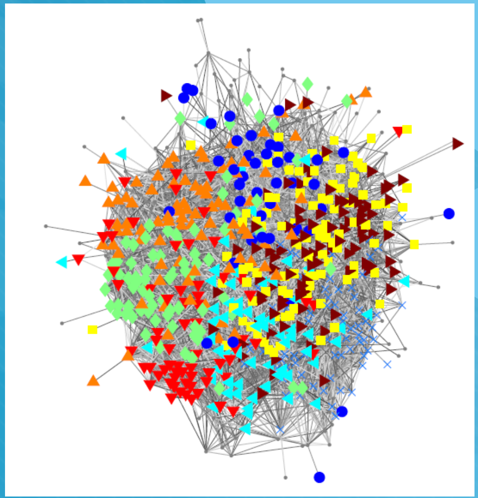


The Physics of Social Networks

Mason A. Porter
Mathematical Institute, University of Oxford

Some of my Collaborators

Sumeet Agarwal, Dani Bassett, Thomas Callaghan, Jean Carlson, Charlotte Deane, Mariano Beguerisse Díaz, Dan Fenn, James Fowler, Mark Fricker, Christi Frost, James Gleeson, Neil Johnson, Nick Jones, Eric Kelsic, Dom Kerr, Anna Lewis, Kevin Macon, Trent McCotter, Mark McDonald, Sergey Melnik, Peter Mucha, Sean Myers, Mark Newman, Yulian Ng, JP Onnela, Ye Pei, Scott Powers, Stephen Reid, Thomas Richardson, Puck Rombach, Serguei Saavedra, Amanda Traud, Casey Warmbrand, Jon Ward, Andrew Waugh, Stacy Williams, Nicholas Wymbs, Yan Zhang

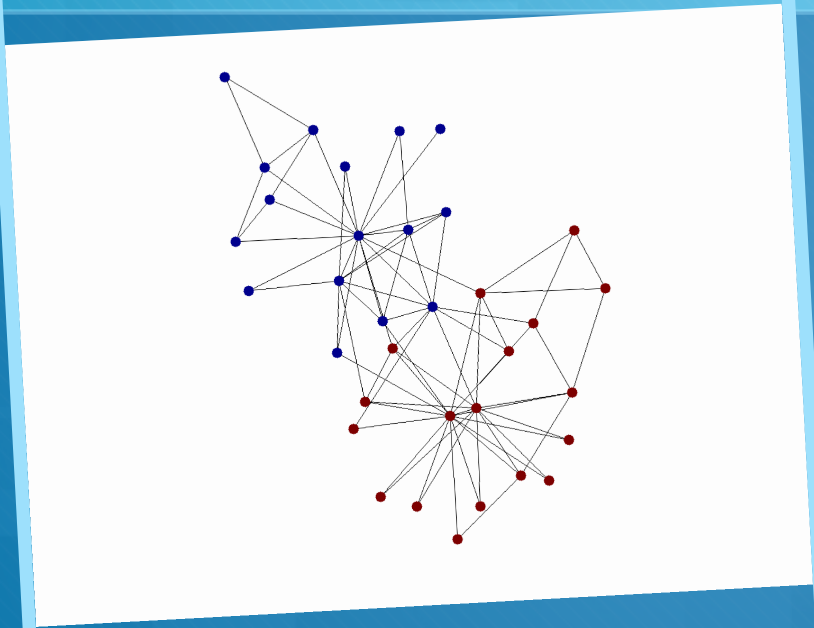
White = undergraduate students when they started doing research with me

Subjects represented include: business, mathematics, medicine, neuroscience, physics, plant sciences, political science, statistics, systems biology, people who work at HSBC bank

Outline

- Introduction
- What is network science?
- Community Structure in Networks
- Example: Facebook Networks
- More Examples
- Conclusions

What is a network?



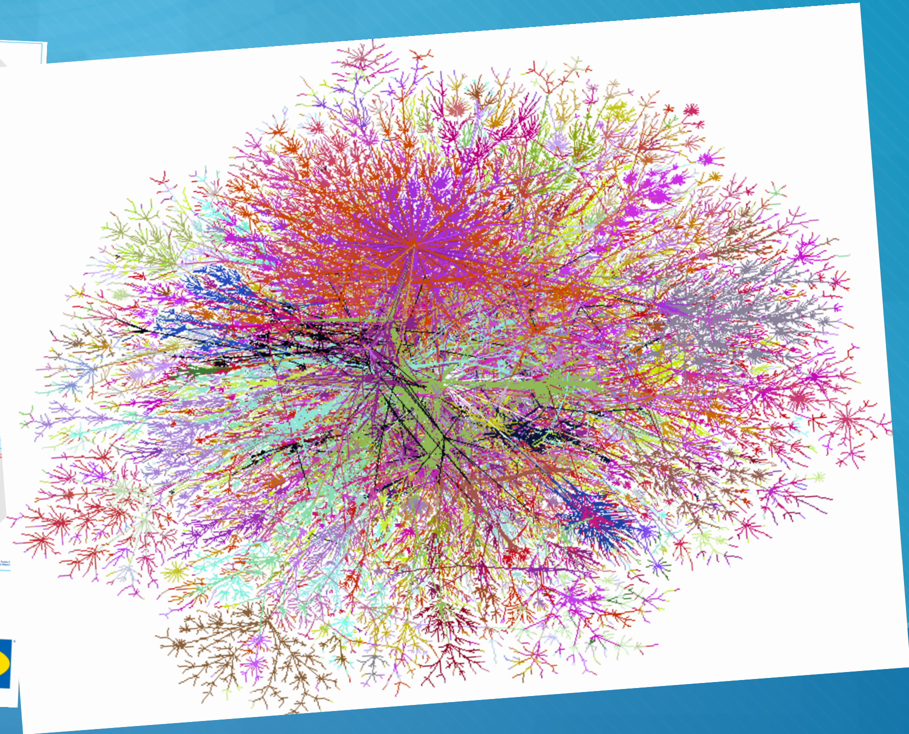
A network consists of *nodes* representing entities. Nodes are connected by *edges* representing ties between the entities.

Examples:

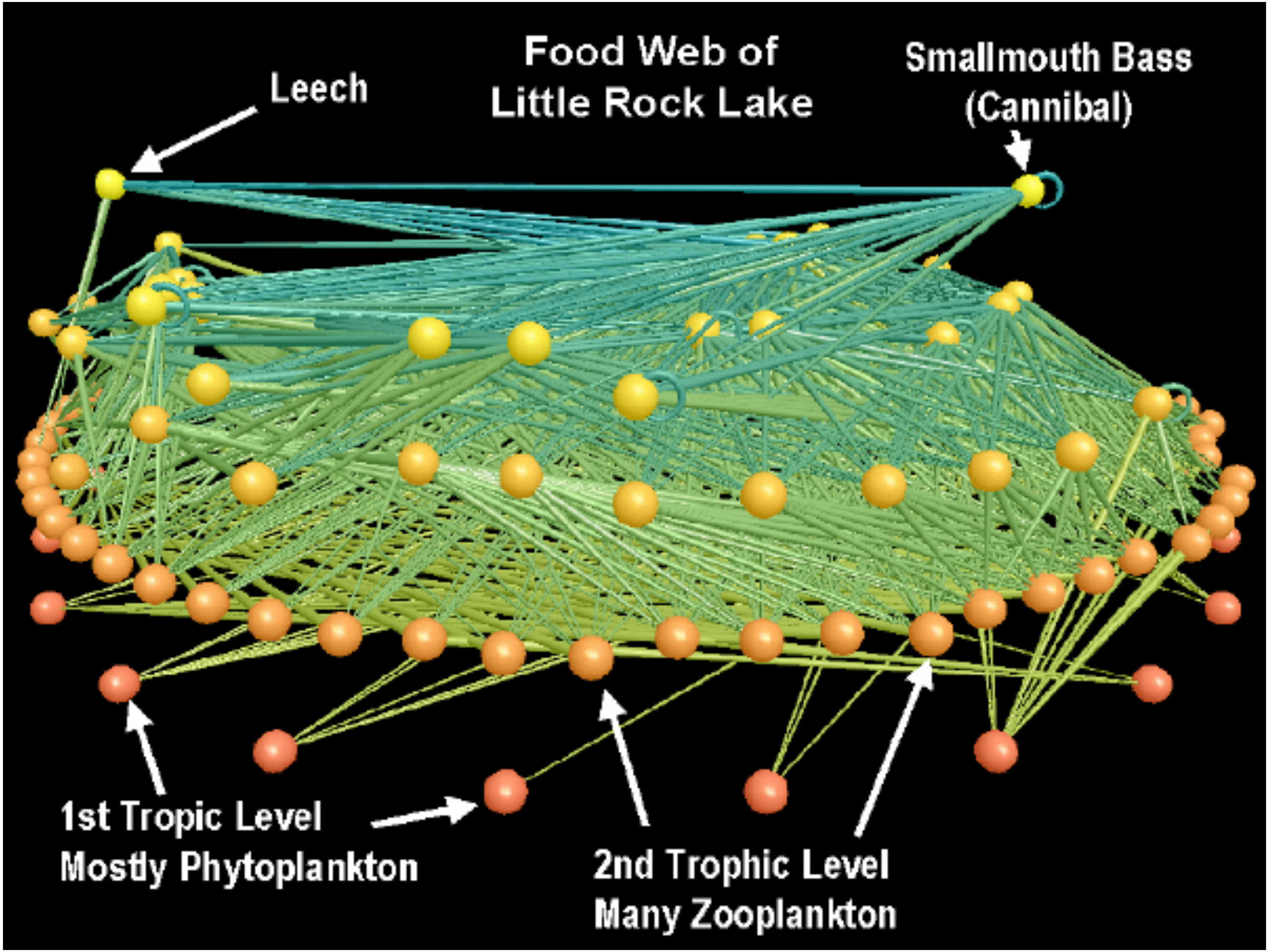
Individuals connected by Facebook "friendships".

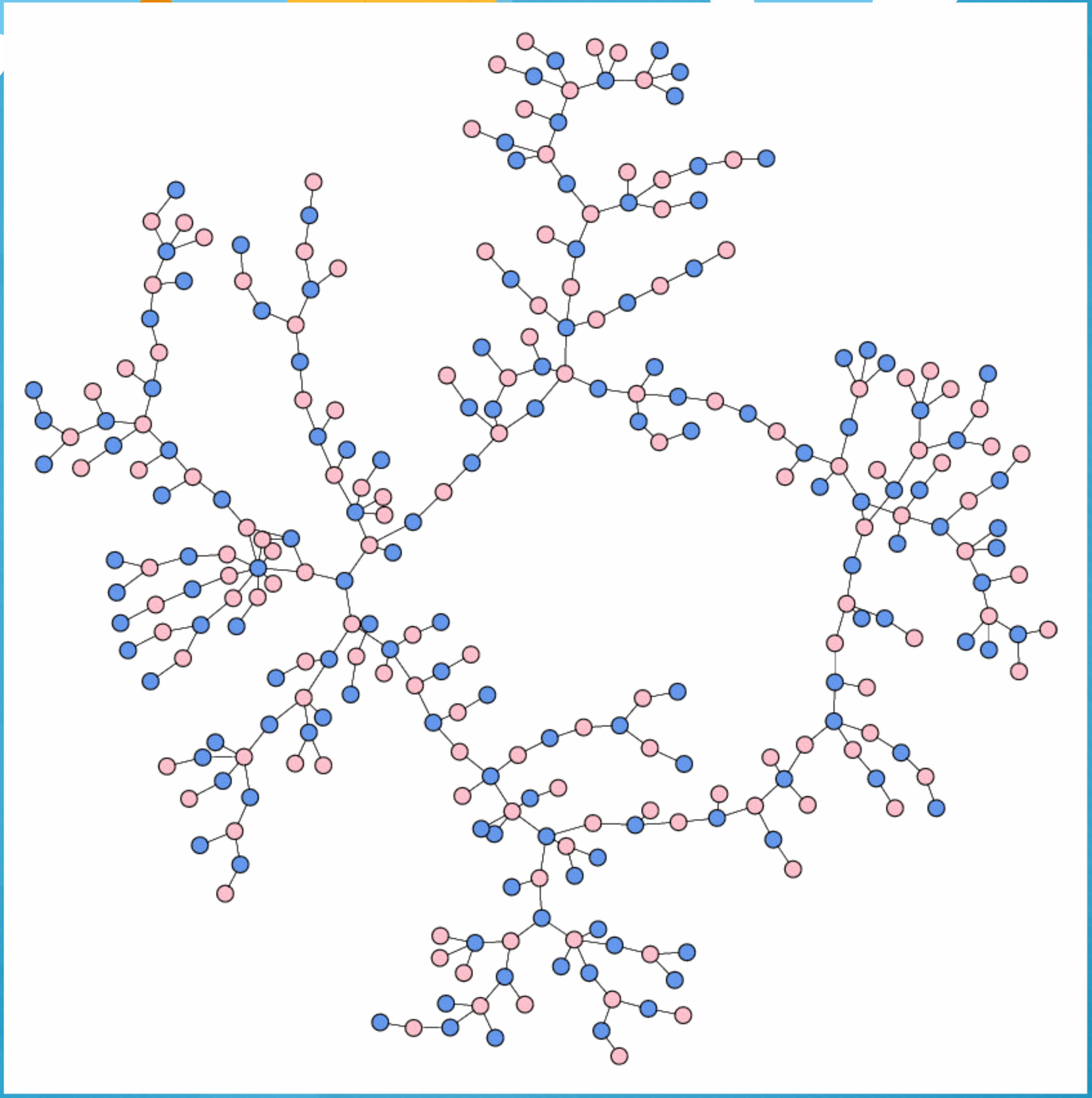
Web pages connected by hyperlinks.

