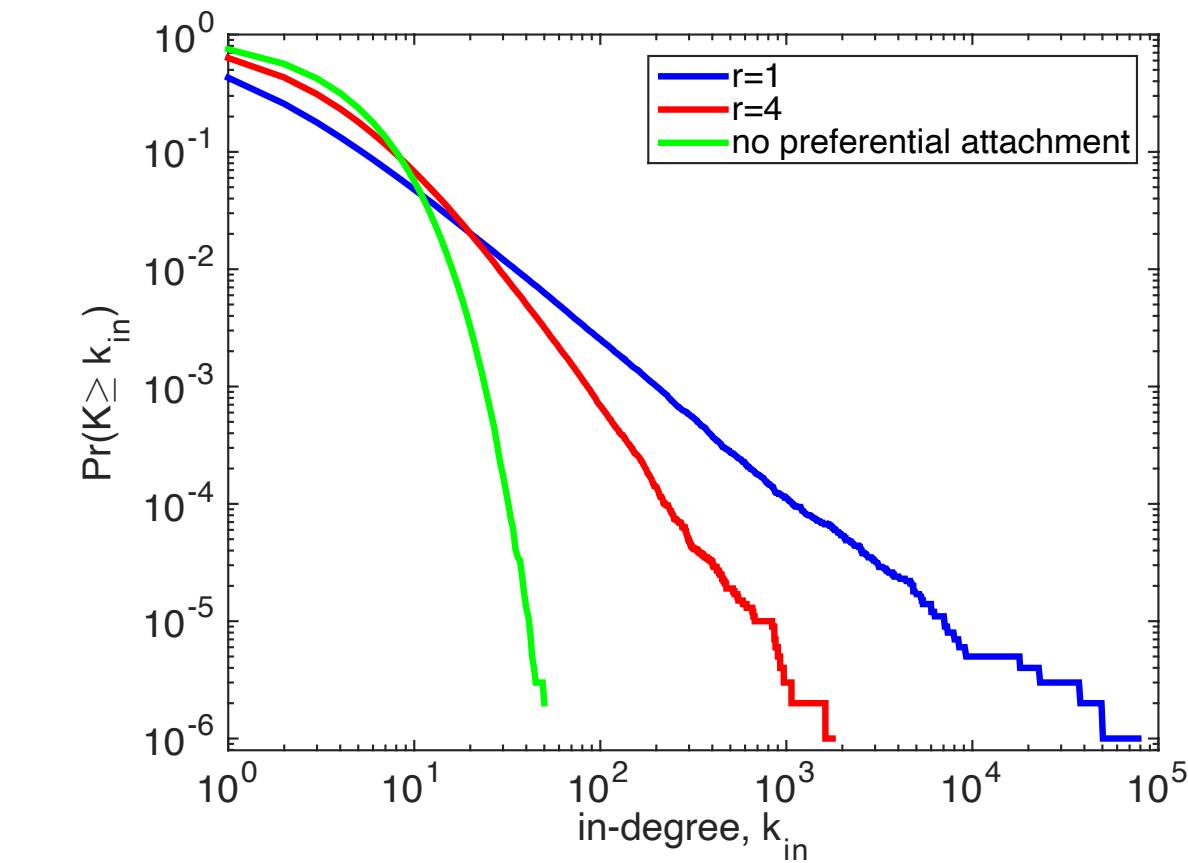
lessons learned

what works well

4. simple simulations—explore dynamics vs structure & numerical experiments



simulate epidemics (SIR) on planted partitions



simulate Price's model

lessons learned

what works well

4. team projects—teamwork & exploring their (your!) own ideas

Infinite Jest: An Elegant Hairball

K. Hunter Wapman
Brian Lubars
Carl Mueller

Design and analysis of trophallaxis network in honey bees*

Ganesh Chandra Satish, Golnar Gharooni Fard, and Rahul Chowdhury
University of Colorado, Boulder, CO, USA

Communities across U.S. Congress cosponsorship networks

Alexander Ray,^{1,*} David Crosswy,^{1,†} and Aaron Aaeng^{1,‡}

¹Department of Computer Science, University of Colorado, Boulder, CO, USA

Relational Topic Model for Congressional Bills Corpus

You Lu
Department of Computer Science
University of Colorado Boulder

Shudong Hao
Department of Computer Science
University of Colorado Boulder

Uncertainty Quantification of Voltage State in Power Systems

Samantha Molnar
CSCI5352 Final Project



The Journal of Open Source Software

DOI: [10.21105/joss.01458](https://doi.org/10.21105/joss.01458)

webweb: a tool for creating, displaying, and sharing interactive network visualizations on the web

K. Hunter Wapman¹ and Daniel B. Larremore^{1, 2}

¹ Department of Computer Science, University of Colorado Boulder, Boulder, CO, USA
² BioFrontiers Institute, University of Colorado Boulder, Boulder, CO, USA

EPJ Data Science
a SpringerOpen Journal



CrossMark

Prestige drives epistemic inequality in the diffusion of scientific ideas

Allison C. Morgan^{1*}, Dimitrios J. Economou¹, Samuel F. Way¹ and Aaron Clauset^{1,2,3}

