Modularity 5.0 :)

[or, if you prefer, *Modularity version exp(iπ/2)*]

Mason A. Porter

Mathematical Institute, University of Oxford

P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, & J.-P. Onnela, *Science*, Vol. 328, No. 5980, 876-878 (2010)
P. J. Mucha & M. A. Porter, *Chaos*, Vol. 20, No. 4, 041108 (2010)
D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, & S. T. Grafton, PNAS, Vol. 118, No. 18, 7641-7646 (2011)

Data Cake

Written on July 21, 2011 by Mark in Metaphors

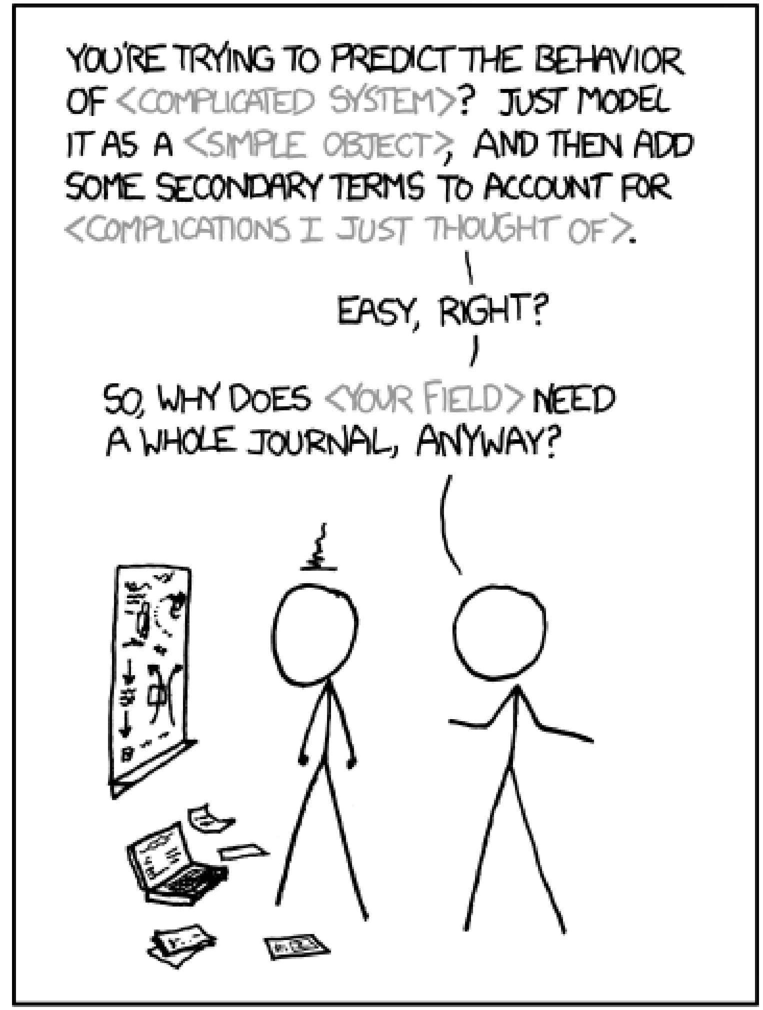
Just a little metaphor I thought you might like...

Goal of the *Science* Paper

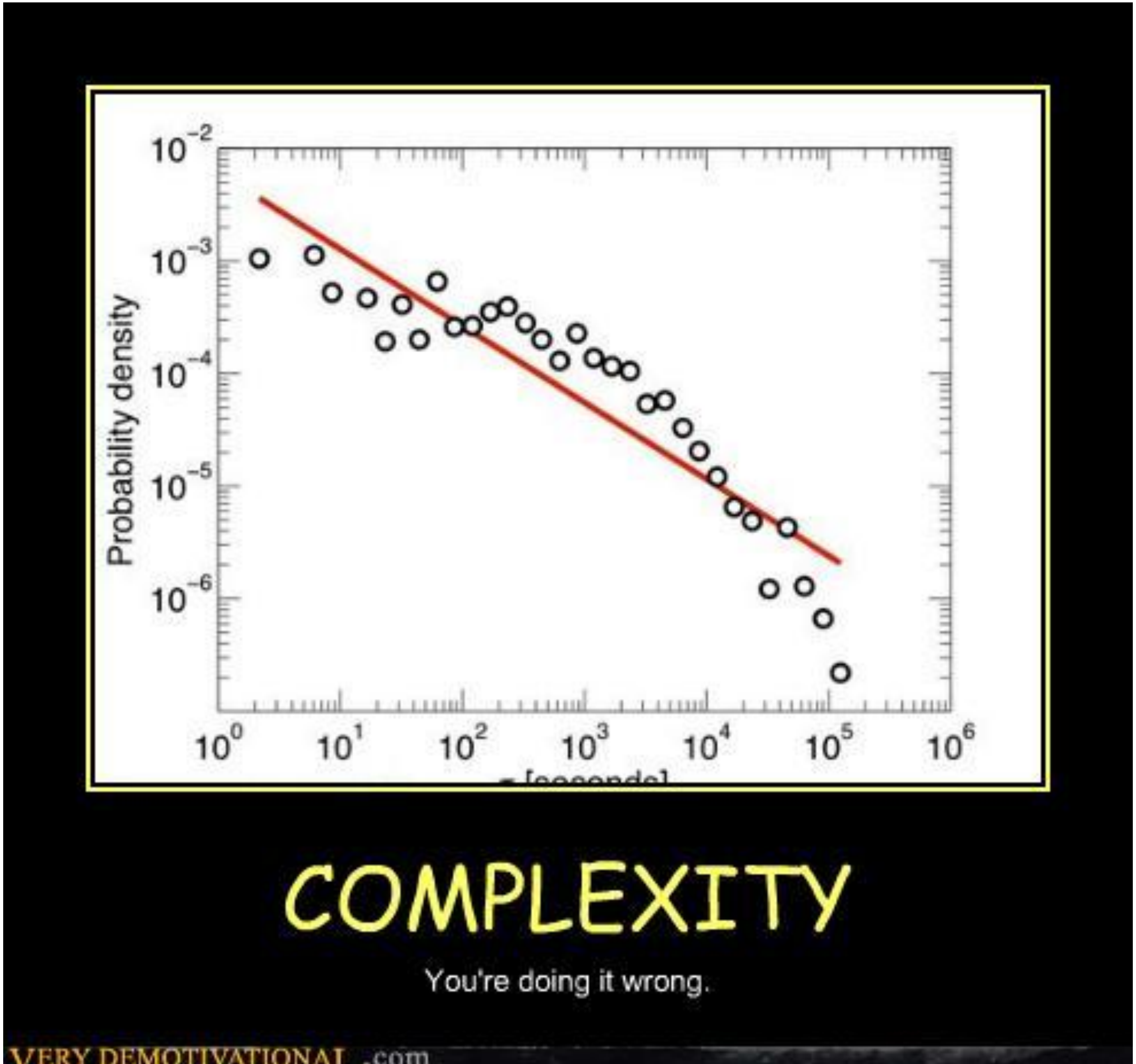
- Extend quality-optimization methods of "community detection" to networks with the following features:
  - *Multiscale:* Consider multiple resolution parameters at once (without sweeping)
  - *Time-Dependent/Longitudinal:* Nodes and edges can change in time
  - *Multiplex:* Multiple types of edges
- ***Goal of this talk:*** Discuss some of the mathematics behind what was used in Bassett et al. (2011).
  - (And I have been asked to wildly speculate a bit, which is a very strange thing for an aggressively cautious person like me to do.)

Outline

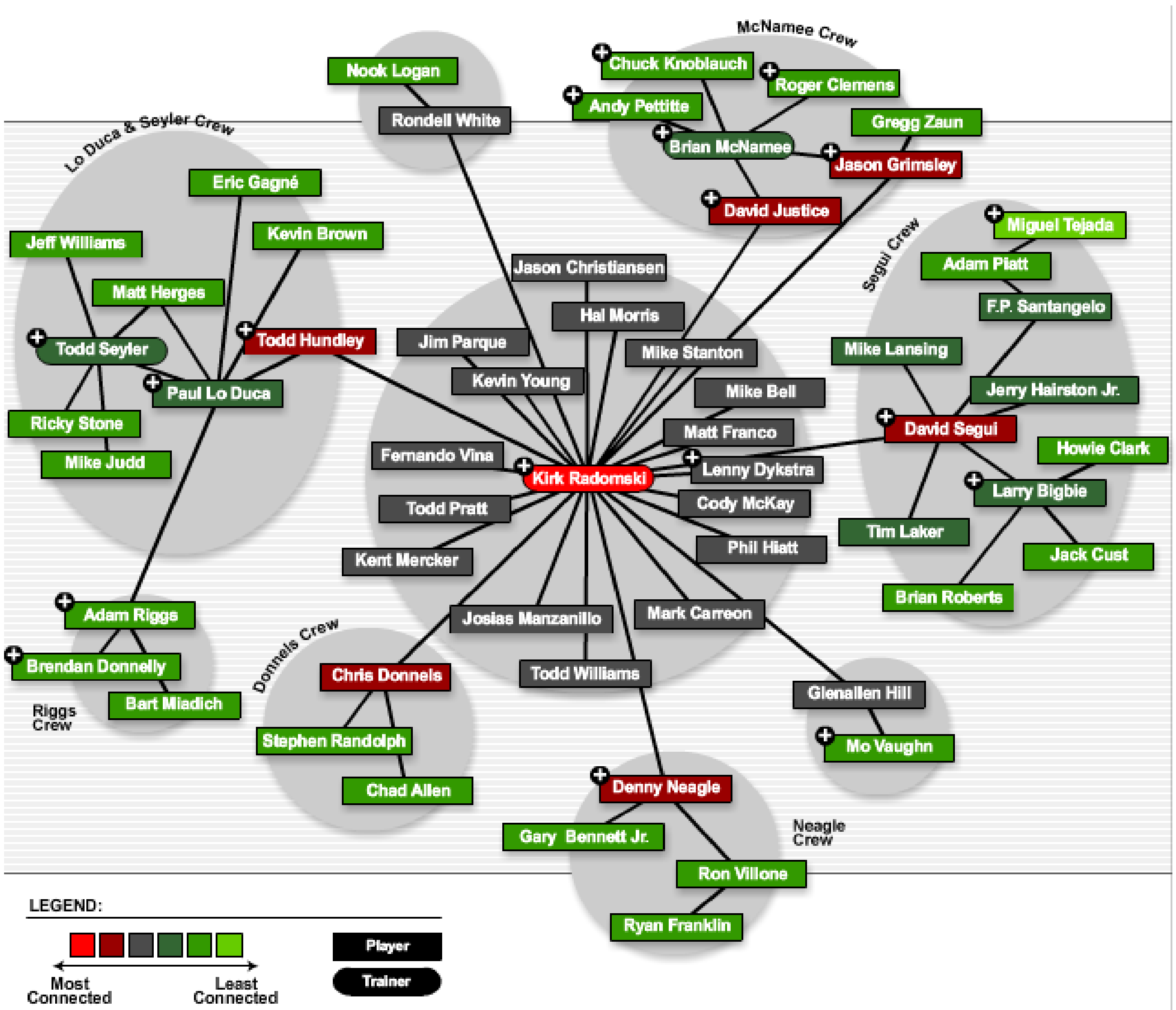
- *Community structure and community detection*
- Multislice networks
- Examples
- Application: fMRI Networks
- Conclusions