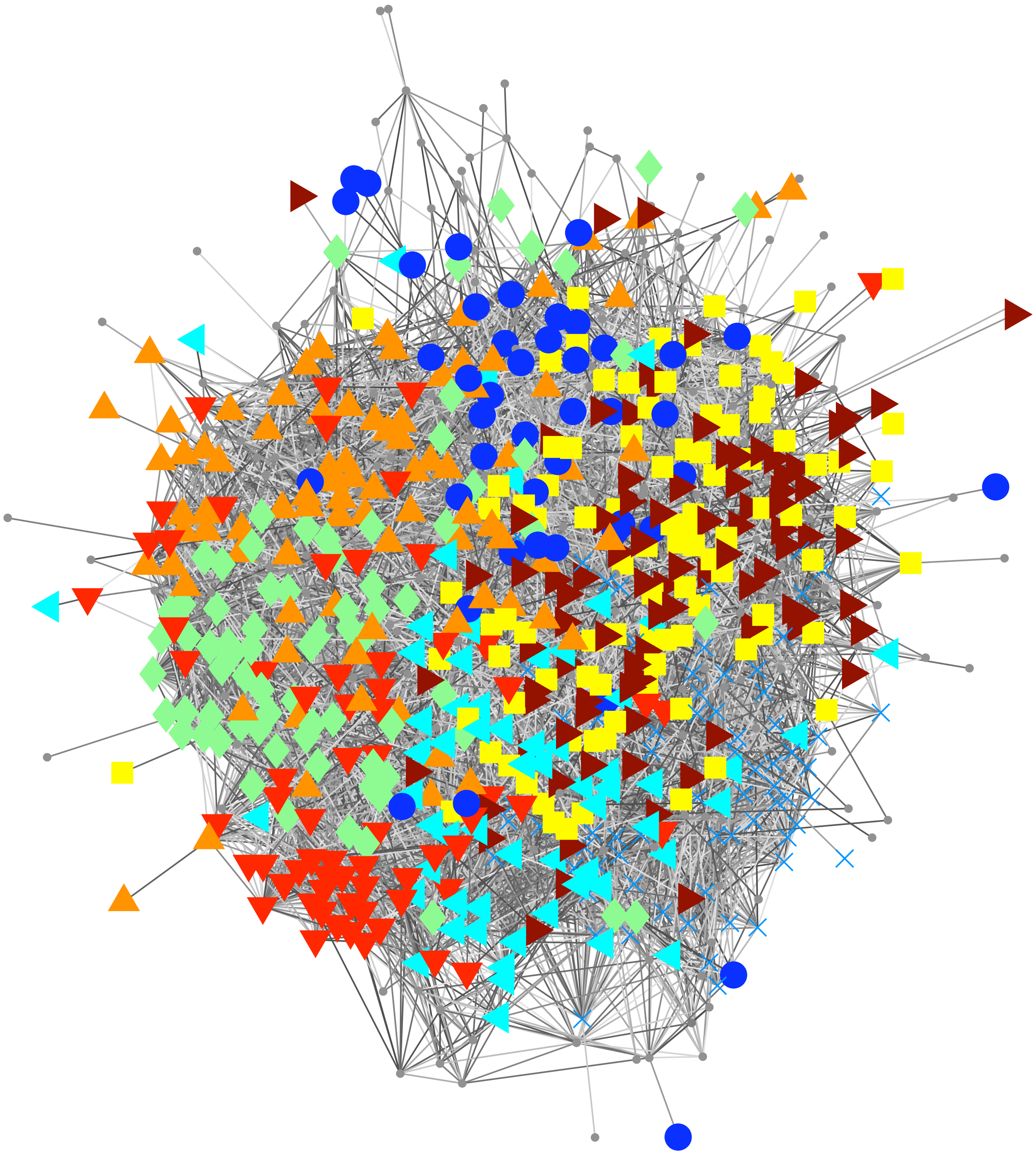
Contiguous cities on a train route.



Networks are Everywhere

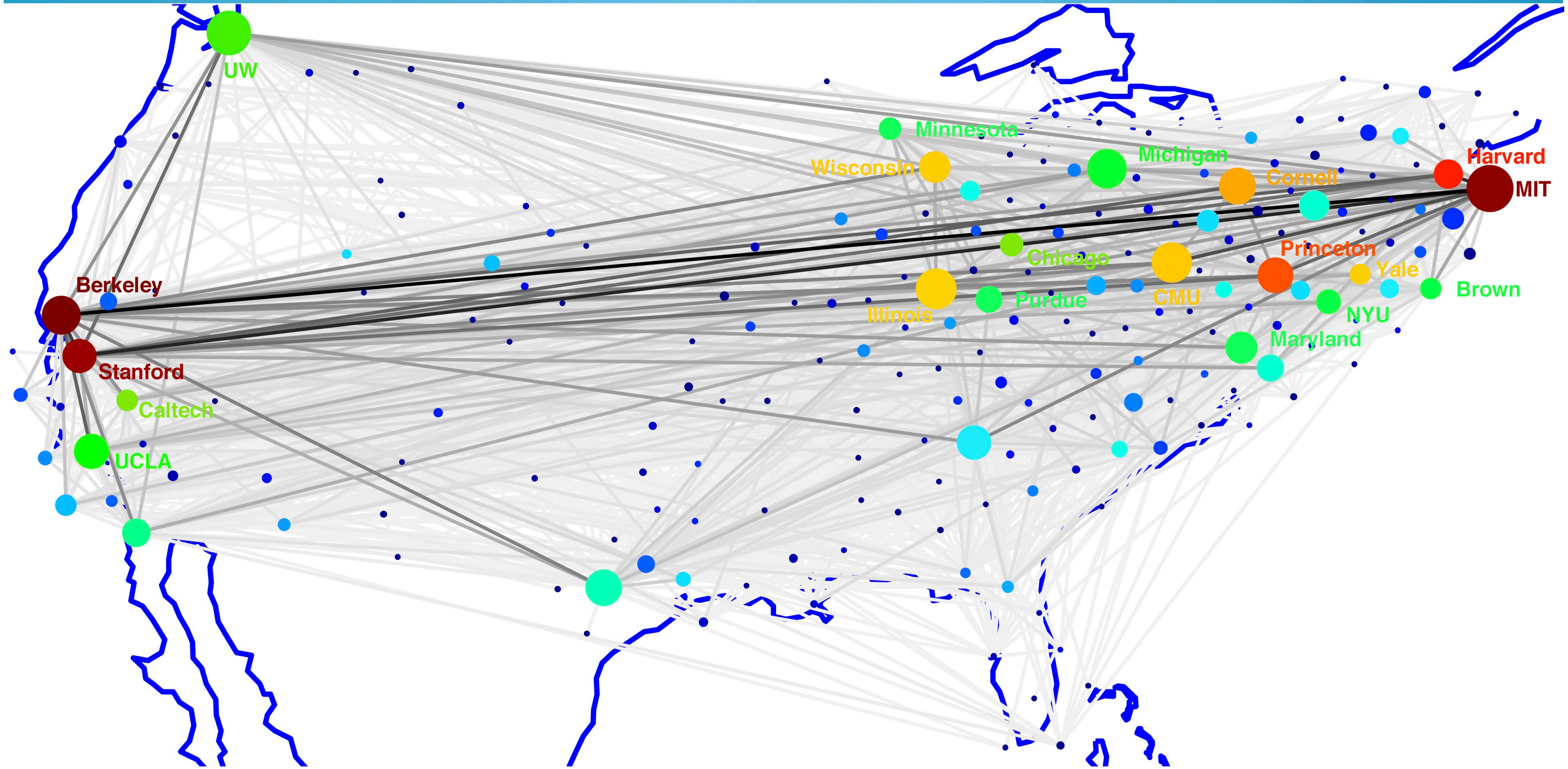

Food Web of
Little Rock Lake
Leech
Smallmouth Bass
(Cannibal)
1st Tropic Level
Mostly Phytoplankton
2nd Trophic Level
Many Zooplankton