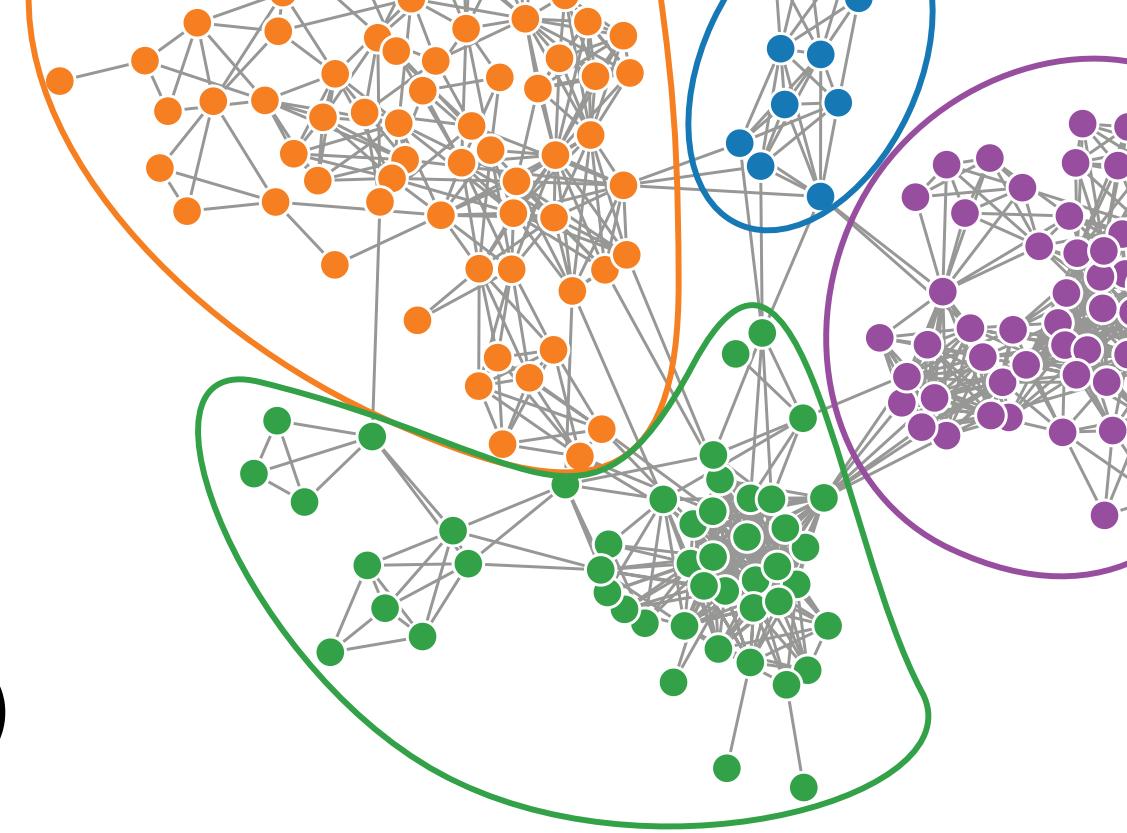
Identifying Politicians in the Panama Papers Network

CSCI 5352 Final Project

Irene Beckman, Santhanakrishnan Ramani, Ruhi Saraf

key takeaways

- **network intuition is hard to develop!**
good intuition draws on many skills
(probability, statistics, computation, causal dynamics, etc.)
- best results come from
 1. exercises to get practice with calculations
 2. practice analyzing diverse real-world networks
 3. conducting out numerical experiments & simulations
- practical tasks are a pedagogical tool (e.g., link and label prediction)
- interpreting the results requires a good *intuition* and to *think like a scientist*
- null models are key concept: is a pattern interesting? what could explain it?
- networks are fun!