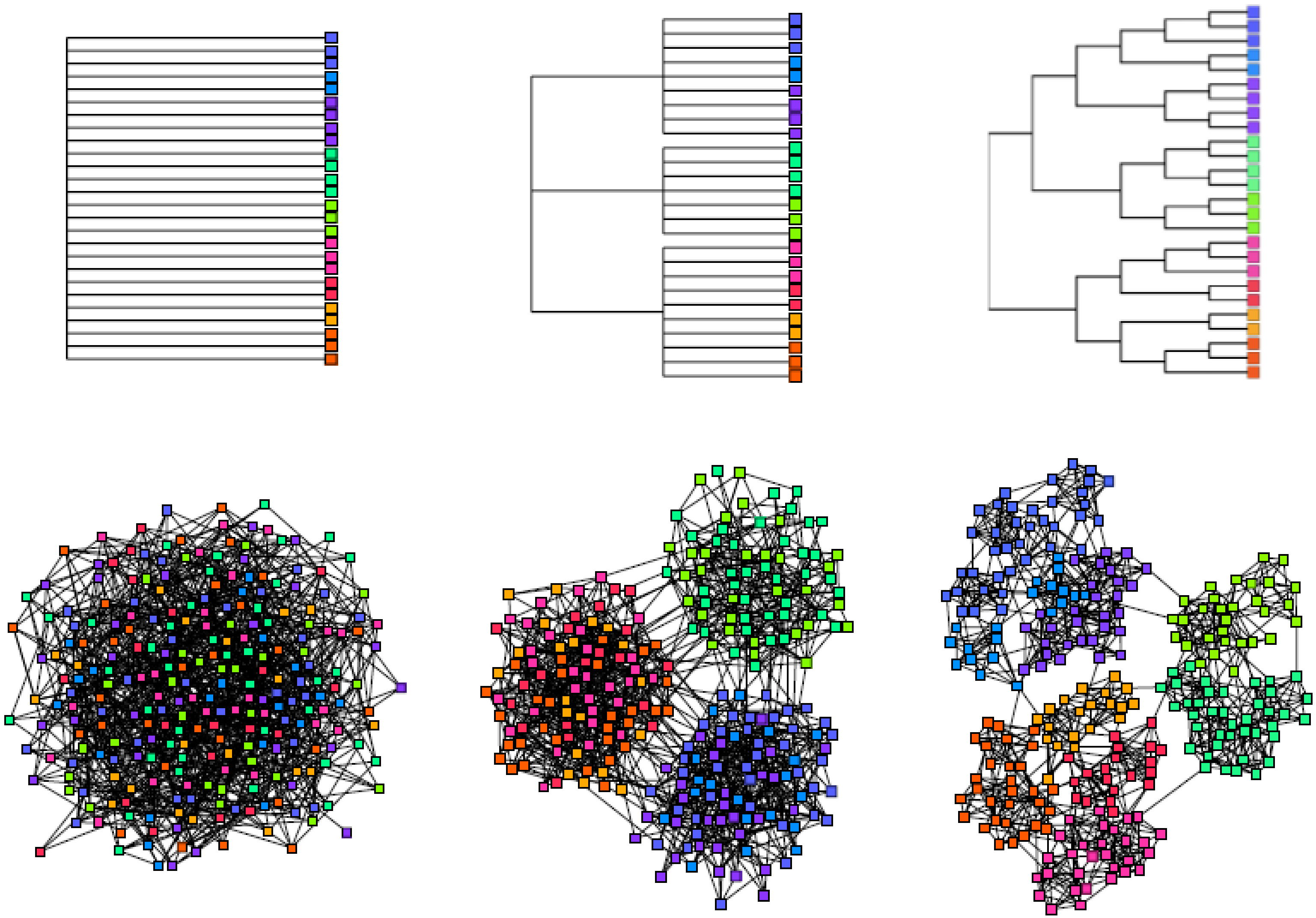
LIBERAL-ARTS MAJORS MAY BE ANNOYING SOMETIMES, BUT THERE'S *NOTHING* MORE OBNOXIOUS THAN A PHYSICIST FIRST ENCOUNTERING A NEW SUBJECT.


Probability density
10^-2
10^-3
10^-4
10^-5
10^-6
10^0
10^1
10^2
10^3
10^4
10^5
10^6
COMPLEXITY
You're doing it wrong.
VERY DEMOTIVATIONAL .com

Community Structure by hand?: Baseball Steroids Networks

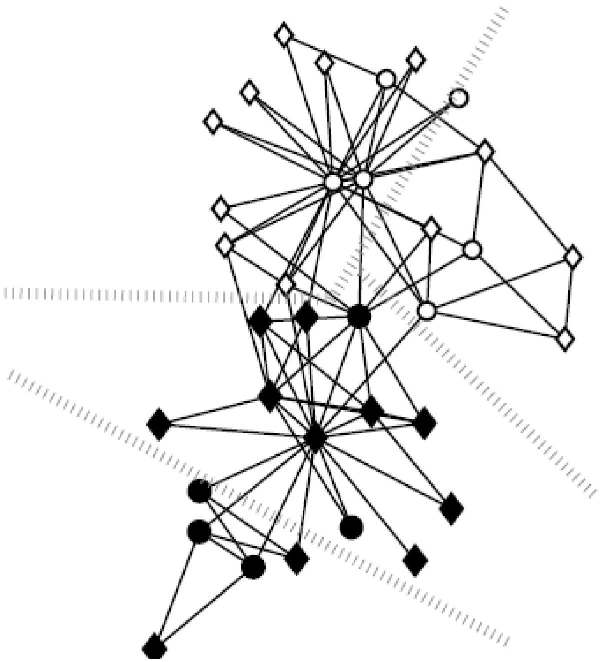


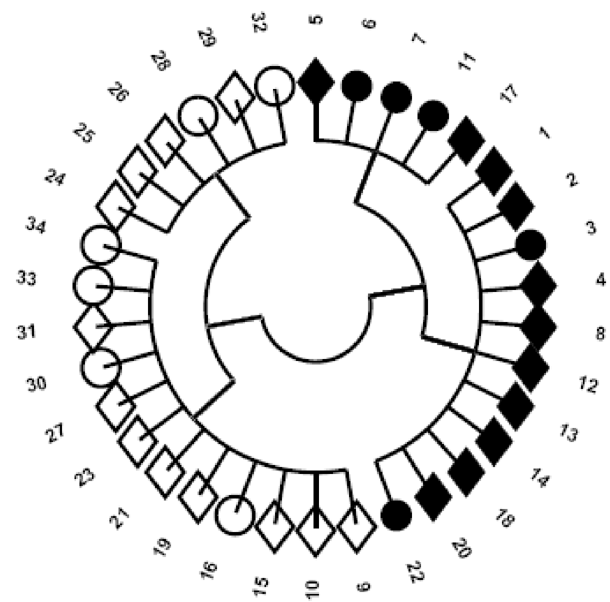
Identifying Communities Algorithmically



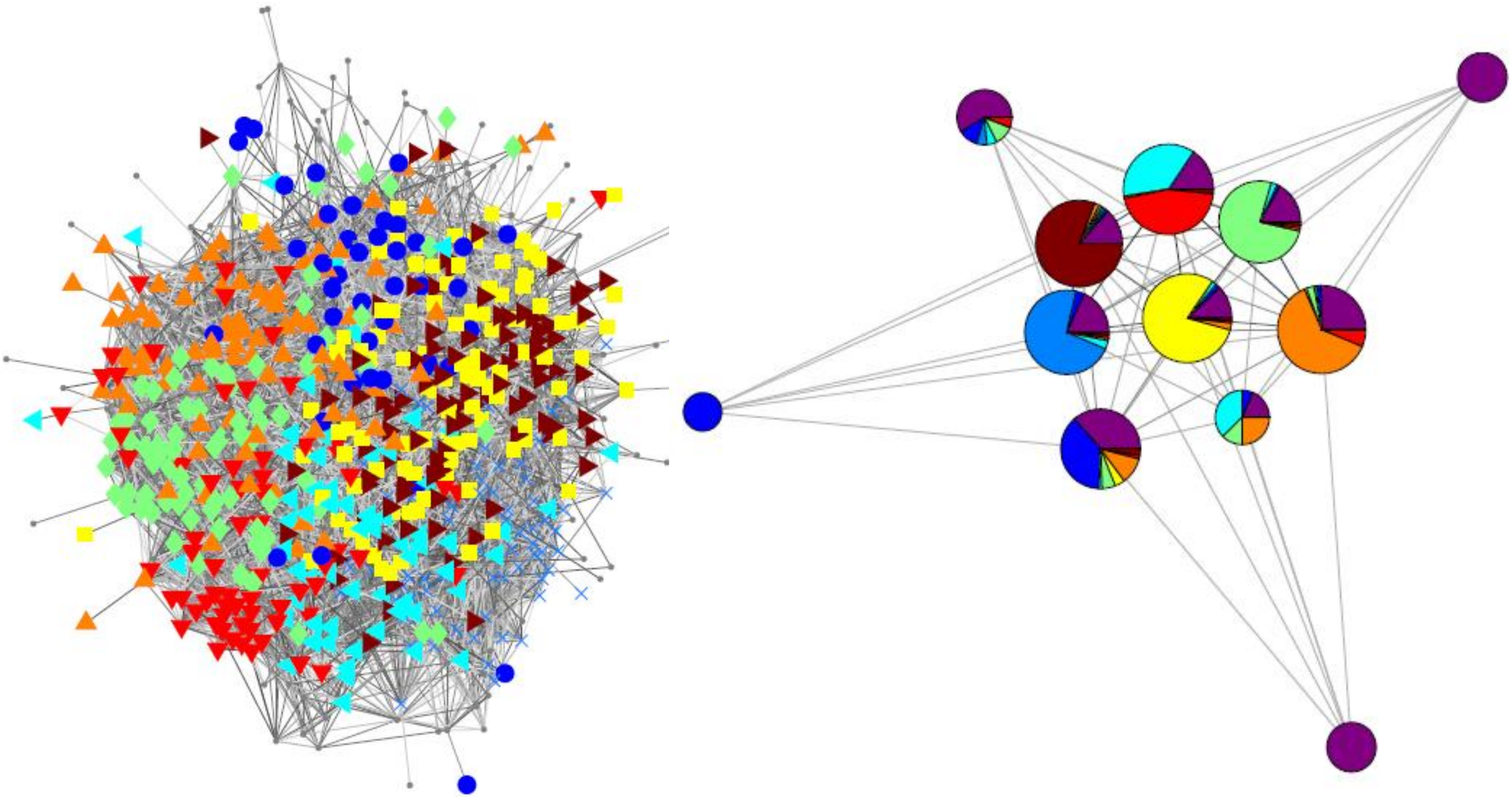
Images from A. Clauset, C. Moore, & M. E. J. Newman (*Nature*, 2008)