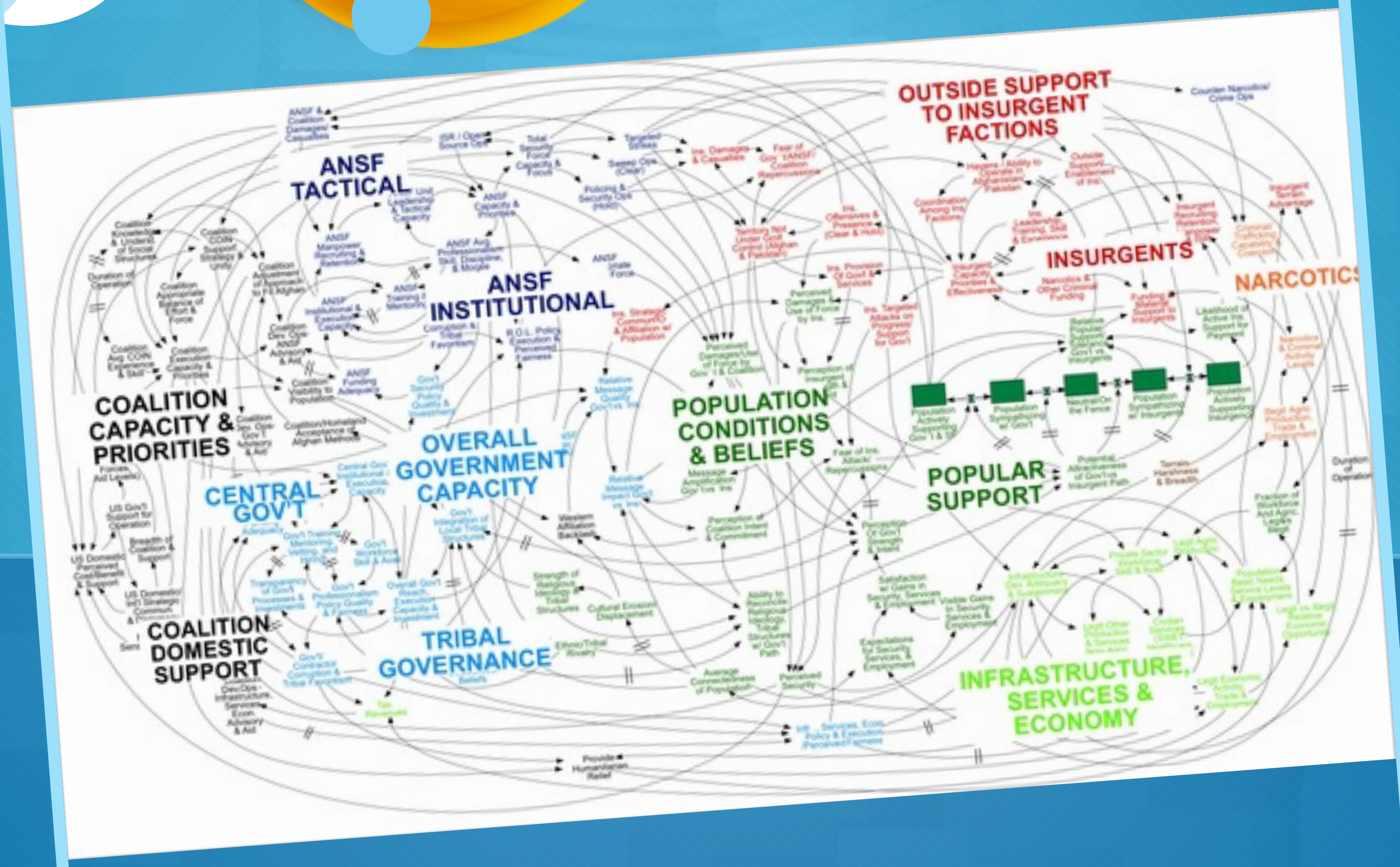




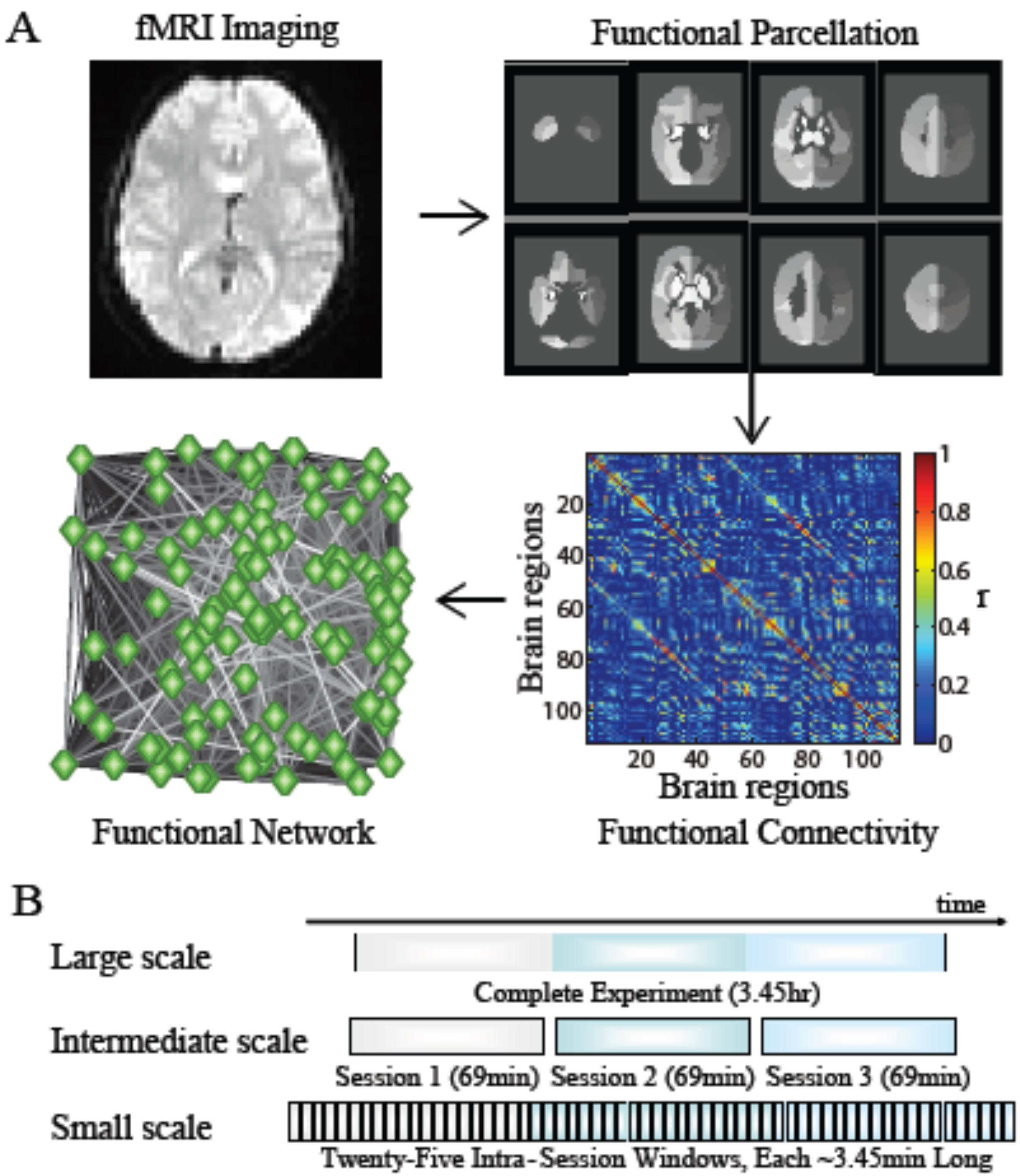
S. Myers, P. J. Mucha, and MAP [2011]. "Mathematical Genealogy & Department Prestige", to appear in *Chaos* (Gallery of Nonlinear Images).

Mathematical Genealogy Project

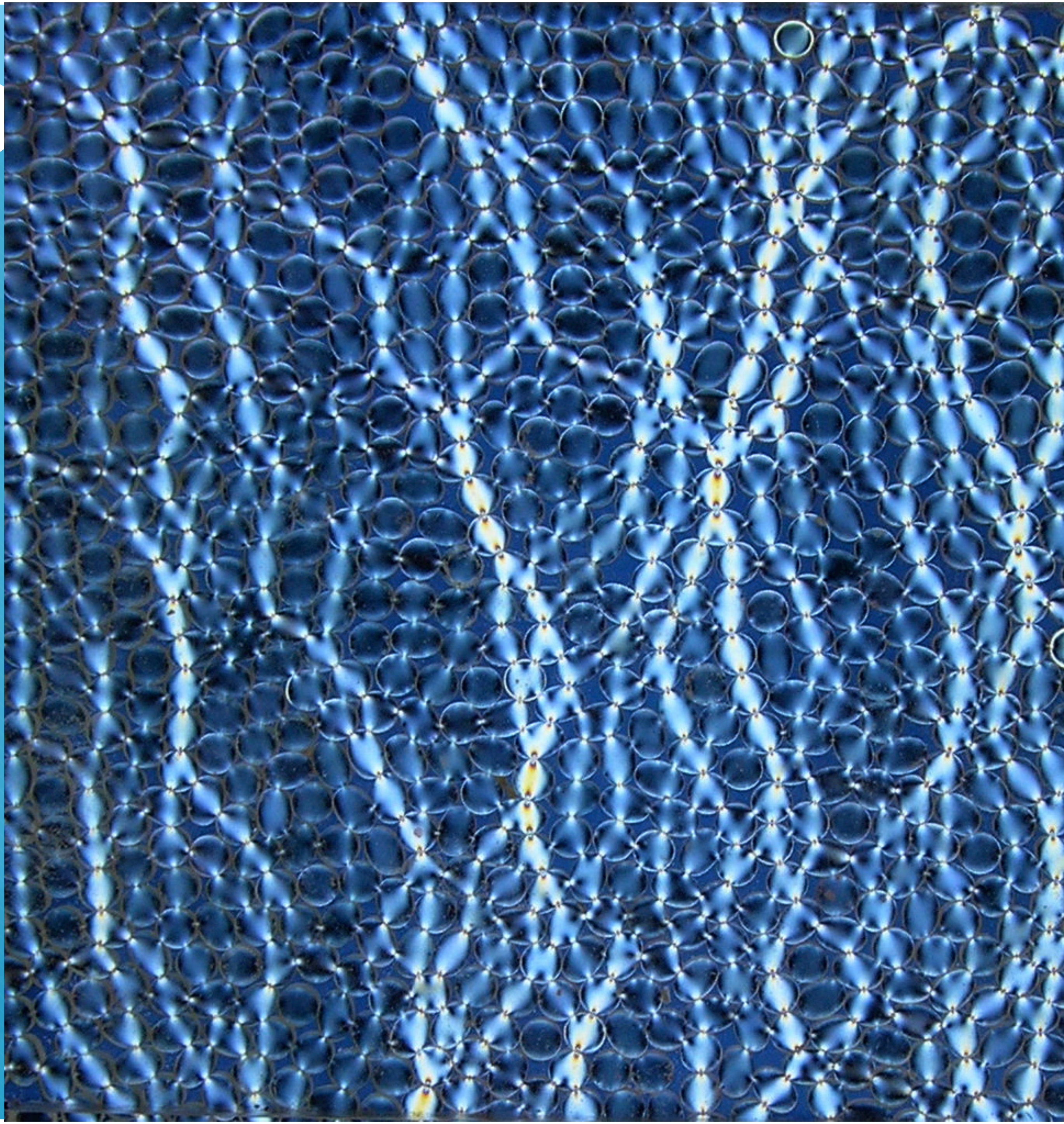




"When we understand this slide, we'll have won the war."



A
fMRI Imaging
Functional Parcellation
Functional Network
Brain regions
20
40
60
80
100
1
0.8
0.6
0.4
0.2
0
r
Brain regions
Functional Connectivity
B
time
Large scale
Complete Experiment (3.45hr)
Intermediate scale
Session 1 (69min)
Session 2 (69min)
Session 3 (69min)
Small scale
Twenty-Five Intra-Session Windows, Each ~3.45min Long



Types of Networks

- **Binary networks:** 1 if there is a connection and 0 if there isn't
- **Weighted networks:** Some value if there is a connection (representing strength of connection) and otherwise 0
- **Directed networks**
- **Bipartite networks:** only nodes of different types are connected to each other (e.g., an actor connected to a movie in which he/she appeared)
- **More …**

Representing a Network

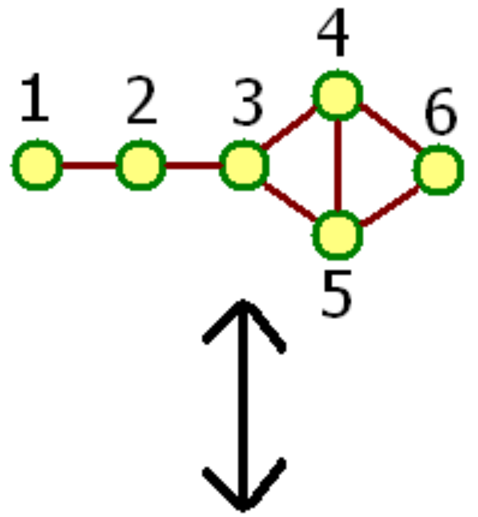
Adjacency matrix A

This example: binary ("unweighted")

$A_{ij} = 1$ if there is a connection between nodes i and j

$A_{ij} = 0$ if no connection

How do we generalize the mathematical representation to weighted, directed, and bipartite examples?



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

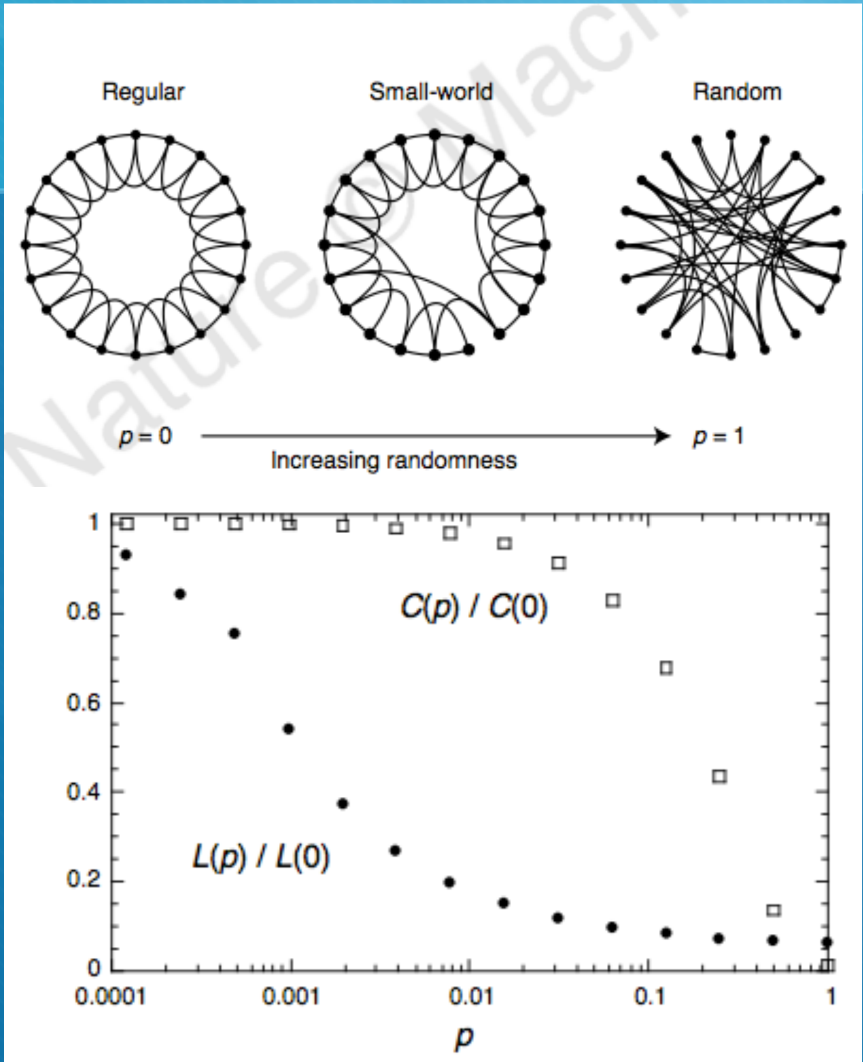
Goals of Network Science

- 1. Basic principles
  - Microscopic, mesoscopic, and macroscopic structures
- 2. Function = structure + dynamics
  - Dynamics *on* networks, dynamics *of* networks, interactions between the two
    - Time-evolution, robustness, etc.
- 3. Application
  - Inference/prediction of structure, demographics, etc.
  - Manipulation and design

Basic Principles

- **Microscopic structure**
  - Properties and roles of individual nodes and edges
  - E.g., local clustering properties, node roles
- **Mesoscopic and mascroscopic structure**
  - Summary statistics
    - Caveats: sensitive to noise and perturbations, etc.
  - Modules and hierarchies
    - *Community structure*
    - Caveats: myriad ways to compute them, choice of “null model”, etc.