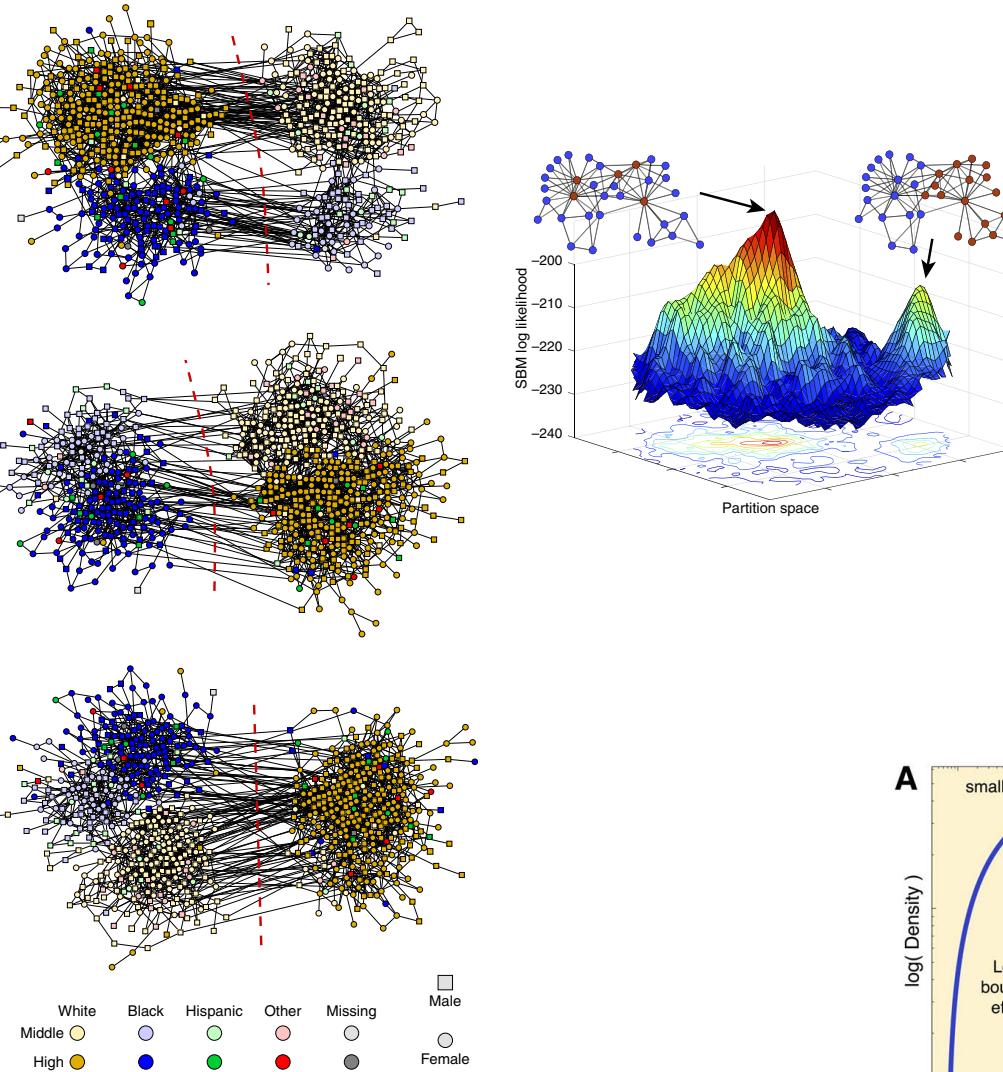


about me

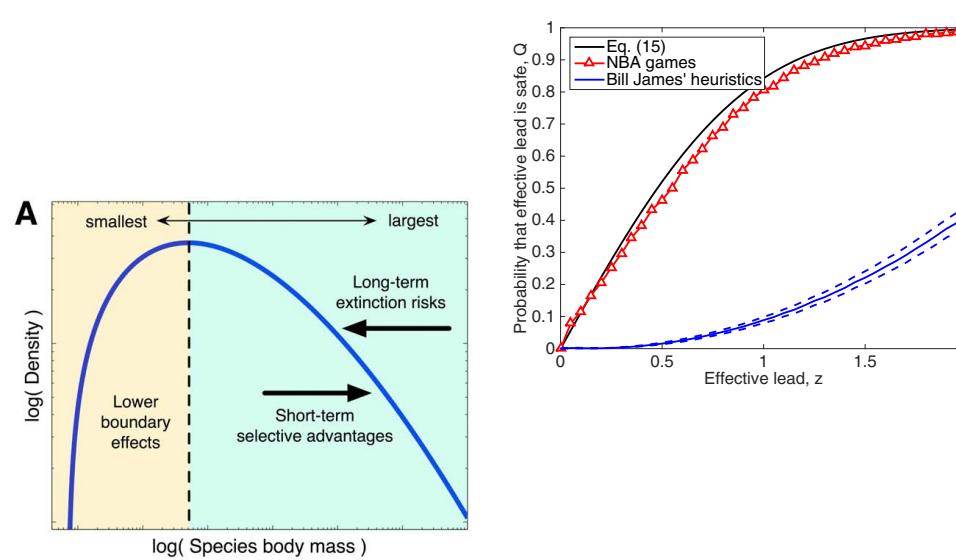
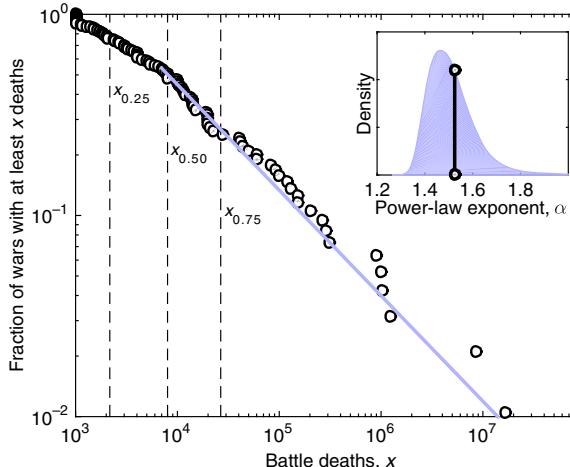


Aaron Clauset
Associate Professor
Computer Science + BioFrontiers Institute
aaronclauset.github.io

computational methods for
network analysis



complex social and
biological systems



inequality and the spread of ideas in science