Simple (and Infamous) Benchmark: Zachary Karate Club



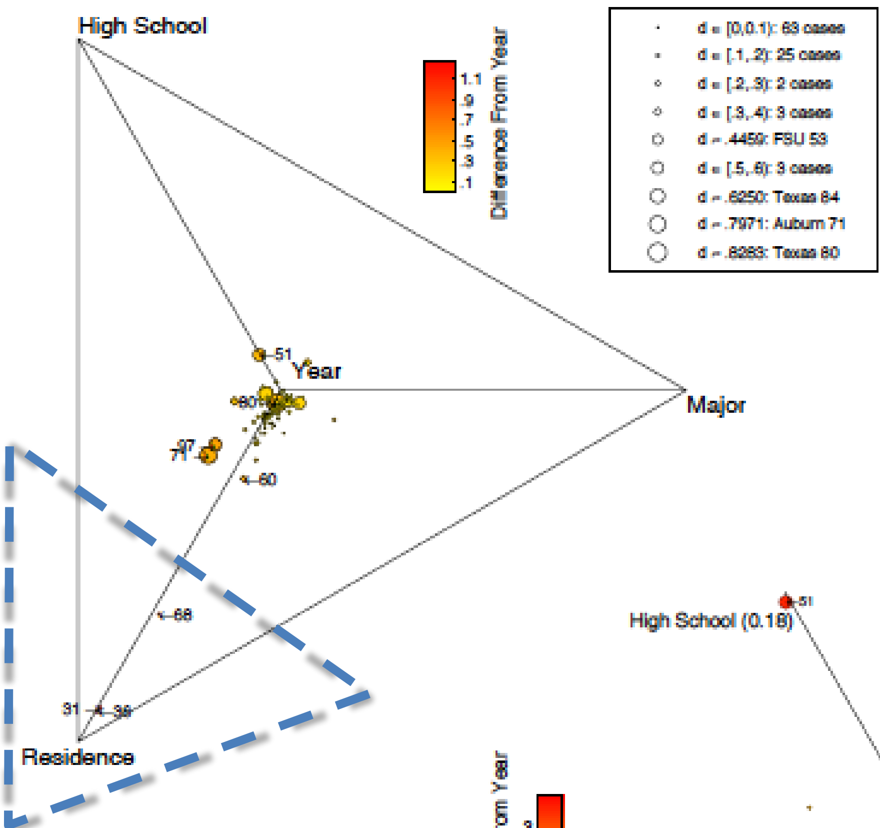


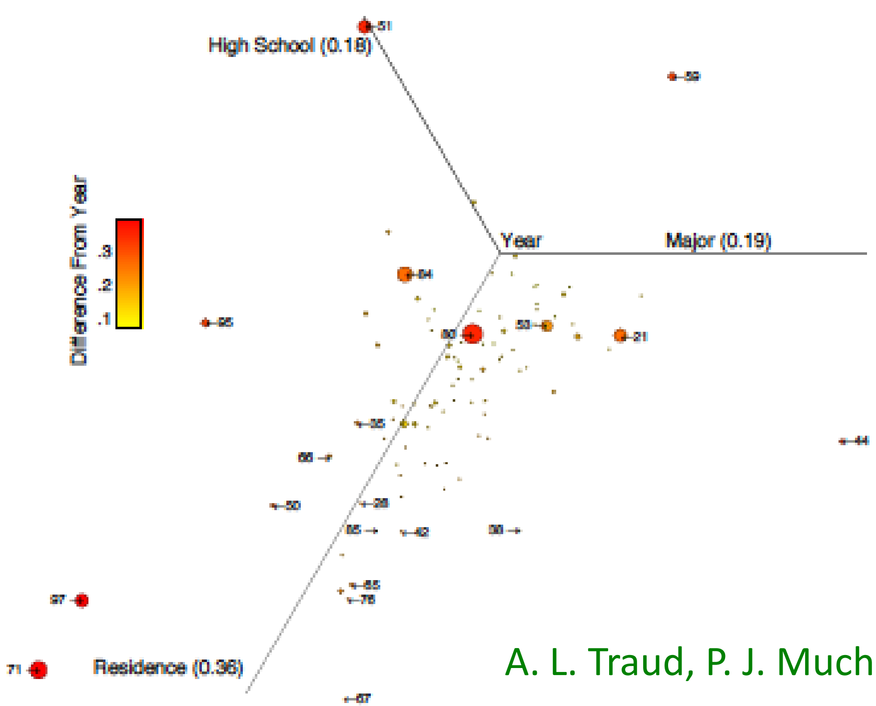
Facebook Friendship Networks



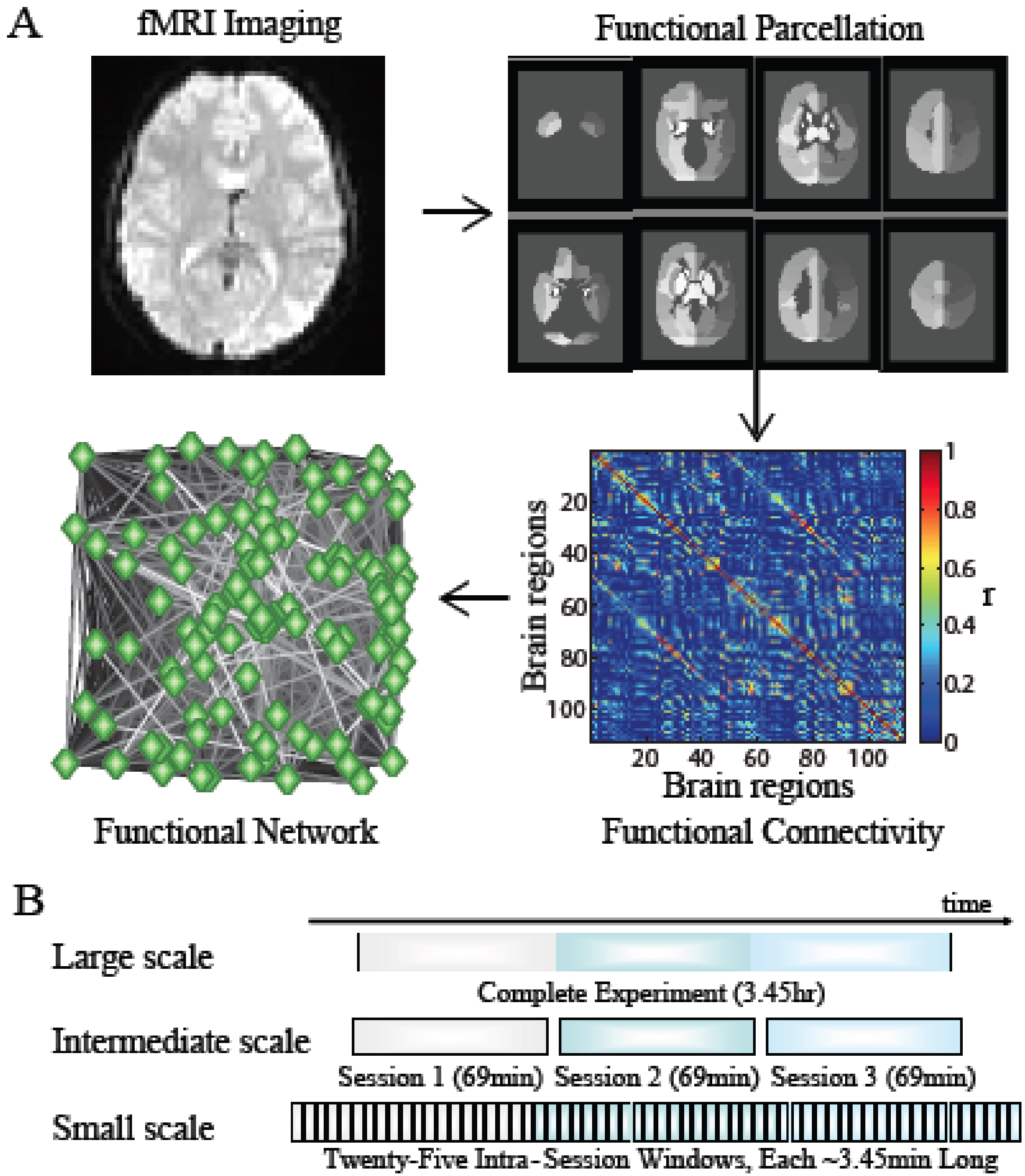
A. L. Traud, E. D. Kelsic, P. J. Mucha, & MAP, *SIAM Review*, Vol. 53, No. 3, 526-543 (2011; arXiv:0809.0960)

Coarse observables rather than specific node assignments to communities





A. L. Traud, P. J. Mucha, & MAP, arXiv: 1102:2166

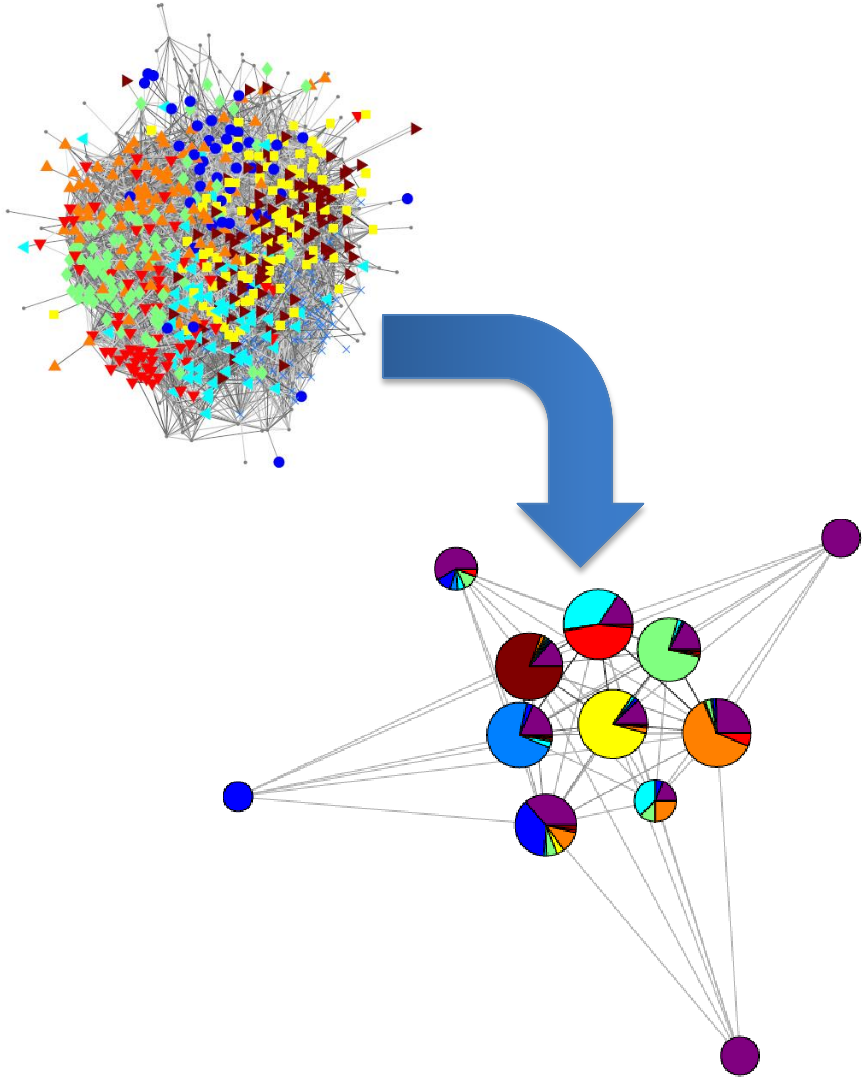

A
fMRI Imaging
Functional Parcellation
Brain regions
20
40
60
80
100
Brain regions
Functional Connectivity
1
0.8
0.6
0.4
0.2
0
r
Functional Network
B
time
Large scale
Complete Experiment (3.45hr)
Intermediate scale
Session 1 (69min)
Session 2 (69min)
Session 3 (69min)
Small scale
Twenty-Five Intra-Session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

Preliminaries

- "Hard/rigid" versus "soft/fuzzy/overlapping" clustering
- A *community* should describe a "cohesive group" of nodes
  - Tons of algorithms available
- Usual notion: more intra-community edges than one would expect at random
  - *But what does "at random" mean?*
- Review articles
  - "Communities in Networks," M. A. Porter, J.-P. Onnela & P. J. Mucha, *Notices of the American Mathematical Society* **56**, 1082-1097 & 1164-1166 (2009).
  - "Community Detection in Graphs," S. Fortunato, *Physics Reports* **486**, 75-174 (2010).

Network Communities

- Communities = Cohesive groups/modules/mesoscopic structures
  - In stat phys, you try to derive macroscopic and mesoscopic insights from microscopic information
- Community structure consists of complicated interactions between modular (horizontal) and hierarchical (vertical) structures
- communities have denser set of Internal links relative to some null model for what links are present at random
  - "Modularity" 5.0



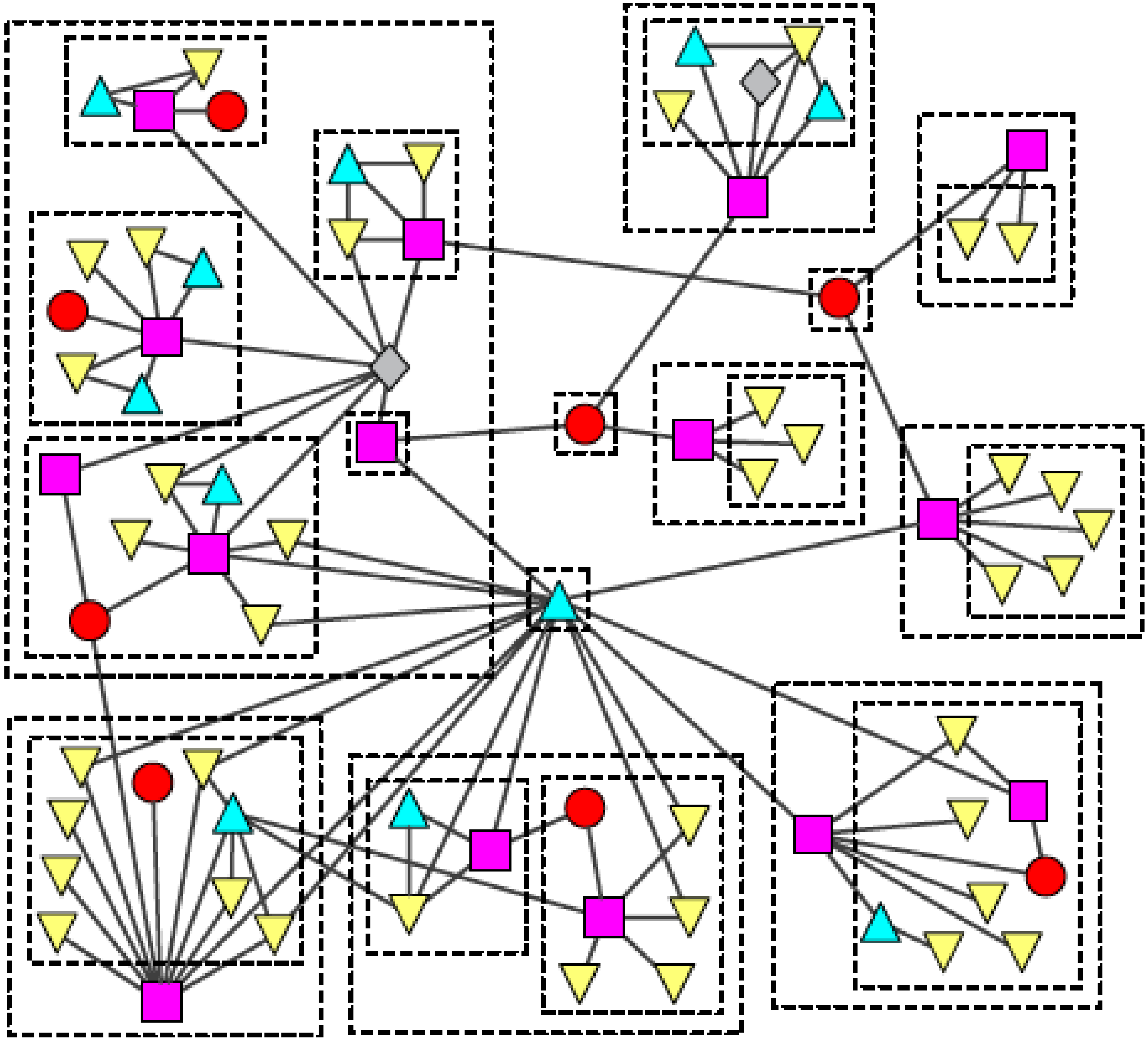


Image from A. Clauset, C. Moore, & M. E. J. Newman (*Nature*, 2008)