Small Worlds



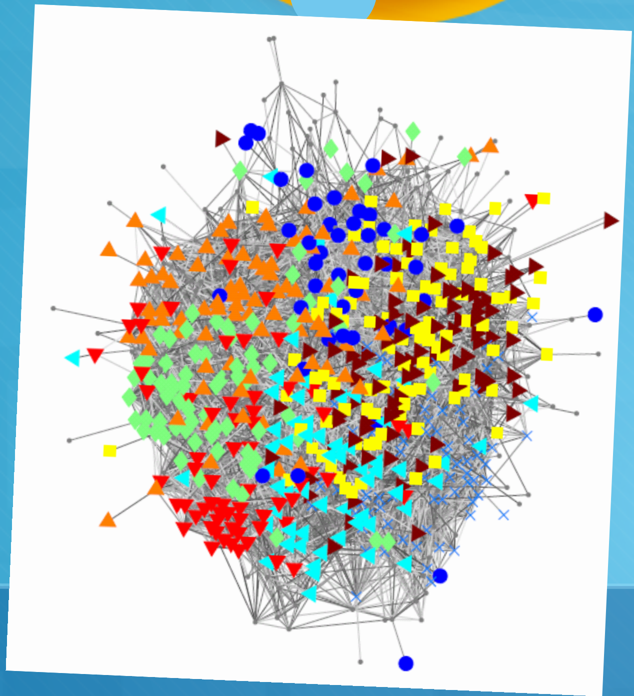
6 degrees of separation (psychologist Stanley Milgram)

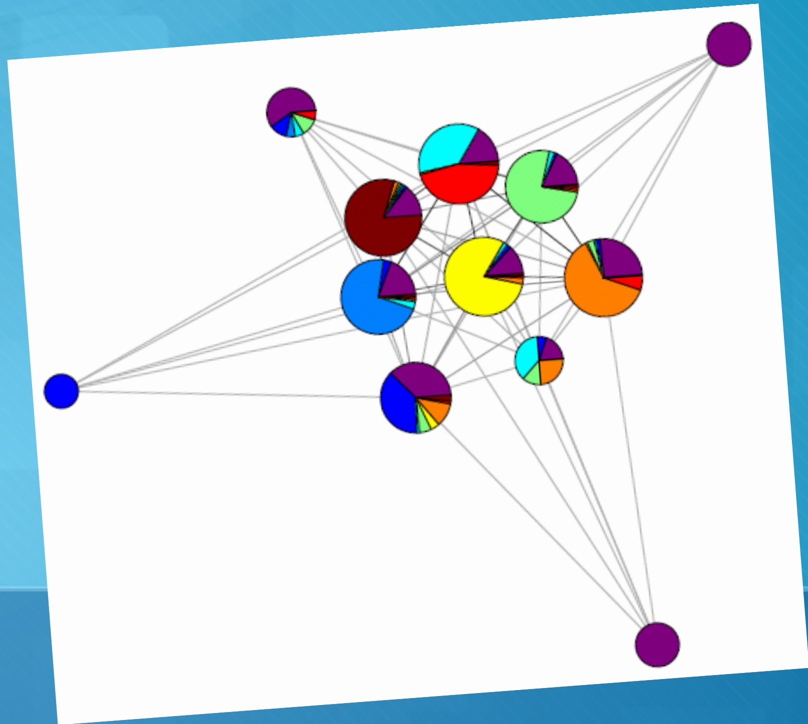
6 degrees of Kevin Bacon

Erdös numbers

Mathematical models developed starting in late 1990s to study this (starting with Watts & Strogatz, 1998)

How to navigate small worlds?





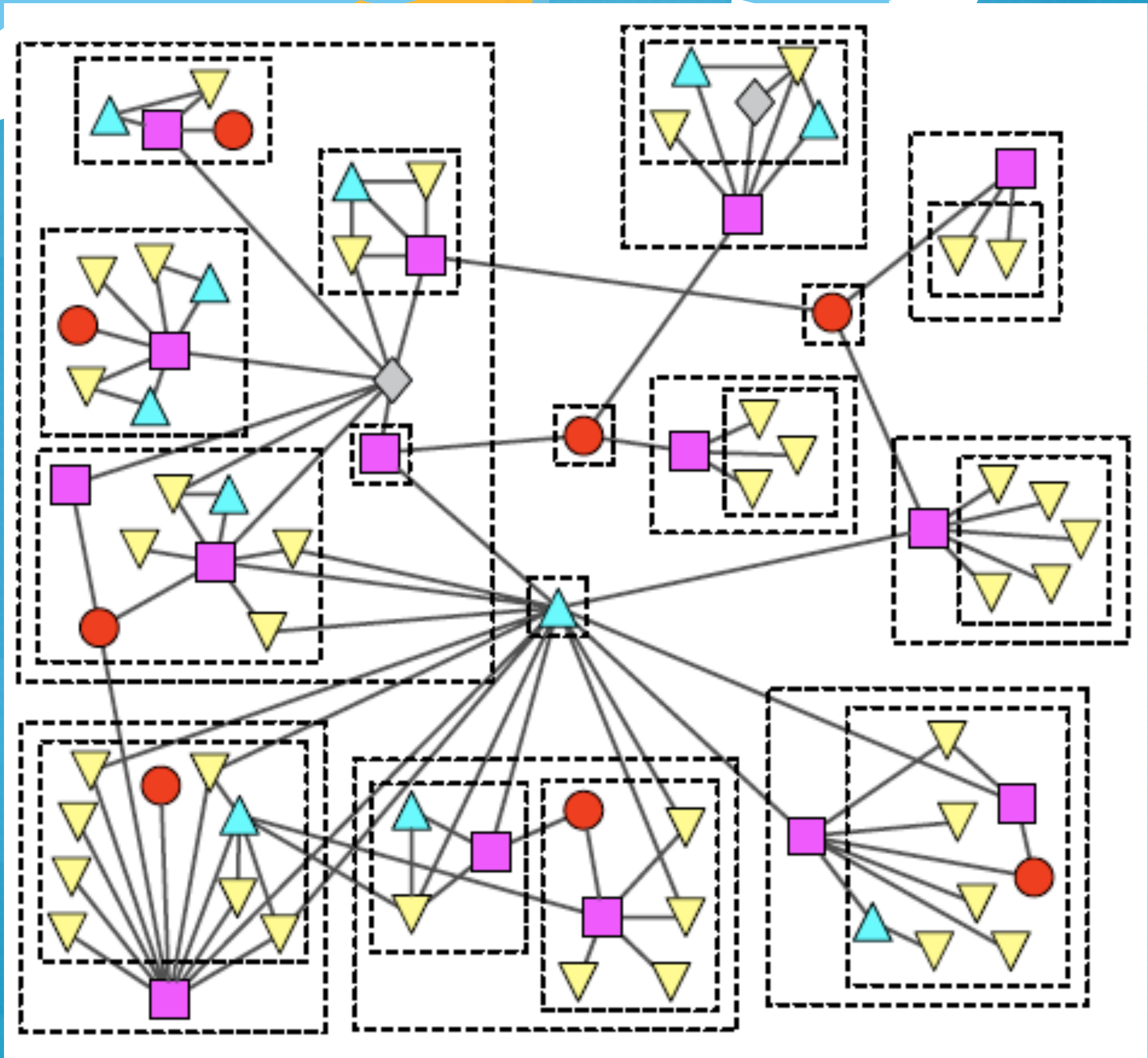
Community Detection (clustering)

Develop and use computer algorithms to group nodes (e.g. circles of friends) in an automated fashion.

The problem is both very difficult and very interesting.

Detecting Communities

- **Communities = Cohesive groups/modules/mesoscopic structures**
  - In statistical physics, one tries to derive macroscopic and mesoscopic insights from microscopic information
- **Community structure is *both* modular *and* hierarchical**
- Communities have larger density of internal ties relative to some null model for what ties are present at random
  - "Modularity"



Detecting Communities

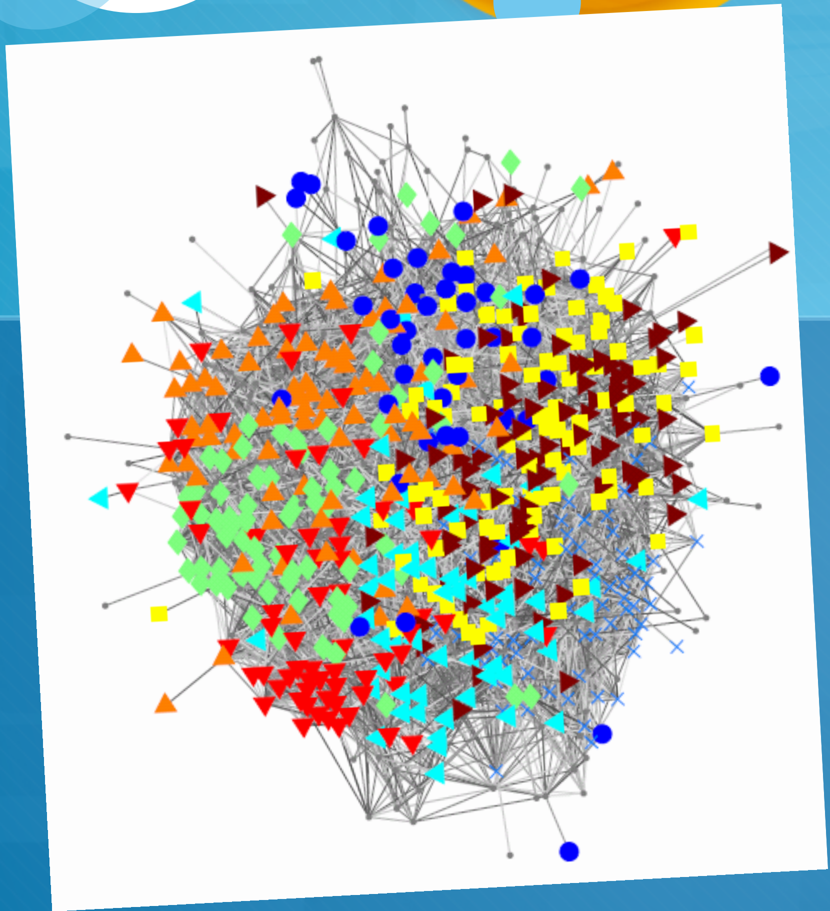
- MAP, J.-P. Onnela, & P. J. Mucha [2009], *Notices of the American Mathematical Society* 56(9): 1082-1097, 1164-1166
- Several types of methods
  - Agglomerative
  - Divisive
  - Local methods
  - Link-based