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

Quality / Modularity

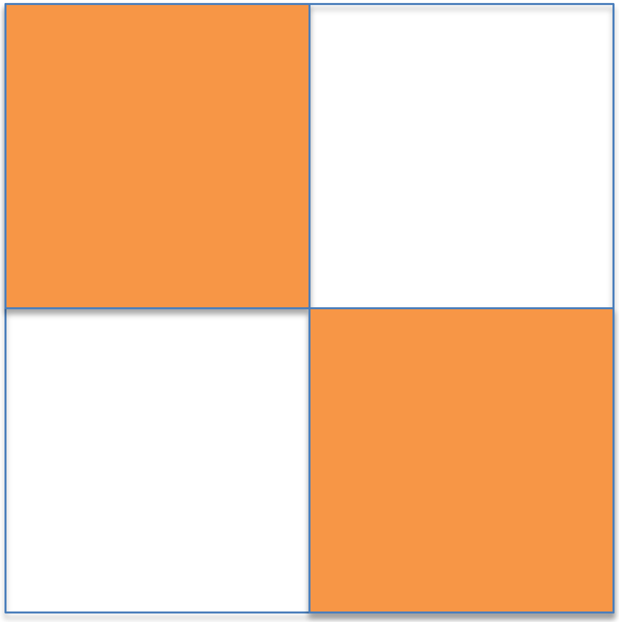
- Popular approach: Use a "modularity" quality function

$$Q = \frac{1}{2W}\sum_{i,j} B_{ij}\delta(C_i, C_j)\,, \quad B_{ij} = A_{ij} - P_{ij}$$

where $\delta(C_i, C_j)$ indicates that the B_{ij} components are only summed over cases in which nodes i and j are classified in the same community. The factor $W = \frac{1}{2}\sum_{ij} A_{ij}$ is the total edge strength in the network (equal to the total number of edges for unweighted networks), where k_i again denotes the strength of node i. In (3.2), P_{ij} denotes the components of a *null model* matrix, which specifies the relative value of intra-community edges in assessing when communities are closely connected [8, 77].

- GOAL: Assign nodes to communities to maximize Q.

Platonic ideal of block structure for "traditional" choice of Q (nested version of this)



- This can be generalized, though vast majority of methods have this in mind…

Example Null Models

(aka: what does “at random” mean?)

- Erdös-Rényi (Bernoulli)

$$P_{ij} = p$$

- Newman-Girvan*

$$P_{ij} = \gamma \frac{k_i k_j}{2W}$$

- Leicht-Newman* (directed)

$$P_{ij} = \gamma \frac{k_i^{in} k_j^{out}}{W}$$

- Barber* (bipartite)

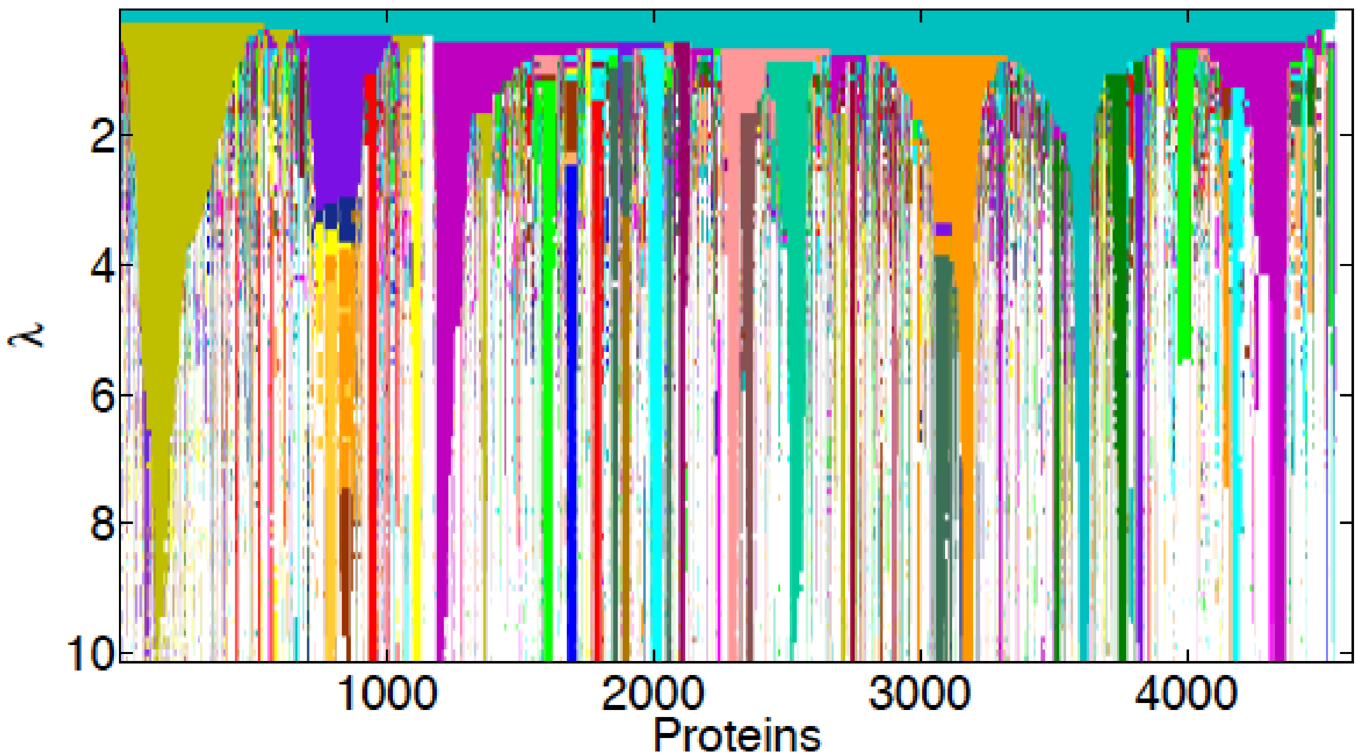
$$P_{ij} = \begin{cases} \gamma \frac{k_i d_j}{W} \\ 0 \end{cases}$$

** With additional resolution parameter γ*

Real Networks: Onion Peeling

Example: Protein-Protein Interaction Networks

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

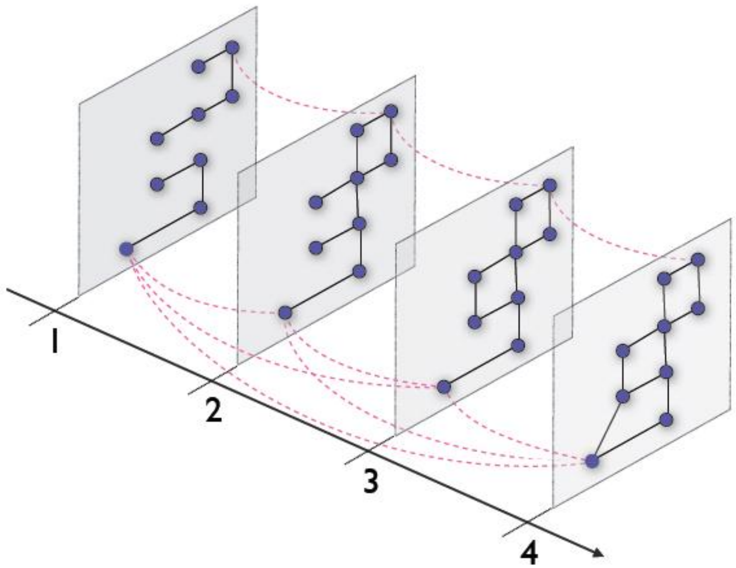


Community Detection: Computational Heuristics

$$Q = \frac{1}{2W} \sum_{i,j} B_{ij} \delta(C_i, C_j)\,, \;\; B_{ij} = A_{ij} - P_{ij}$$

- Cannot guarantee optimal quality without full enumeration of possible partitions
  - NP-hard problem
  - Many algorithms available (simulated annealing, etc.)
  - Need to pick null model appropriate to problem
  - Extreme near-degeneracies in "good" local optima of Q
    - (B. H. Good, Y.-A. de Montjoye, & A. Clauset, *PRE*, 2010)

Multislice Networks



- Typical formulation for studying networks: Static networks, with a single kind of tie, partitioned at a single spatial resolution
  - Also potentially sweep over multiple resolutions (or over multiple static snapshots) but in an ad hoc fashion
- Multislice framework: dynamic, multiplex, and with communities at multiple scales
- Simple idea: Glue common individuals across "slices"

What is an appropriate null model?

$$Q = \frac{1}{2W} \sum_{i,j} B_{ij} \delta(C_i, C_j) , \; B_{ij} = A_{ij} - P_{ij}$$

- Each slice is a network (static, single type) with a specified spatial resolution of interest
- Different slices can mean: different value of resolution parameter, different time snapshot, different type of connection
- Have both intra-slice edges & inter-slice edges
- How to choose a null model?

Quality of Partition via "Stability"

- Idea: use a dynamical process on a network to learn about network structure
  - We build on work of R. Lambiotte, J.-C. Delvenne, & M. Barahona [*arXiv:0812.1770*]
- Quality of a network partition expressed in terms of its "stability" (autocovariance function of an ergodic Markov process on the network):

$$R_{\mathcal{M}}(t) = \sum_{C \in \mathcal{P}} P(C,t) - P(C,\infty)$$

  - P(C,t) = probability, for a given community C, for a random walker to be in that community both initially and at time t
- Stability measures the quality of a partition in terms of the persistence of the dynamics by giving a positive contribution to communities from which a random walker is unlikely to escape with a given time t

Laplacian Dynamics (i.e., random walks)

- Lambiotte, Delvenne, & Barahona [*arXiv:0812.1770*] *derived* modularity from normalized Laplacian dynamics

$$\dot{p}_i = \sum_j \frac{A_{ij}}{k_j} p_j - p_i, \qquad p_i^* = k_i/2m.$$

$$R_{\mathrm{NL}}(t) = \sum_C \sum_{i,j \in C} \left[\left(e^{t(B-I)} \right)_{ij} \frac{k_j}{2m} - \frac{k_i}{2m} \frac{k_j}{2m} \right]. \qquad B_{ij} = A_{ij}/k_j$$

Expansion of matrix exponential to first-order in t recovers Newman-Girvan modularity with resolution $\gamma = 1/t$.

Question: How do we apply this idea to multislice networks?

Generalized Laplacian Dynamics

a) Calculate (to first order in t) the probability of observing an edge between nodes i and j, *conditional on the type of connection necessary to move j → i*
b) Generalize dynamics to include motion along different types of edges
c) Different spreading weights on different types of edges

Multislice Networks

$$k_{js} = \sum_i A_{ijs},\ c_{js} = \sum_r C_{jsr},\ \ \kappa_{js} = k_{js} + c_{js}.$$

$$\dot{p}_{is} = \sum_{jr} (A_{ijs}\delta_{sr} + \delta_{ij} C_{jsr}) p_{jr}/\kappa_{jr} - p_{is}$$

$$\sum_{ij} \left[(\delta_{ij} + tL_{ij})\, p_j^* - \rho_{i|j} p_j^* \right] \delta(c_i, c_j). \qquad \gamma = 1/t$$

$$p_{jr}^* = \kappa_{jr}/(2\mu), \text{ where } 2\mu = \sum_{jr} \kappa_{jr}.$$

$$\rho_{is|jr} p_{jr}^* = \left[\frac{k_{is}}{2m_s} \frac{k_{jr}}{\kappa_{jr}} \delta_{sr} + \frac{C_{jsr}}{c_{jr}} \frac{c_{jr}}{\kappa_{jr}} \delta_{ij} \right] \frac{\kappa_{jr}}{2\mu}$$

$$Q_{\text{multislice}} = \frac{1}{2\mu} \sum_{ijsr} \left\{ \left(A_{ijs} - \gamma_s \frac{k_{is} k_{js}}{2m_s} \right) \delta_{sr} + \delta_{ij} C_{jsr} \right\} \delta(g_{is}, g_{jr})$$

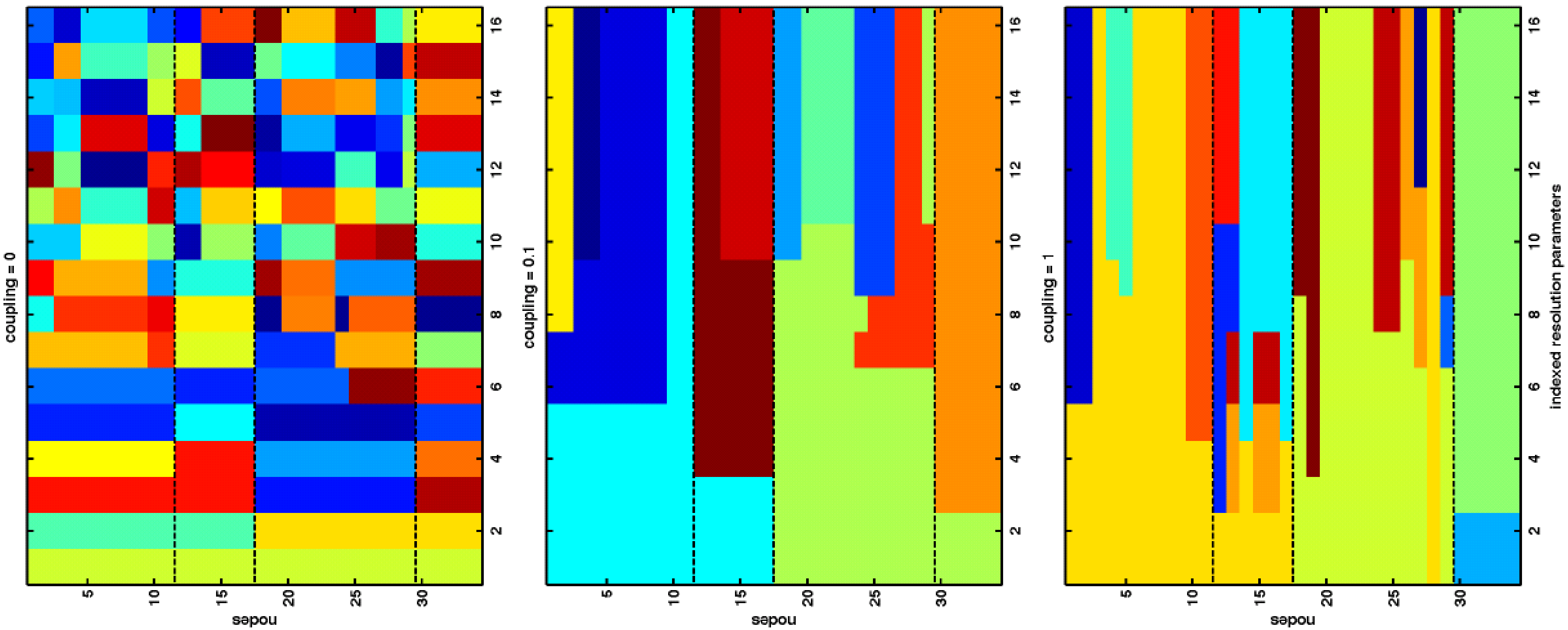
Examples

(proof of concept)

- Zachary Karate Club
  - Multiple resolution parameter values at once ("multiscale")
- Tastes, Ties, & Time
  - Multiplex (multiple edge types)
- 200 years of roll call votes in U.S. Senate
  - Time-dependent

$$Q_{\text{multislice}} = \frac{1}{2\mu} \sum_{ijsr} \left\{ \left(A_{ijs} - \gamma_s \frac{k_{is} k_{js}}{2m_s} \right) \delta_{sr} + \delta_{ij} C_{jsr} \right\} \delta(g_{is}, g_{jr})$$

Zachary Karate Club



$$C_{jsr} = \{0, \omega\}$$

Tastes, Ties, & Time

- Data from Lewis *et al*. 2008
  - "not-Harvard" data set
- First wave of private northeastern school
- Edge types:
  - Facebook friends
  - Picture friends
  - Roommates
  - Housing Groups

ω	#Communities	#Communities per Individual			
		1	2	3	4
0	1036	0	0	0	1640
0.1	122	230	664	611	135
0.2	66	326	805	415	94
0.3	49	430	792	354	64
0.4	36	522	770	302	46
0.5	31	645	695	276	24
1	16	1640	0	0	0

Roll Call Voting Networks

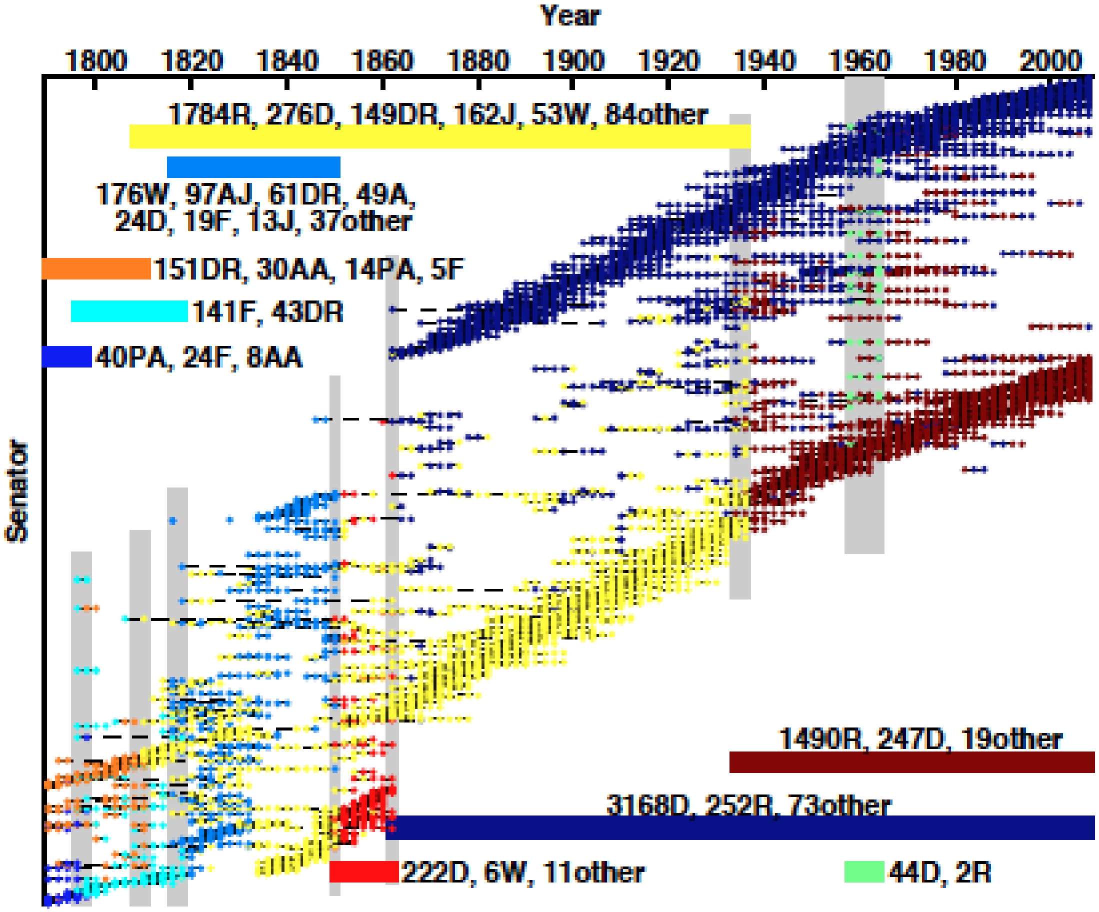
into an $n \times n$ *adjacency matrix* A, with elements $A_{ij} \in [0,1]$ representing the extent of voting agreement between legislators i and j, with elements defined here by

$$A_{ij} = \frac{1}{b_{ij}} \sum_k \alpha_{ijk} \,, \tag{1}$$

where α_{ijk} equals 1 if legislators i and j voted the same on bill k and 0 otherwise and b_{ij} is the total number of bills on which both legislators voted. The matrix A encodes a network of weighted affiliations between legislators, with weights determined by the similarity of their roll-call records

- A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, & M. A. Porter [2011], *arXiv:0907.3509* (without multislice formulation)
- Modularity Q as a measure of polarization
- Can calculate how closely each legislator is tied to their community (e.g., by looking at magnitude of corresponding component of leading eigenvector of modularity matrix if using a spectral optimization method)
- Medium levels of optimized modularity as a predictor of majority turnover
  - By contrast, leading political science measure doesn't give statistically significant indication
- One network slice for each two-year Congress

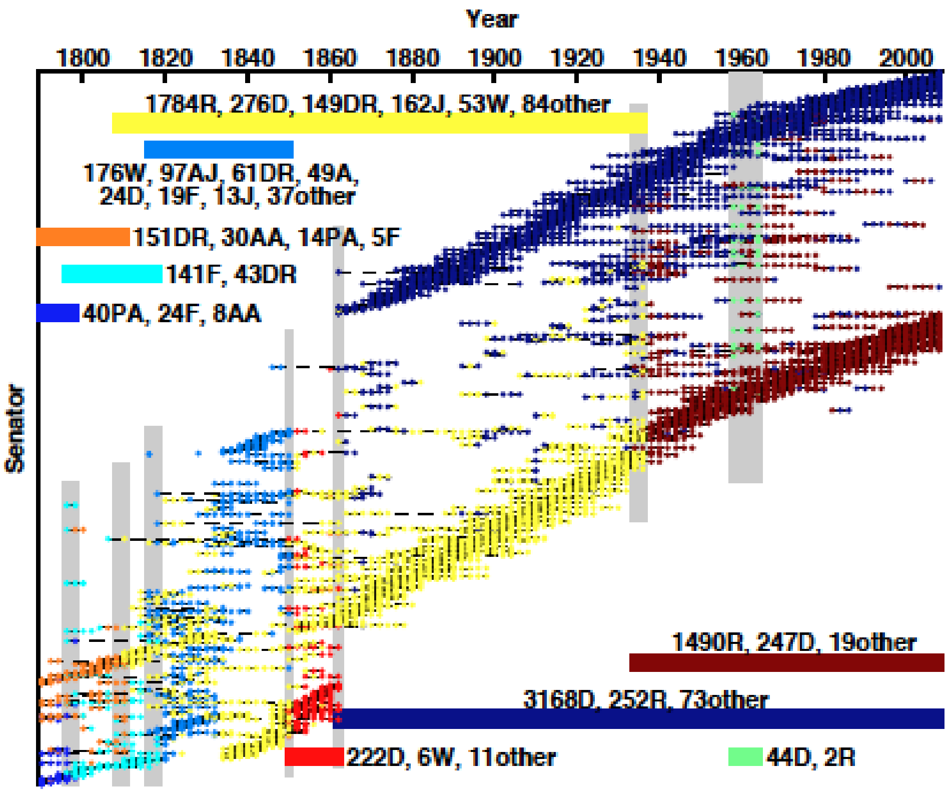
Multislice formulation:
110 Senates (220 years)



Nominal party affiliations:

- Pro-Administration (PA)
- Anti-Administration (AA)
- Federalist (F)
- Democratic-Republican (DR)
- Whig (W)
- Anti-Jackson (AJ)
- Adams (A)
- Jackson (J)
- Democratic (D)
- Republican (R)

110 Senates

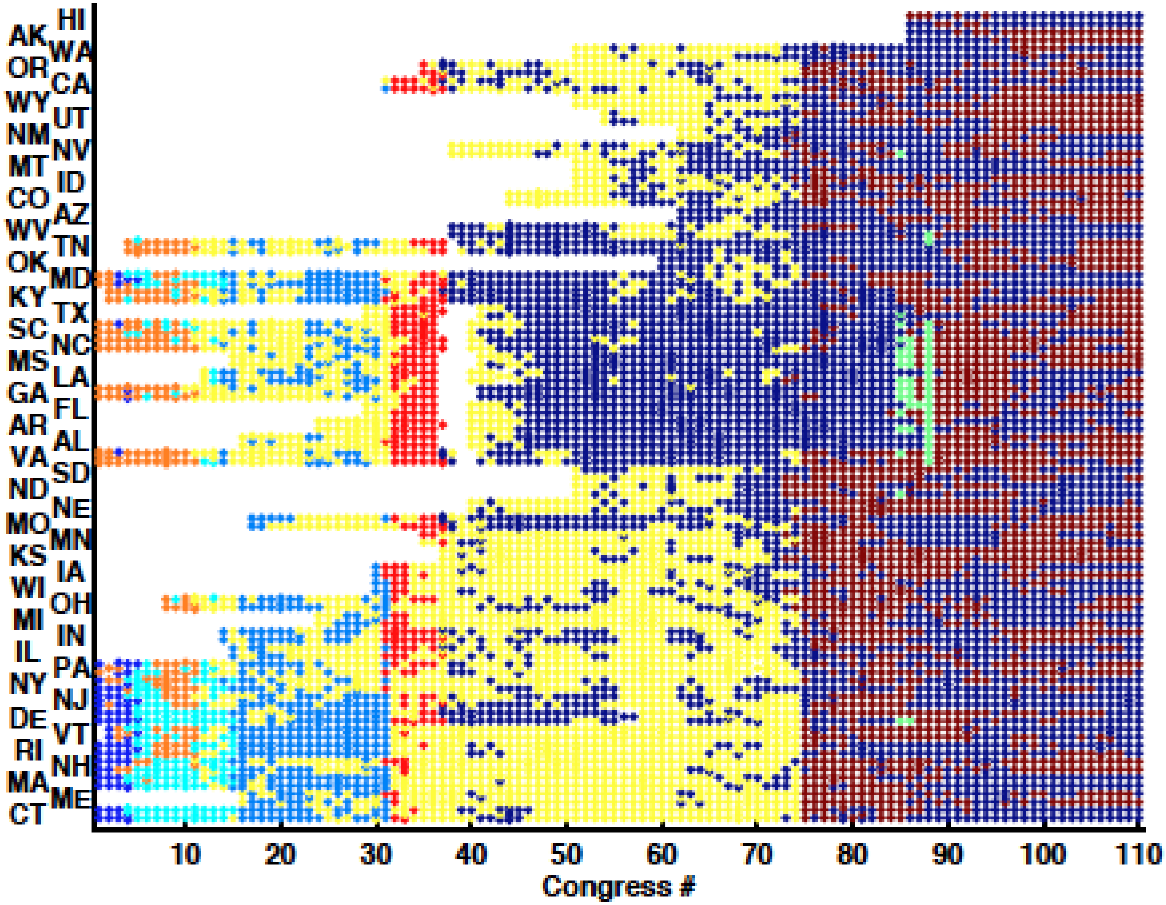


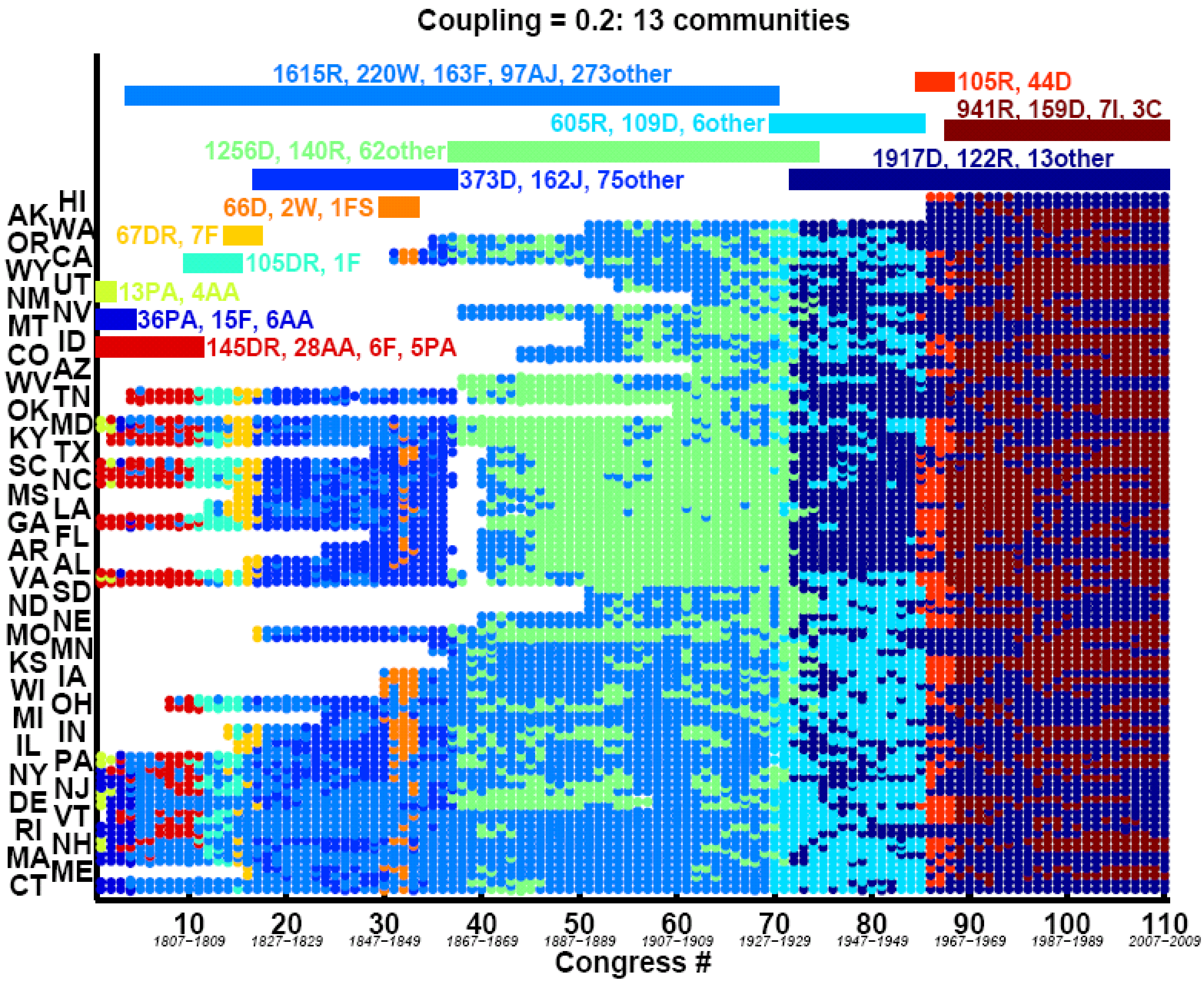
Gray areas:

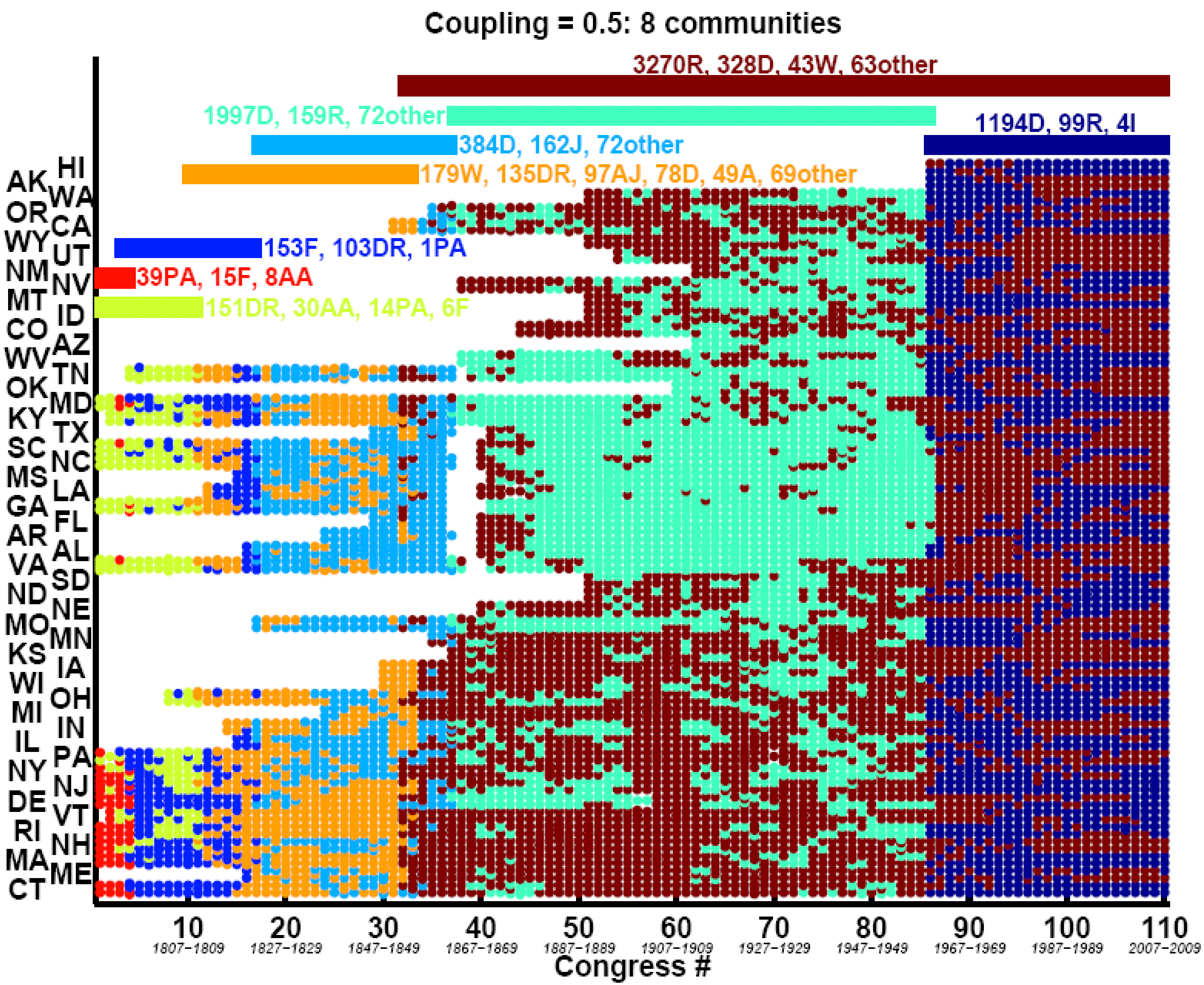
- 4th and 5th: First with political parties
- 10th and 11th: Vice President Aaron Burr's indictment for treason
- 14th and 15th: Changing structures in Democratic-Republican party
- 31st: Compromise of 1850
- 37th: Beginning of the American Civil War
- 73rd and 74th: Landslide 1932 election amidst the Great Depression
- 85th to 88th: Brought the major American civil rights acts

Gray areas: 3 communities exist at the same time (9 communities in total; ω = 0.5)