Modularity

$$H = -\sum_{ij} J_{ij}\delta(\sigma_i, \sigma_j)$$

- **Minimize Potts Hamiltonian H**
  - σ_i = community assignment (spin state) of node i
  - $J_{ij} > 0$ → "ferromagnetic" interaction between i & j → nodes i and j try to be in the same state
  - $J_{ij} < 0$ → "antiferromagnetic" interaction between i & j → nodes i and j try to be in different states
- **Modularity optimization**
  - A_{ij} = adjacency matrix
  - $W = (1/2)\Sigma_{ij}A_{ij}$ = sum of all edge weights
  - P_{ij} = prob(i connected to j) in null model
    - Newman-Girvan: $p_{ij} = k_i k_j/(2W)$, where $k_i = \Sigma_j A_{ij}$ = total edge weight of node i
    - "Resolution parameter": use $\lambda^* p_{ij}$

$$J_{ij} = \frac{A_{ij} - p_{ij}}{W}$$



Facebook Networks

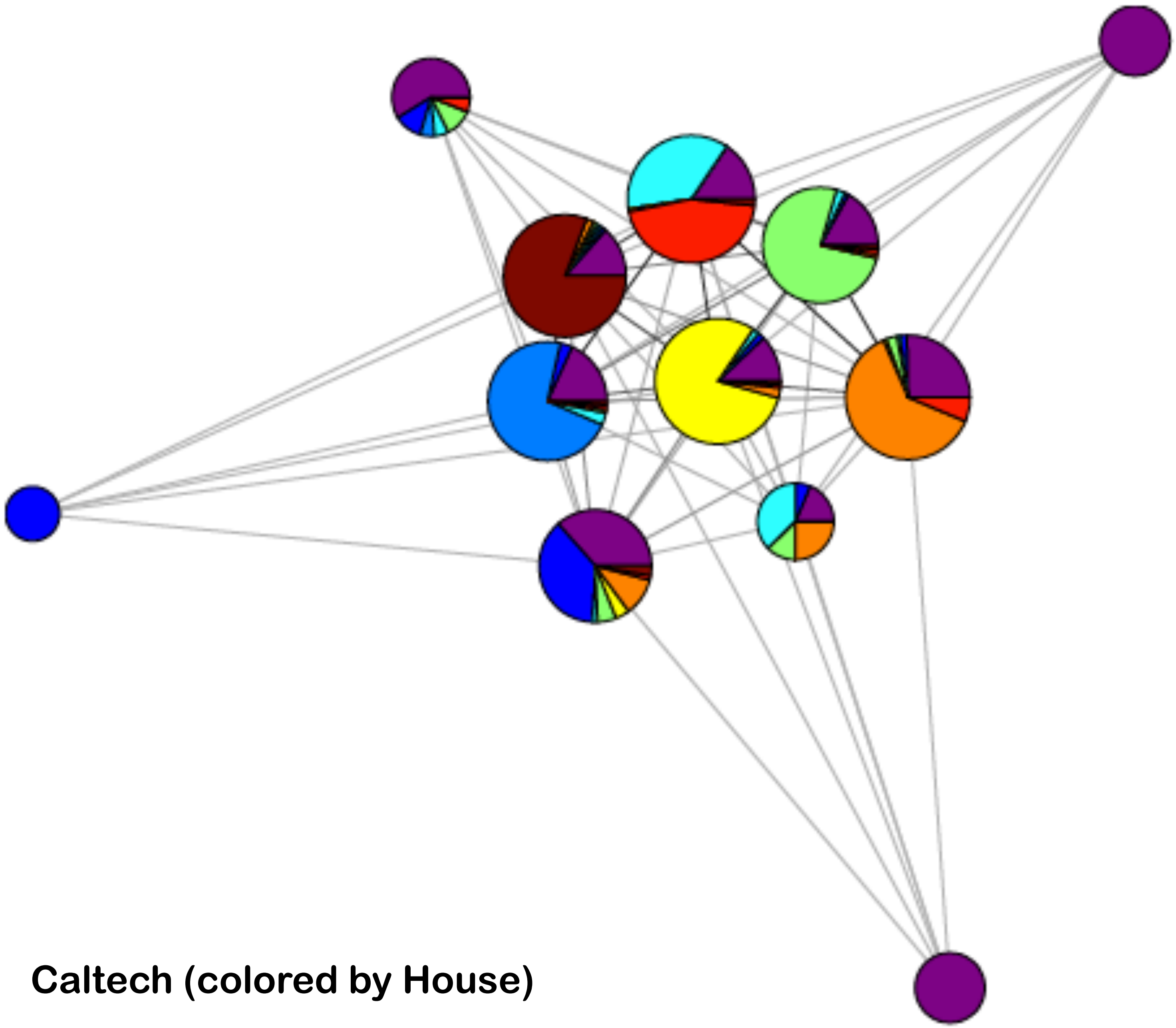
Nodes = individuals

Links = self-identified friendships (1 or 0)

Data

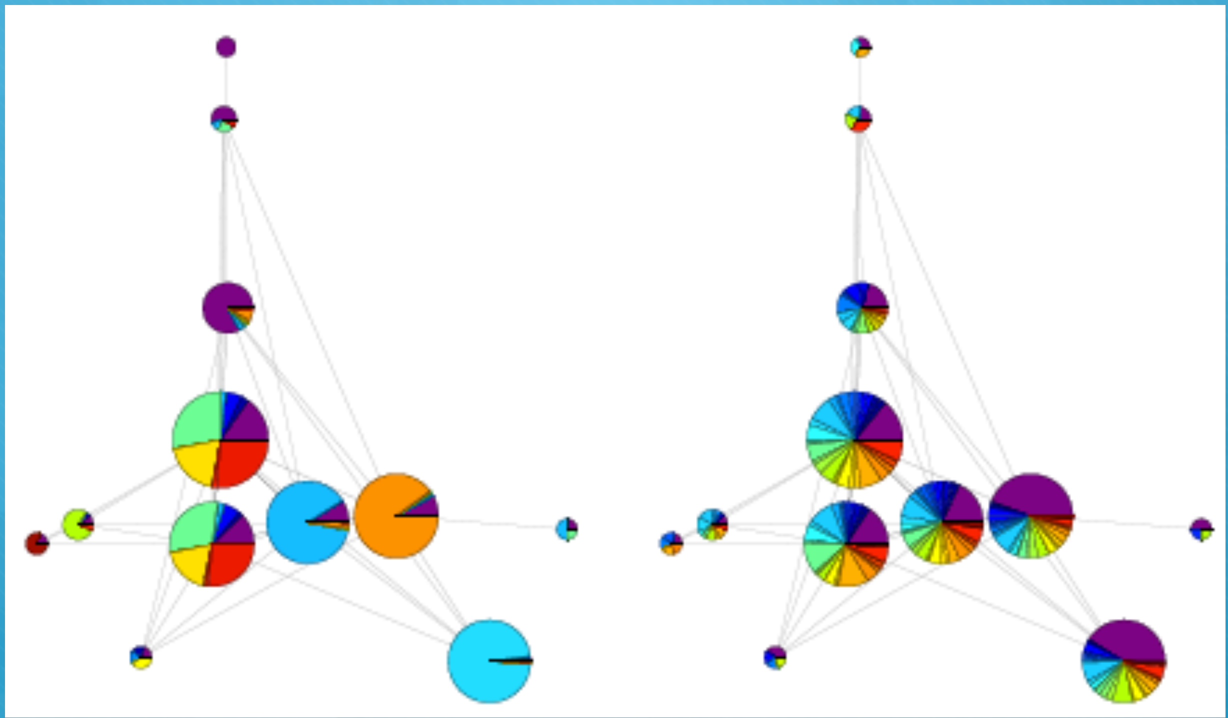
- 100 different universities (full networks)
- Single-time snapshot: September 2005
  - Facebook was university-only
- Self-reported demographics
  - Gender, class year, high school, major, dormitory/"House"
- Provided by Adam D'Angelo & Facebook

A. L. Traud, E. D. Kelsic, PJM, & MAP [2008], to appear in *SIAM Review* (arXiv: 0809.0690)
ALT, C. Frost, PJM, & MAP [2009], *Chaos* 19(4): 041104 (Gallery of Nonlinear Images)
ALT, PJM, & MAP, arXiv:1102.2166



Caltech (colored by House)