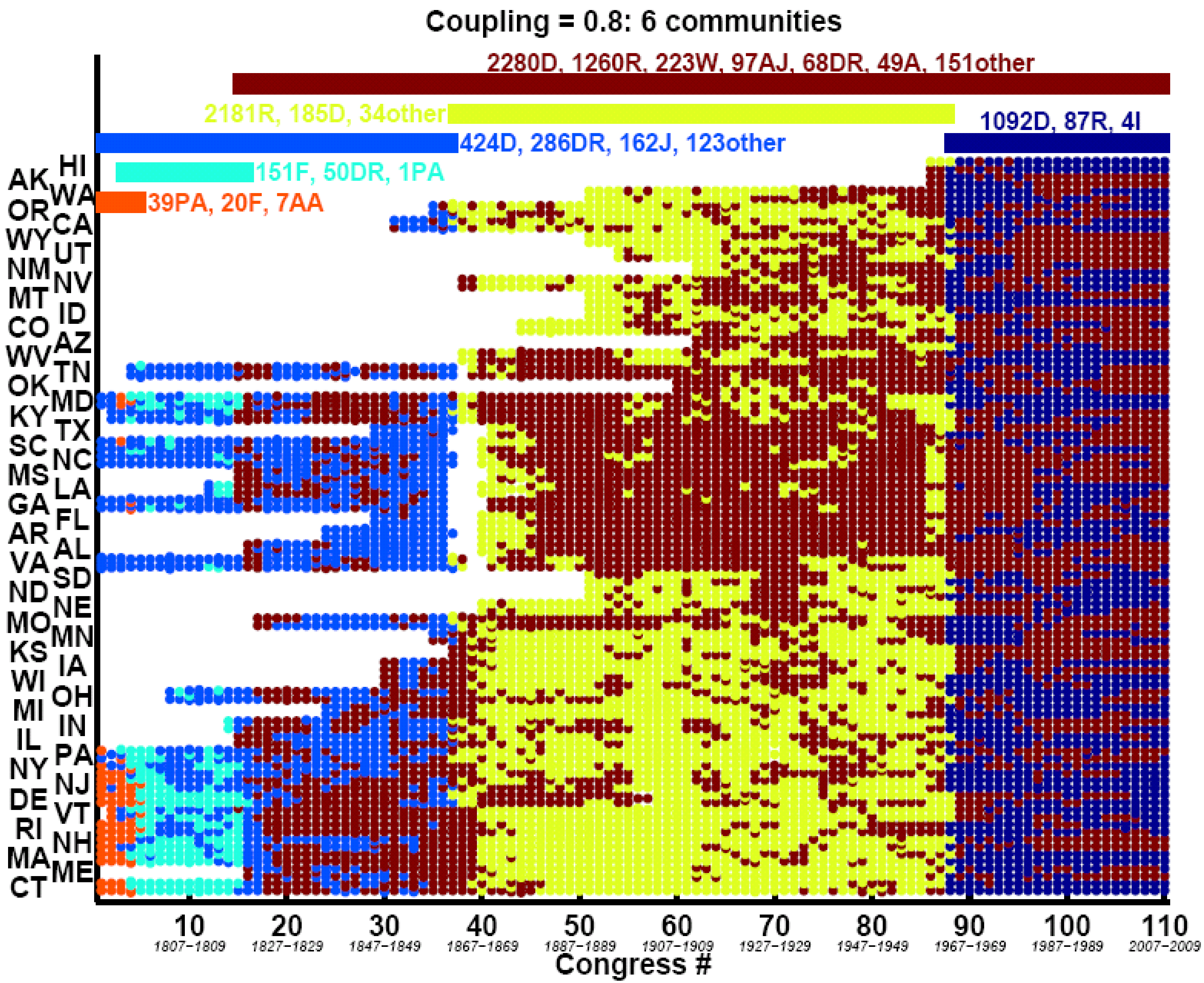
Arranged by state…



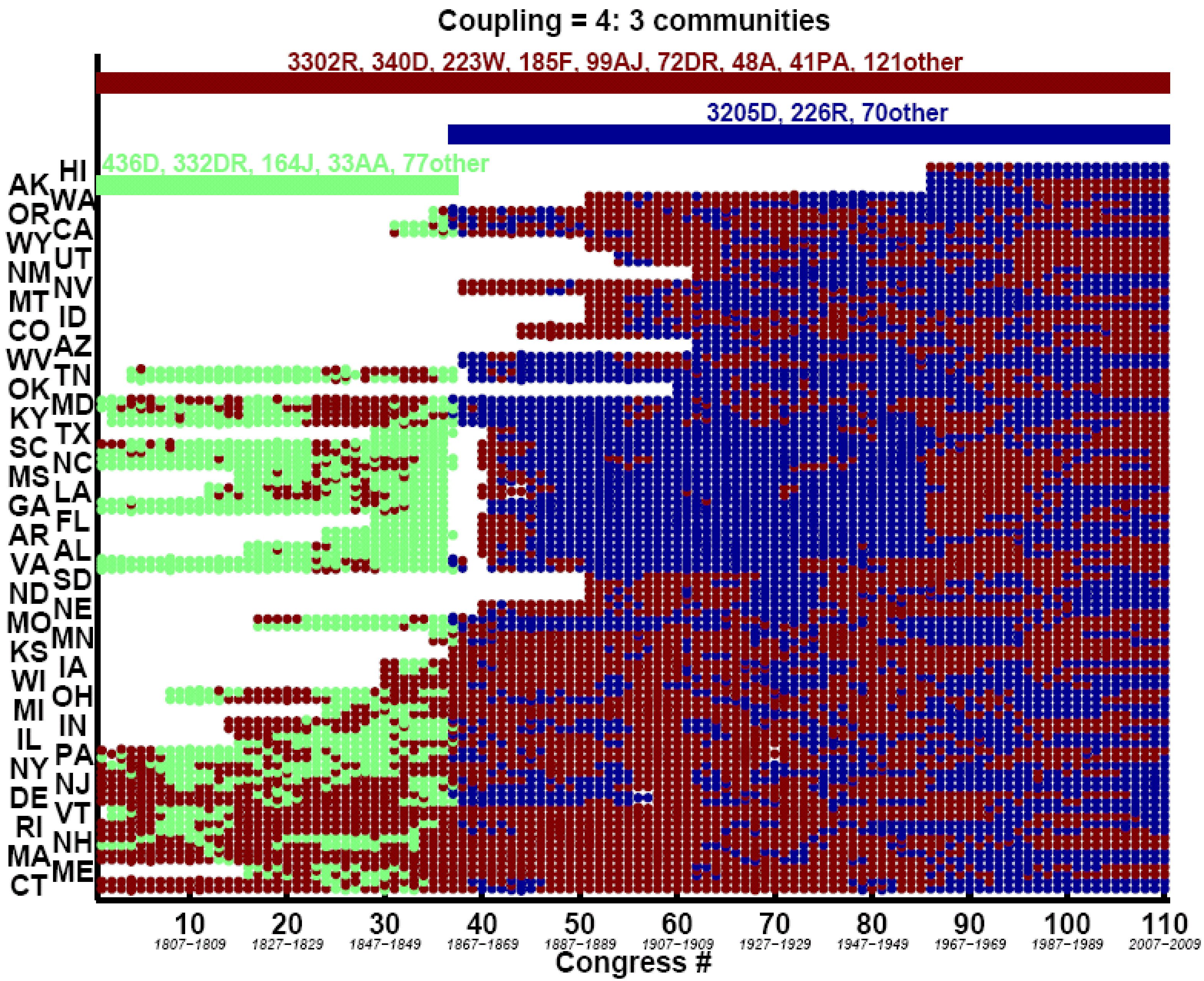

Coupling = 0.2: 13 communities
1615R, 220W, 163F, 97AJ, 273other
105R, 44D
941R, 159D, 7I, 3C
605R, 109D, 6other
1256D, 140R, 62other
1917D, 122R, 13other
373D, 162J, 75other
66D, 2W, 1FS
67DR, 7F
105DR, 1F
13PA, 4AA
36PA, 15F, 6AA
145DR, 28AA, 6F, 5PA
AK
OR
WY
NM
MT
CO
WV
OK
KY
SC
MS
GA
AR
VA
ND
MO
KS
WI
MI
IL
NY
DE
RI
MA
CT
HI
WA
CA
UT
NV
ID
AZ
TN
MD
TX
NC
LA
FL
AL
SD
NE
MN
IA
OH
IN
PA
NJ
VT
NH
ME
10
1807–1809
20
1827–1829
30
1847–1849
40
1867–1869
50
1887–1889
60
1907–1909
70
1927–1929
80
1947–1949
90
1967–1969
100
1987–1989
110
2007–2009
Congress #



Coupling = 0.5: 8 communities
3270R, 328D, 43W, 63other
1997D, 159R, 72other
1194D, 99R, 4I
384D, 162J, 72other
179W, 135DR, 97AJ, 78D, 49A, 69other
153F, 103DR, 1PA
39PA, 15F, 8AA
151DR, 30AA, 14PA, 6F
AK
OR
WY
NM
MT
CO
WV
OK
KY
SC
MS
GA
AR
VA
ND
MO
KS
WI
MI
IL
NY
DE
RI
MA
CT
HI
WA
CA
UT
NV
ID
AZ
TN
MD
TX
NC
LA
FL
AL
SD
NE
MN
IA
OH
IN
PA
NJ
VT
NH
ME
10
1807–1809
20
1827–1829
30
1847–1849
40
1867–1869
50
1887–1889
60
1907–1909
70
1927–1929
80
1947–1949
90
1967–1969
100
1987–1989
110
2007–2009
Congress #



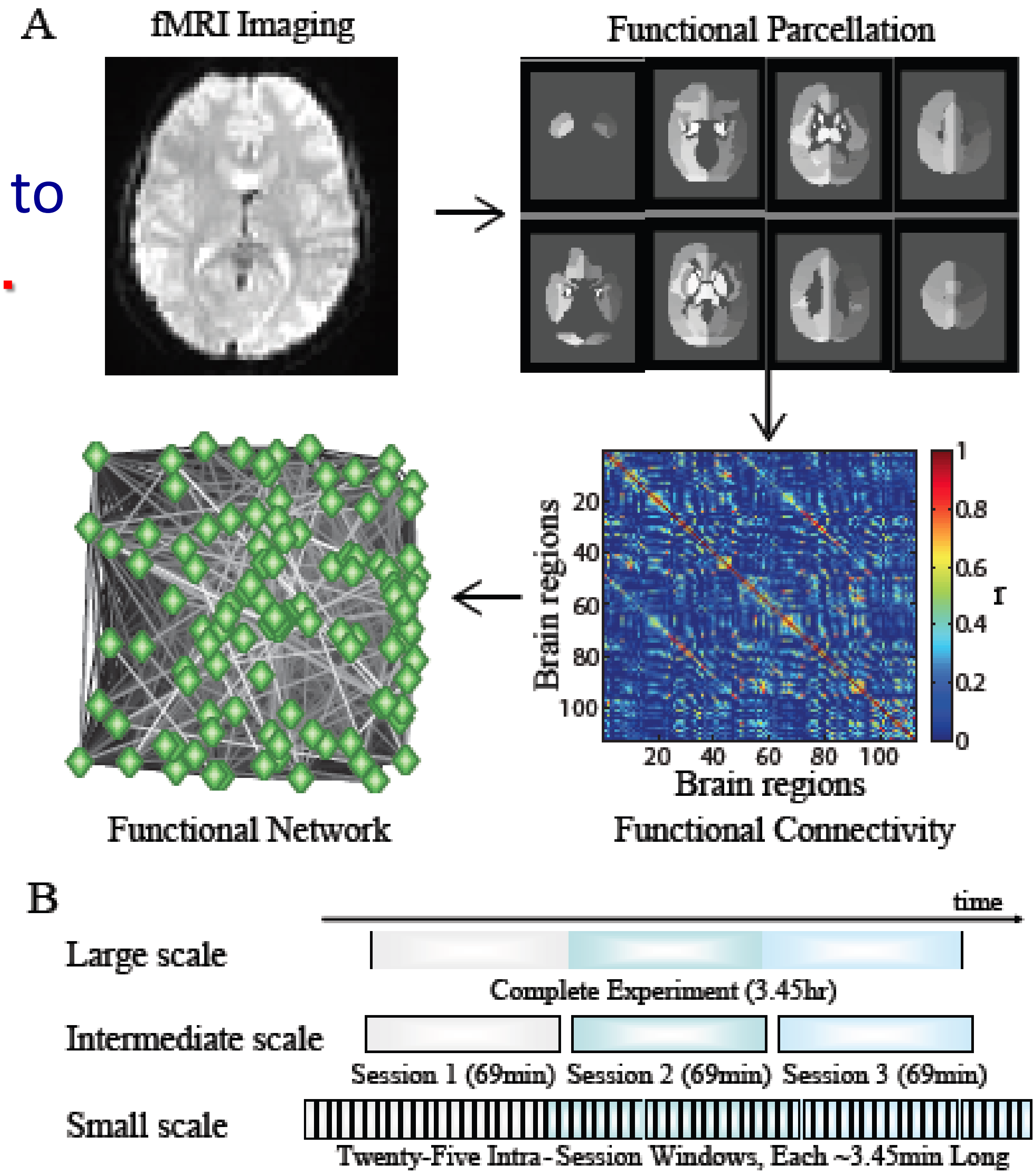
Coupling = 0.8: 6 communities
2280D, 1260R, 223W, 97AJ, 68DR, 49A, 151other
2181R, 185D, 34other
1092D, 87R, 4I
424D, 286DR, 162J, 123other
151F, 50DR, 1PA
39PA, 20F, 7AA
AK HI
OR WA
WY CA
NM UT
MT NV
CO ID
WV AZ
OK TN
KY MD
SC TX
MS NC
GA LA
AR FL
VA AL
ND SD
MO NE
KS MN
WI IA
MI OH
IL IN
NY PA
DE NJ
RI VT
MA NH
CT ME
10 1807–1809
20 1827–1829
30 1847–1849
40 1867–1869
50 1887–1889
60 1907–1909
70 1927–1929
80 1947–1949
90 1967–1969
100 1987–1989
110 2007–2009
Congress #