Princeton: Class Year & Major

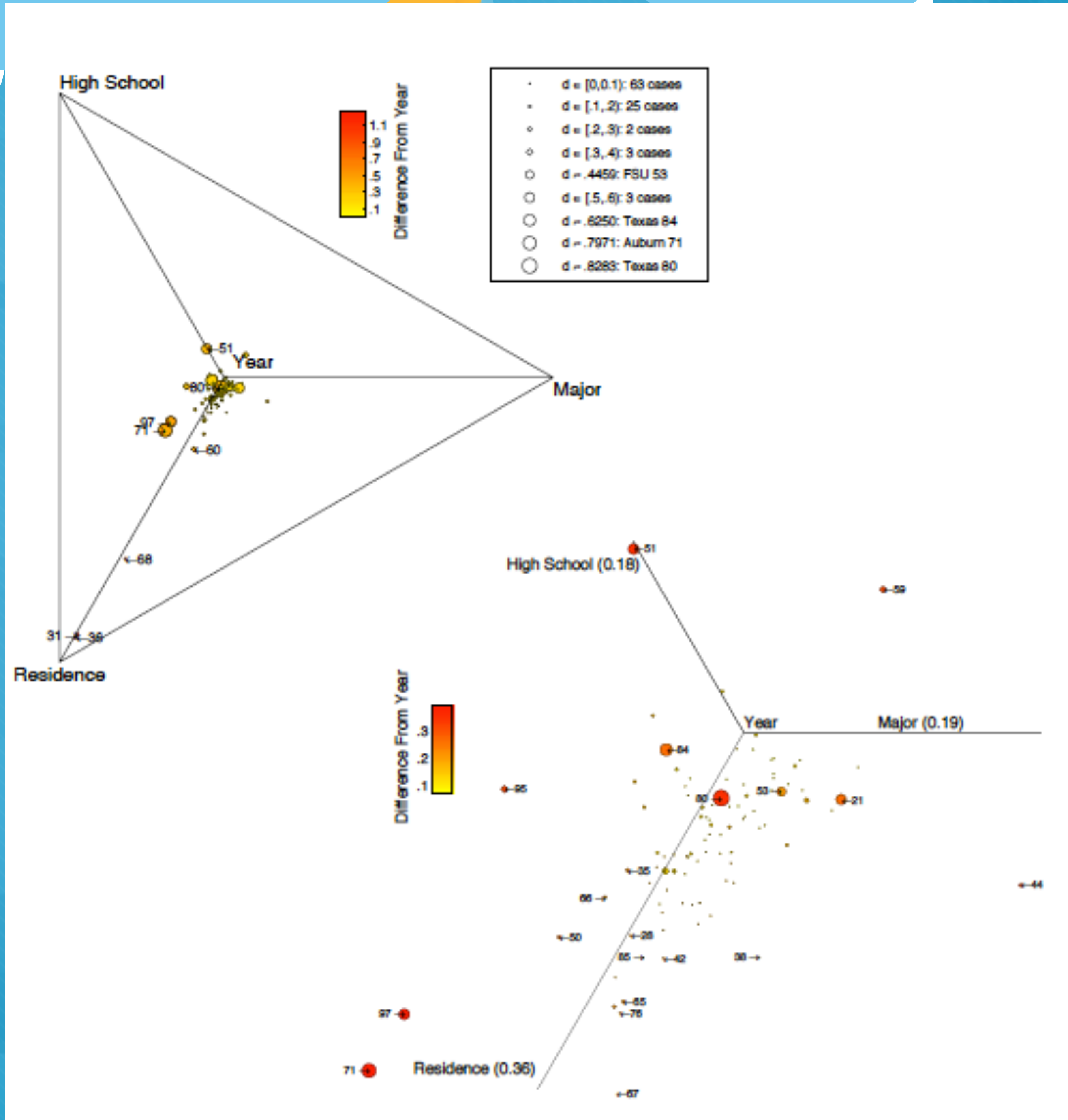


Quantitative Comparisons

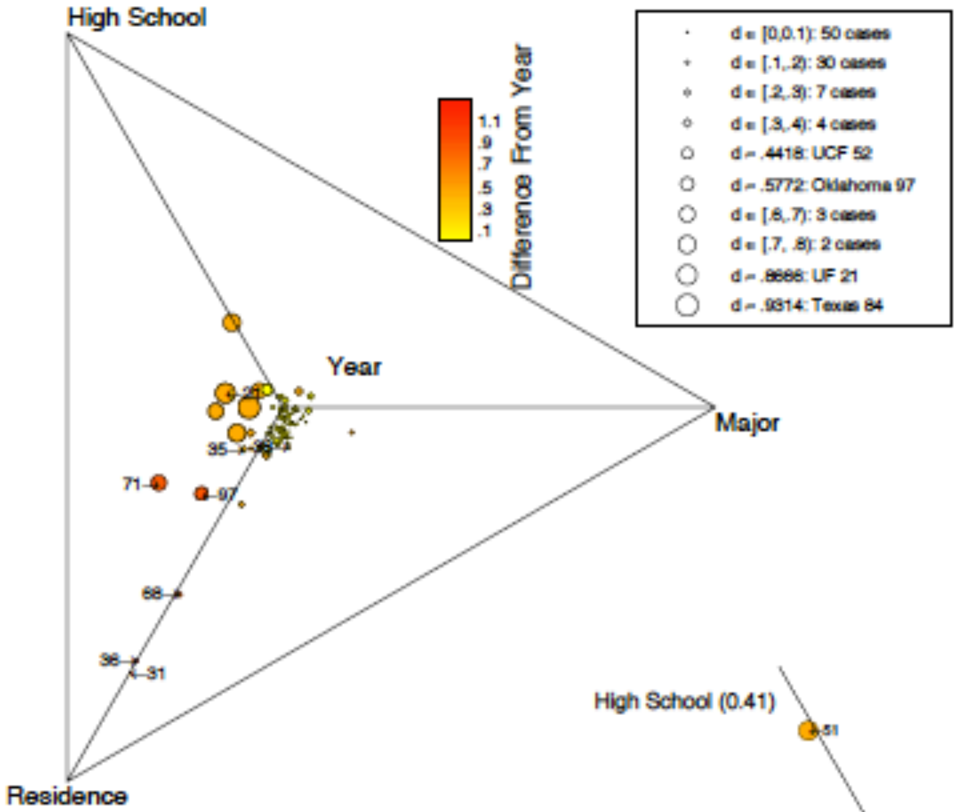
- Princeton example: Is this random? Is it correlated? Visually, it's not clear!
- My collaborators and I have used quantitative methods (e.g. permutation tests) to study the community structure of the Facebook networks from the 100 US universities.

How do universities organize?

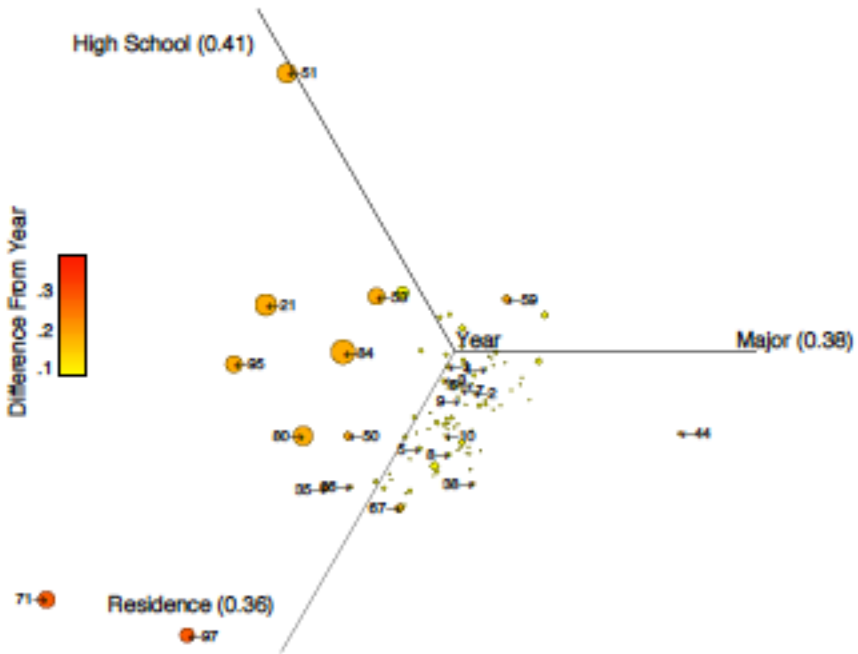
- Houses are important at Caltech (reality check for methodology)
- High school is more important at large state universities
- Class year is the most important factor at most universities and dorm is often a very strong secondary factor
- Major has varying importance at different universities
- Our work suggests interesting future research projects for social scientists



High School
Year
Major
Residence
Difference From Year
d ∈ [0,0.1): 63 cases
d ∈ [.1,.2): 25 cases
d ∈ [.2,.3): 2 cases
d ∈ [.3,.4): 3 cases
d = .4459: FSU 53
d ∈ [.5,.6): 3 cases
d = .6250: Texas 84
d = .7971: Auburn 71
d = .8283: Texas 80
High School (0.18)
Major (0.19)
Residence (0.36)

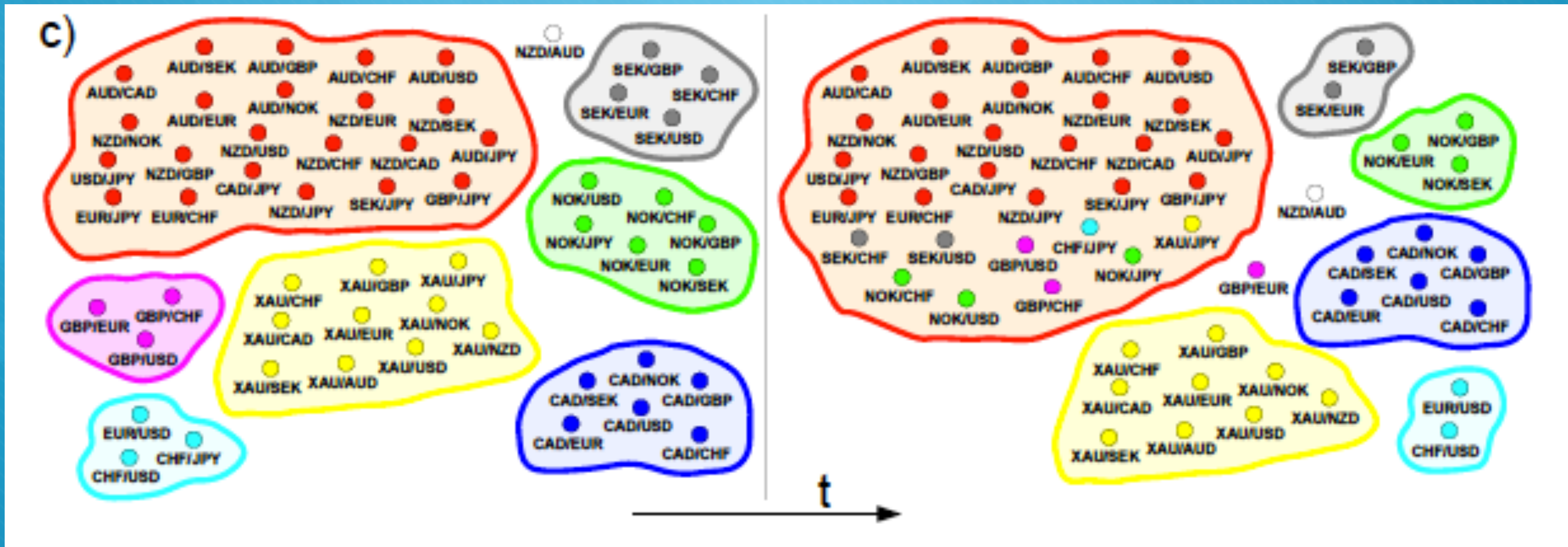


High School
Year
Major
Residence
Difference From Year
1.1
.9
.7
.5
.3
.1
d ∈ [0,0.1): 50 cases
d ∈ [.1,.2): 30 cases
d ∈ [.2,.3): 7 cases
d ∈ [.3,.4): 4 cases
d = .4418: UCF 52
d = .5772: Oklahoma 97
d ∈ [.6,.7): 3 cases
d ∈ [.7, .8): 2 cases
d = .8666: UF 21
d = .9314: Texas 84



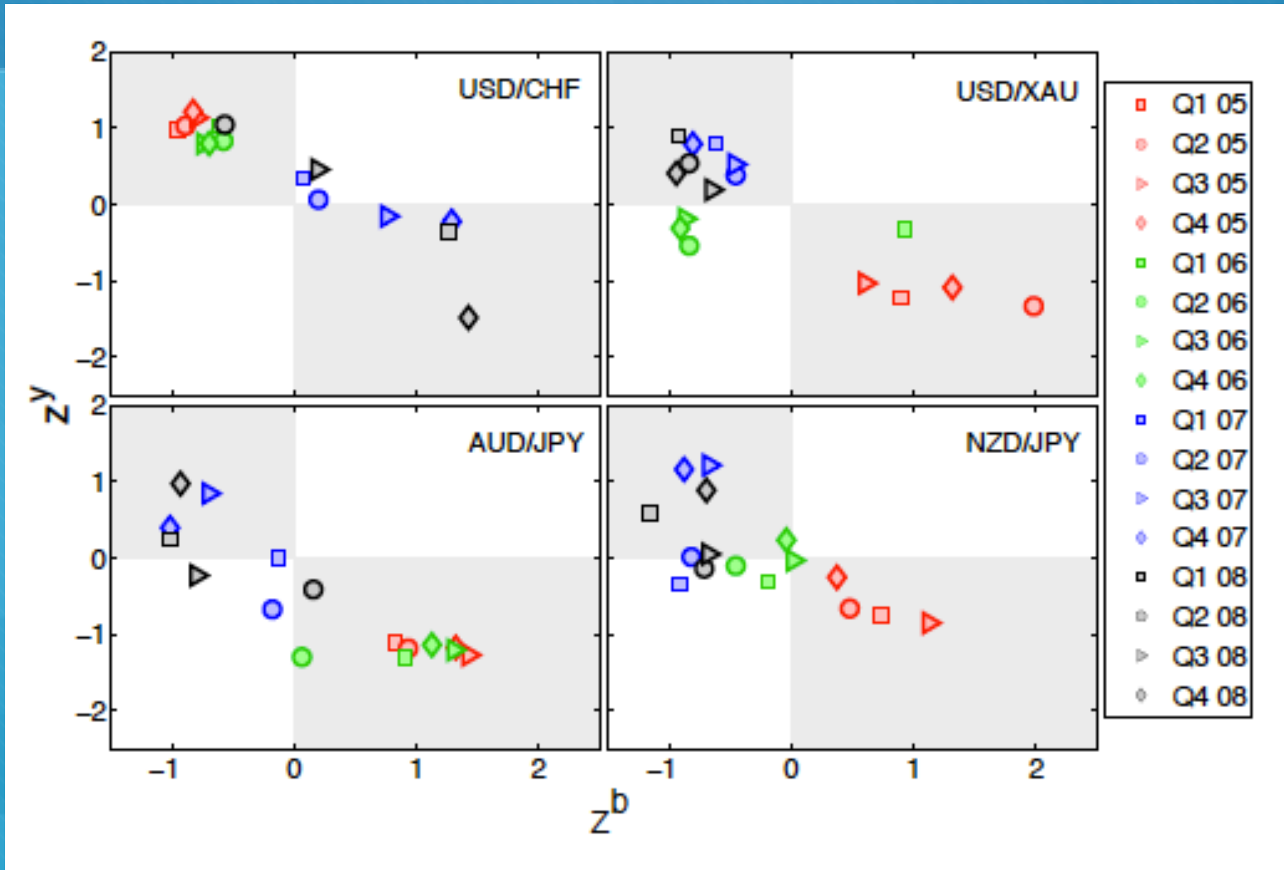
High School (0.41)
Year
Major (0.38)
Residence (0.36)
Difference From Year
.3
.2
.1

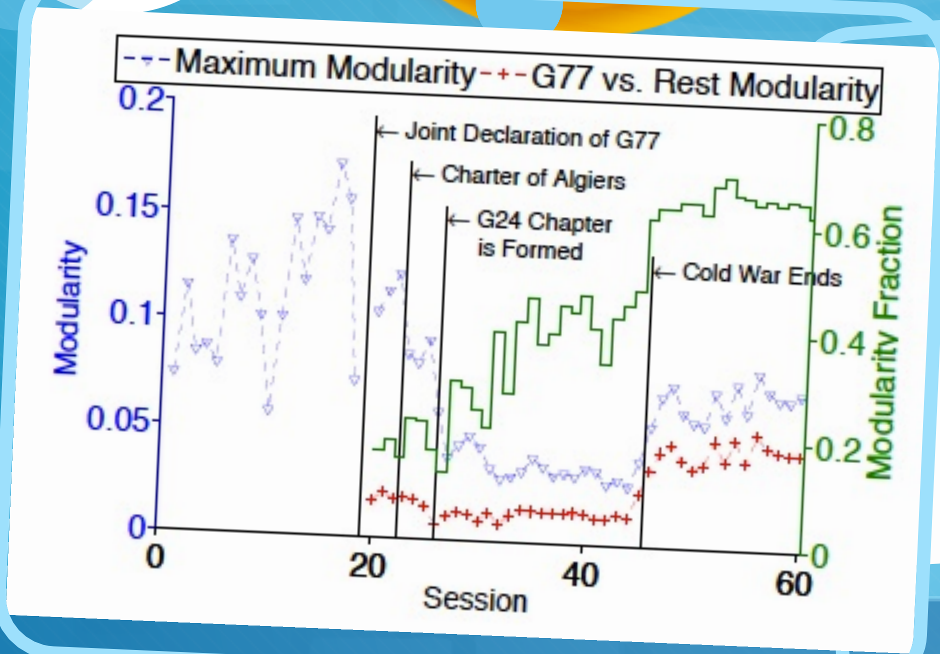
Application to Finance

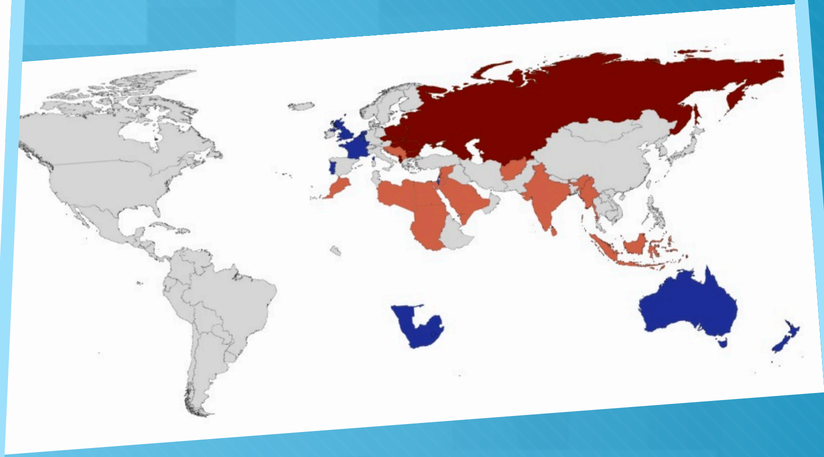


D. J. Fenn, MAP, M.McDonald, S. Williams, N. F. Johnson, and N. S. Jones [2009] *Chaos* **19**(3), 033119
DJF, MAP, PJM, MM, SW, NFJ, and NSJ [2010], arXiv: 0905.4912

Some exchange rates changed roles in the network right when the Credit Crunch began!





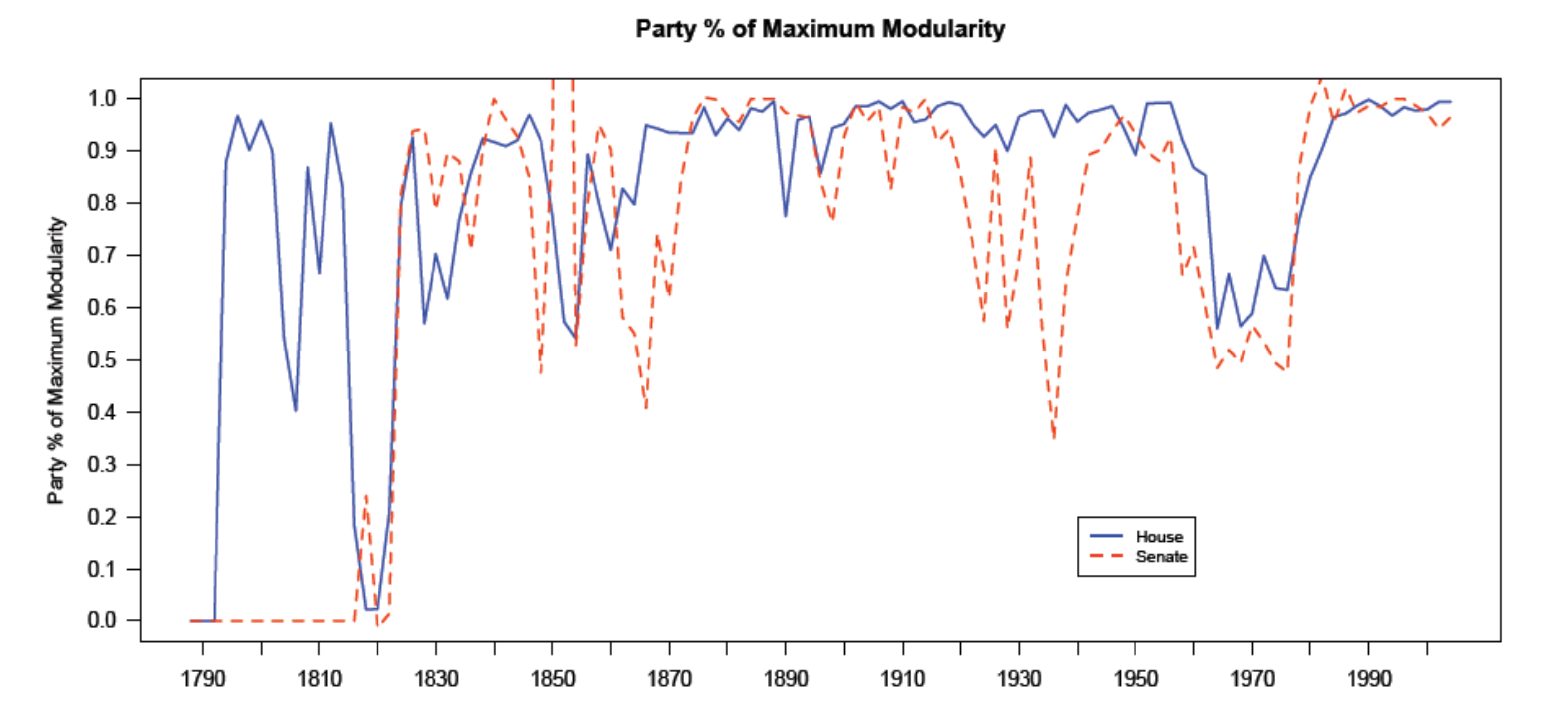


Voting Networks

Example: Voting on resolutions in the United Nations General Assembly.

Kevin T. Macon, PJM, and MAP [2011], arXiv:1010.3757

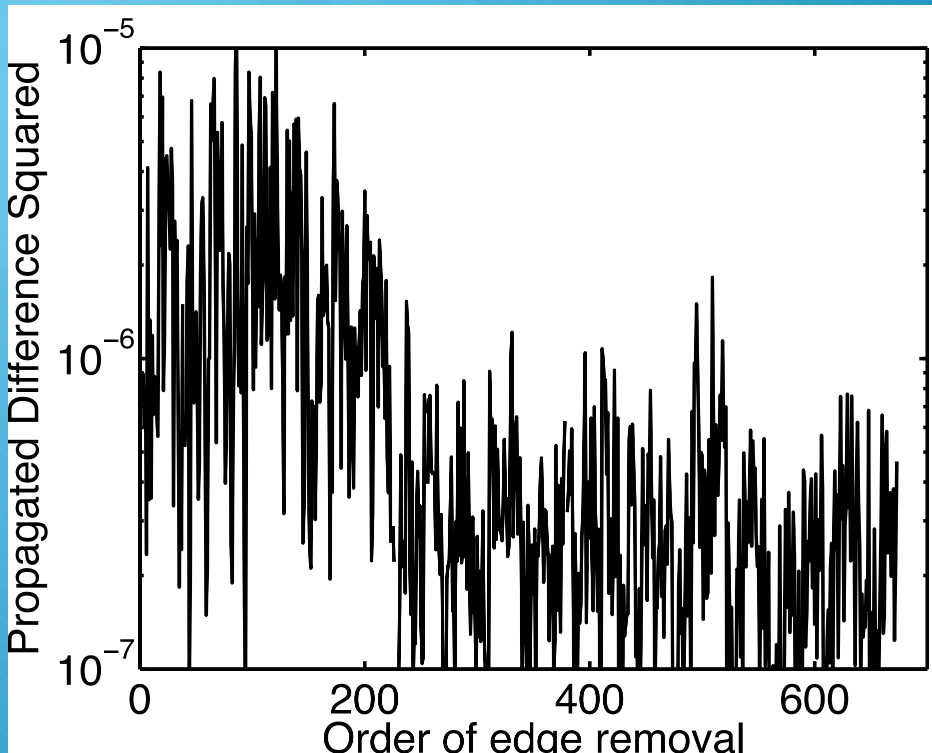
Political Realignments

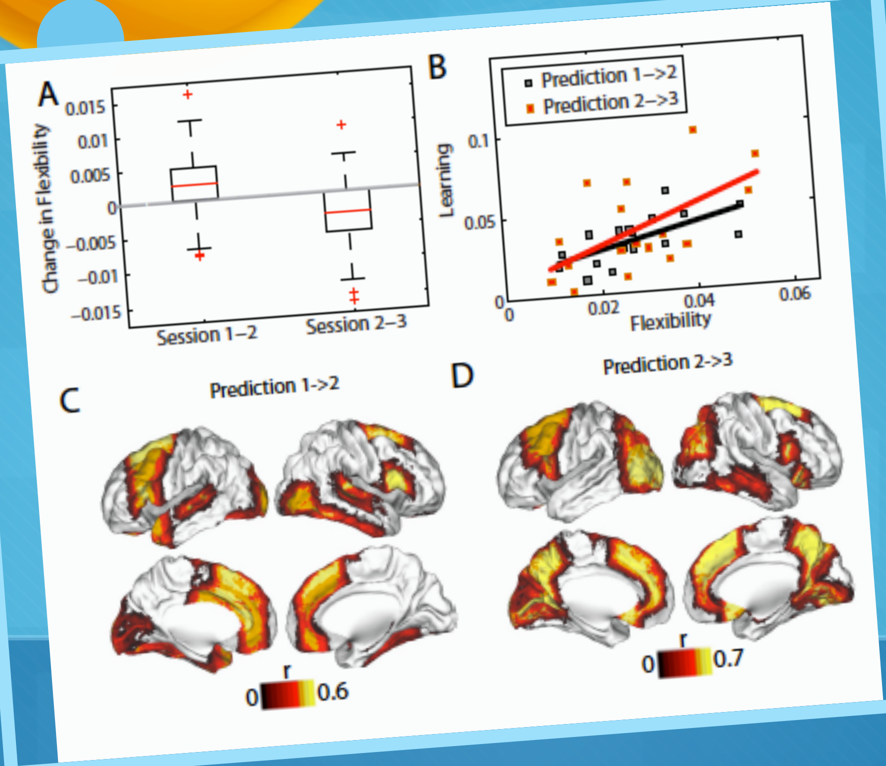


A. S. Waugh, L. Pei, J. H. Fowler, PJM, and MAP [2010], arXiv:0907.3509

Ranking NCAA Football Teams

- T. Callaghan, PJM, & MAP [2007], *American Mathematical Monthly* 114 (9), 761-777
- Effect of communities on rankings obtained by random walk on the game schedule graph
- Important games removed early in Girvan-Newman betweenness-based community detection algorithm



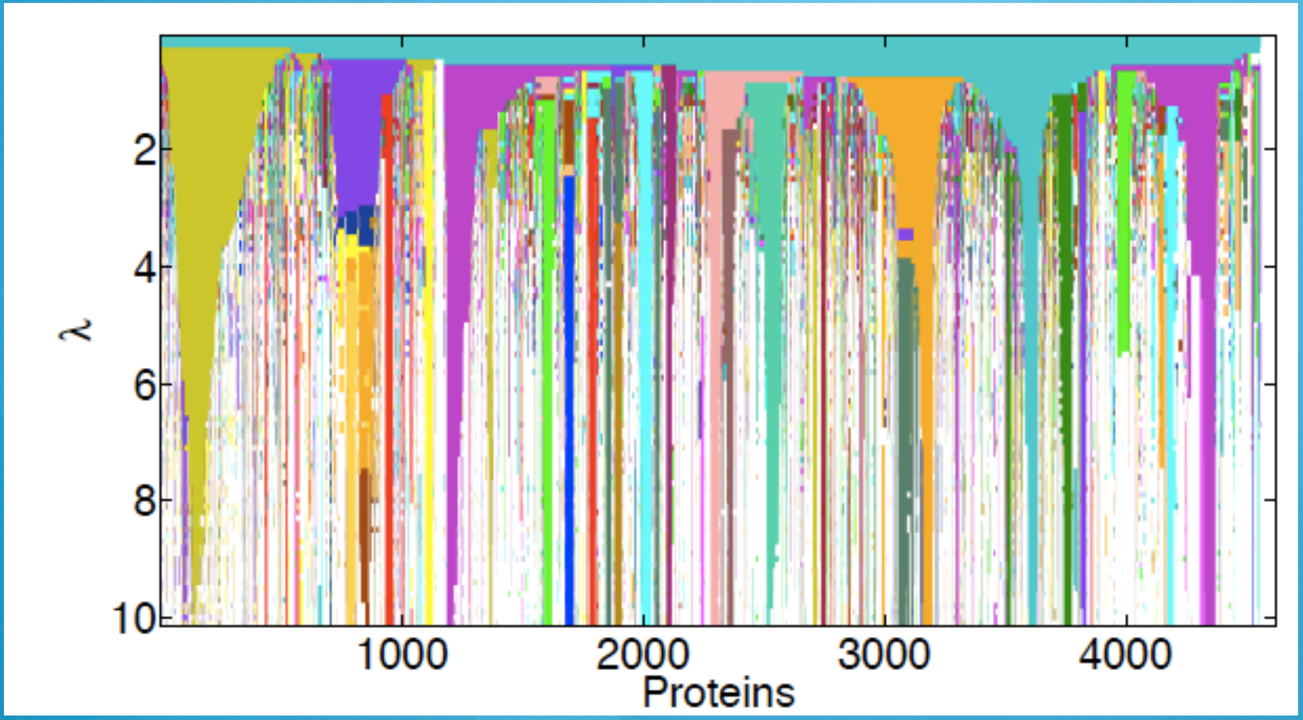


Dynamic Reconfiguration of Human Brain Networks During Learning

D. S. Bassett, N. F. Wymbs, MAP, PJM, J. M. Carlson, and S. T. Grafton, *Proceedings of the National Academy of Sciences,* Vol. 118, No. 18: 7641—7646, 2011.

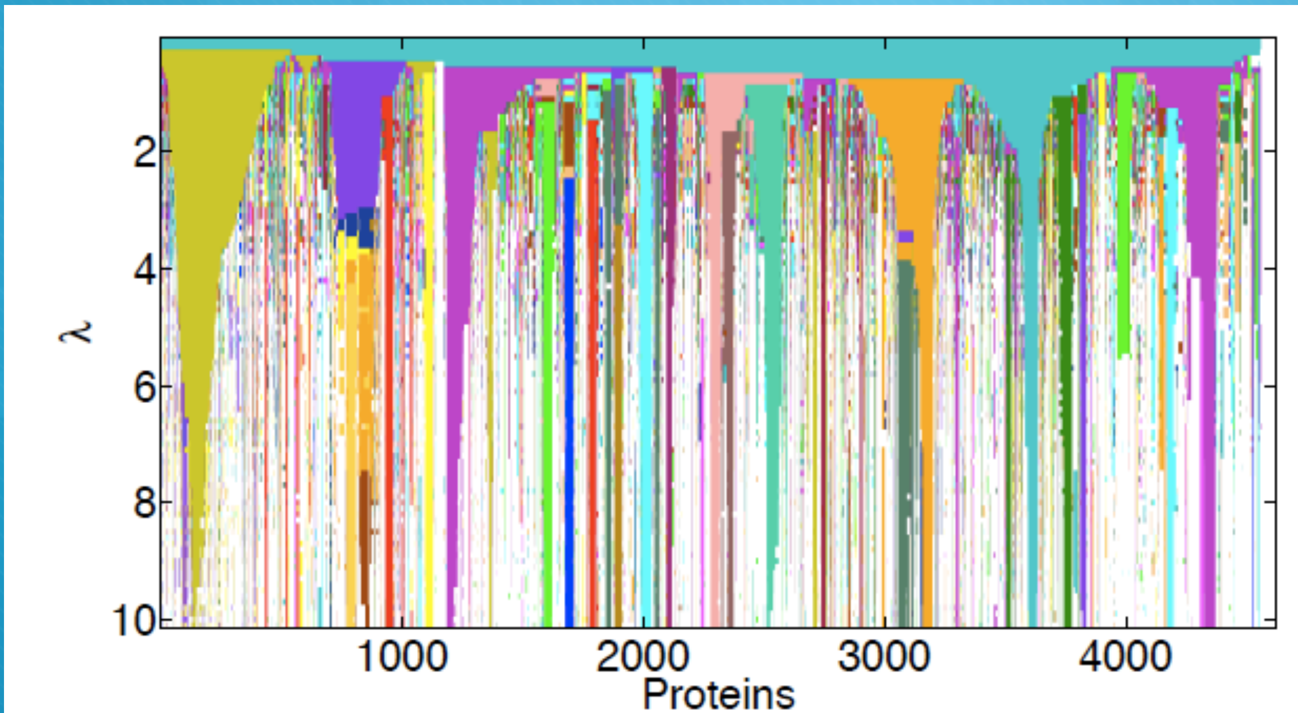
Protein-Protein Interaction Networks

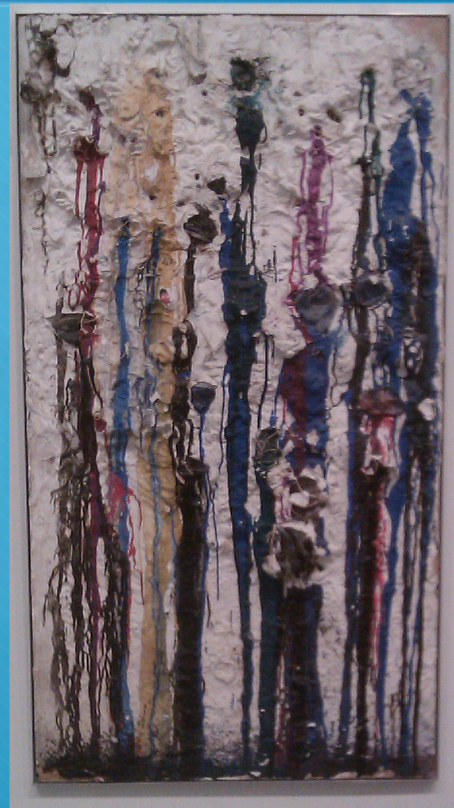
A. C. F. Lewis, NSJ, MAP, & C. M. Deane [2010] *BMC Systems Biology* 4: 100



Protein-Protein Interaction Networks

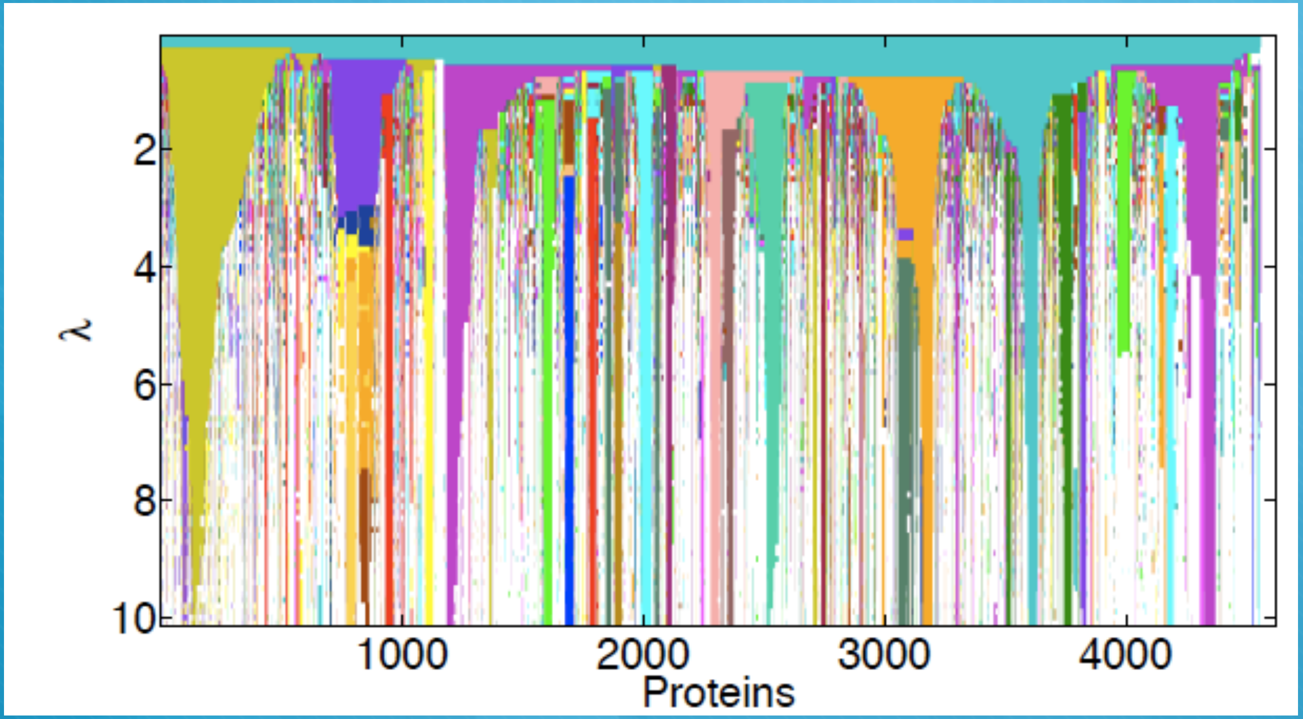
A. C. F. Lewis, NSJ, MAP, & C. M. Deane [2010] *BMC Systems Biology* 4: 100

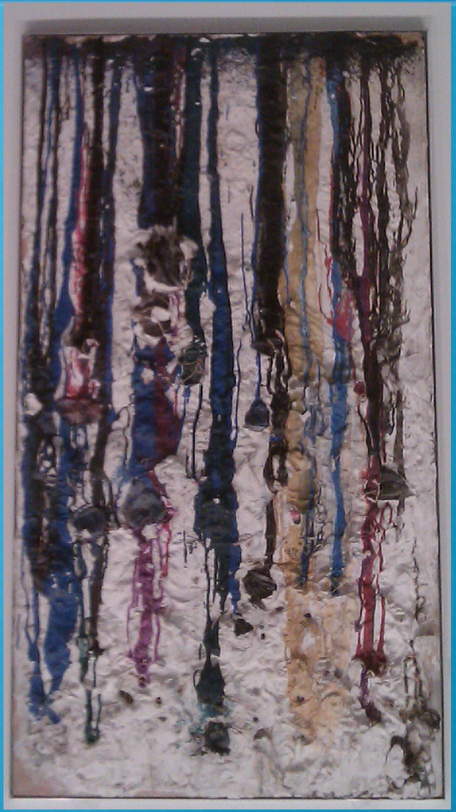




Protein-Protein Interaction Networks

A. C. F. Lewis, NSJ, MAP, & C. M. Deane [2010] *BMC Systems Biology* **4**: 100





Some References

- M. E. J. Newman [2008] *Physics Today* 61(11): 33—38
- S. H. Strogatz [2001] *Nature* 410, 268—276.
- M. E. J. Newman [2010] Networks: An Introduction
- MAP, J.-P. Onnela, and P. J. Mucha [2009] *Notices of the American Mathematical Society* 56(9): 1082—1097, 1164—1166

Conclusions

- **Network science** is a fascinating subject that draws from physics, mathematics, computer science, sociology, biology, and numerous other fields.
- Studying **community structure** networks can lead to interesting insights in a diverse array of applications.
- A **statistical physics** perspective: Start with local information and appropriately coarse-grain/average over things to find global insights.


YOU'RE TRYING TO PREDICT THE BEHAVIOR OF <COMPLICATED SYSTEM>? JUST MODEL IT AS A <SIMPLE OBJECT>, AND THEN ADD SOME SECONDARY TERMS TO ACCOUNT FOR <COMPLICATIONS I JUST THOUGHT OF>.
EASY, RIGHT?
SO, WHY DOES <YOUR FIELD> NEED A WHOLE JOURNAL, ANYWAY?
LIBERAL-ARTS MAJORS MAY BE ANNOYING SOMETIMES, BUT THERE'S *NOTHING* MORE OBNOXIOUS THAN A PHYSICIST FIRST ENCOUNTERING A NEW SUBJECT.