Coupling = 4: 3 communities
3302R, 340D, 223W, 185F, 99AJ, 72DR, 48A, 41PA, 121other
3205D, 226R, 70other
436D, 332DR, 164J, 33AA, 77other
AK OR WY NM MT CO WV OK KY SC MS GA AR VA ND MO KS WI MI IL NY DE RI MA CT
HI WA CA UT NV ID AZ TN MD TX NC LA FL AL SD NE MN IA OH IN PA NJ VT NH ME
10 20 30 40 50 60 70 80 90 100 110
1807–1809 1827–1829 1847–1849 1867–1869 1887–1889 1907–1909 1927–1929 1947–1949 1967–1969 1987–1989 2007–2009
Congress #

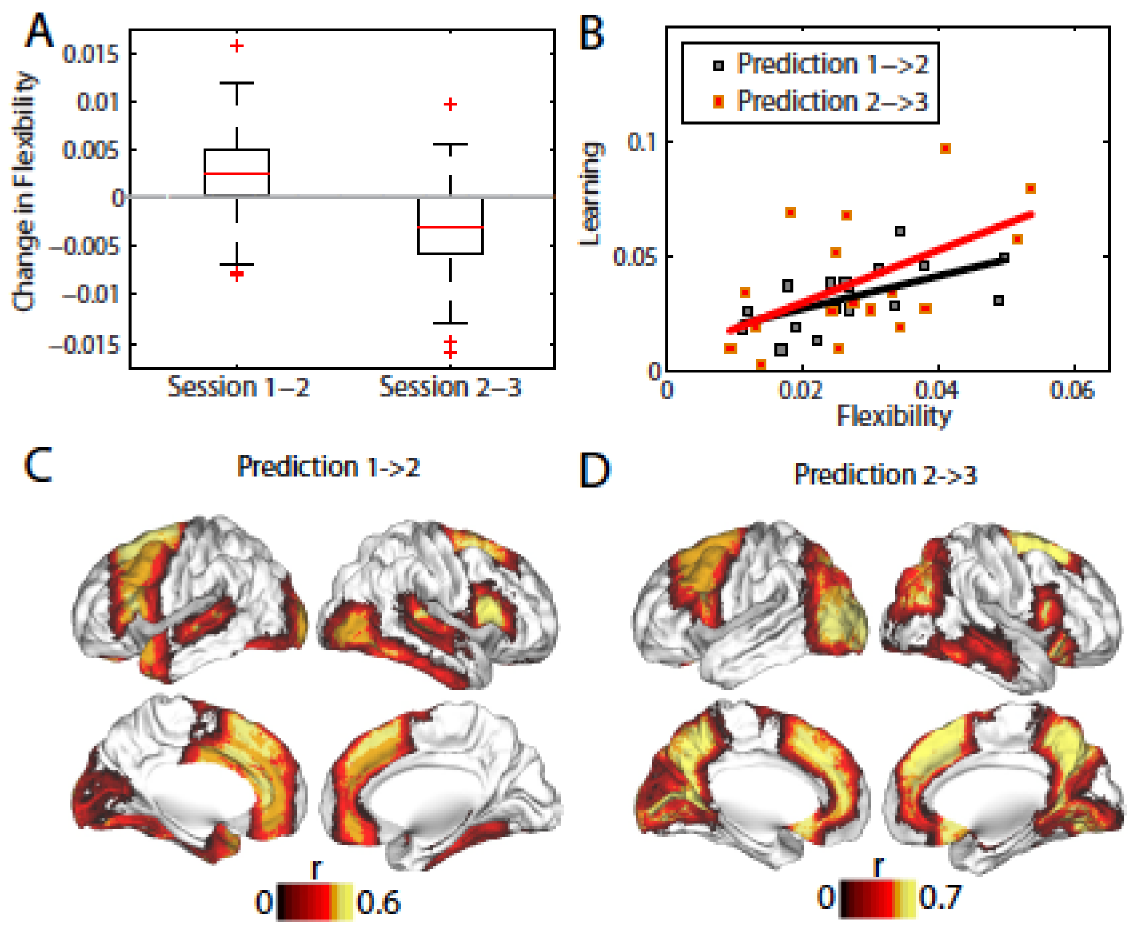
From Congressmen to **Braaaaaaiiiiins…**

- Back to the Bassett et al. (2011) paper that Scott discussed in week 1

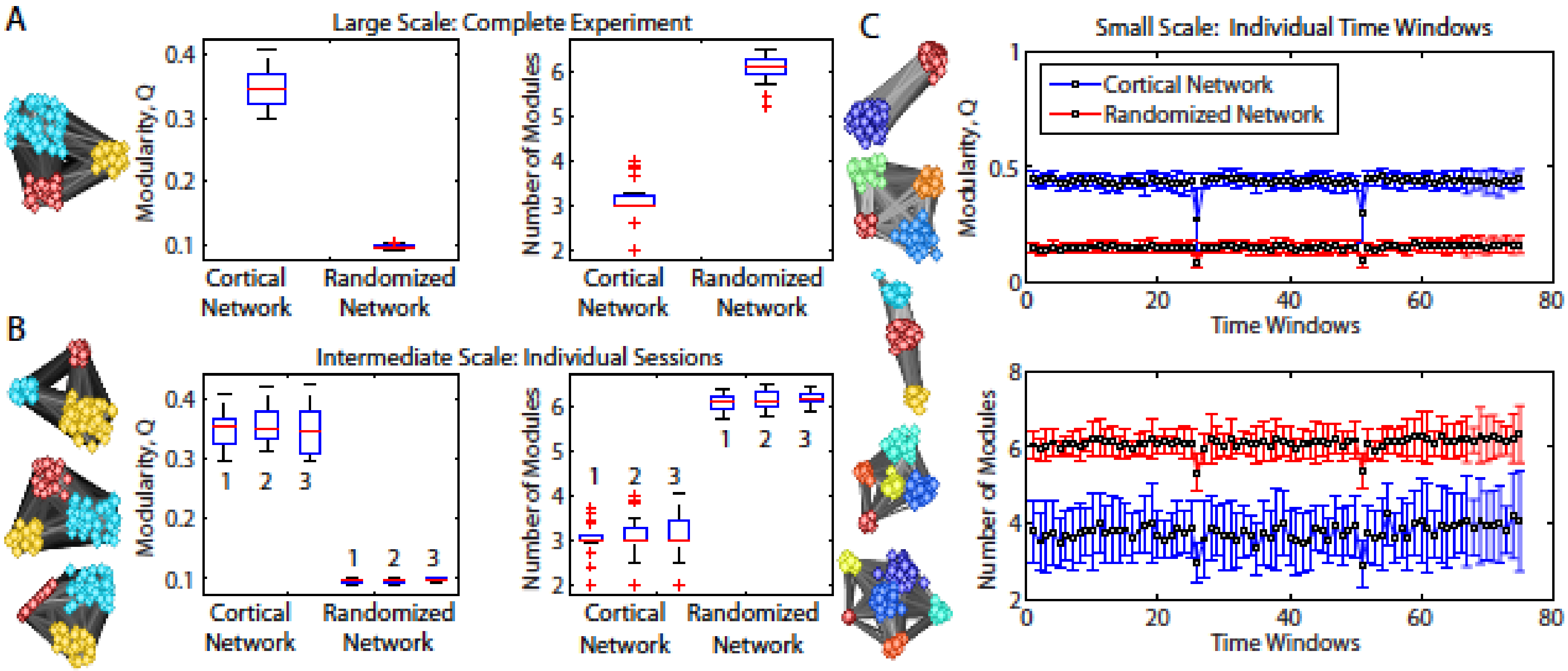


Dynamic Reconfiguration of Human Brain Networks During Learning

- fMRI data: network from correlated time series
- Examine role of modularity in human learning by identifying dynamic changes in modular organization over multiple time scales
- Main result: flexibility, as measured by allegiance of nodes to communities, in one session predicts amount of learning in future sessions

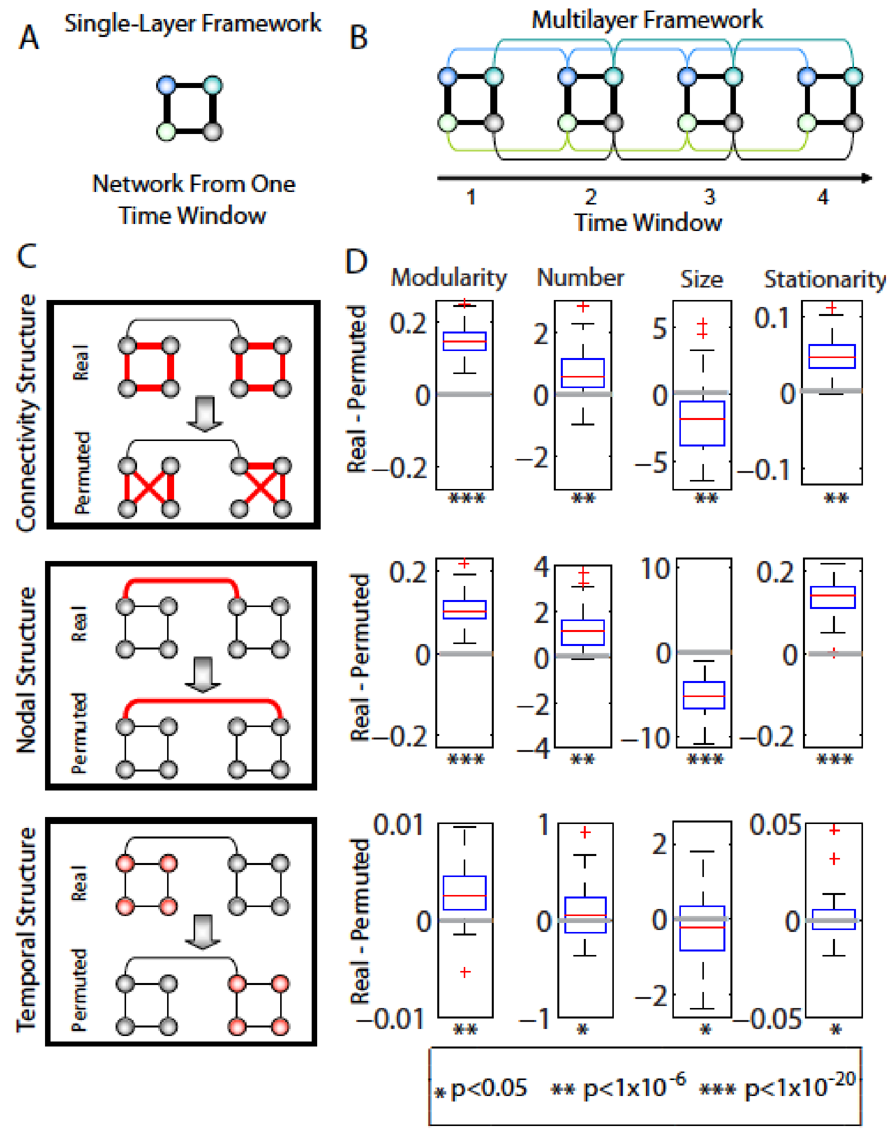


Time Evolution of Static Communities



Dynamic Community Structure

- Investigating community structure in multislice framework requires considering new null models
- Many more details!
  - E.g., Robustness of results to choice of size of time window, size of inter-slice coupling, particular definition of flexibility, complicated modularity landscape (see Good et al, 2010), etc.



Stationarity and Flexibility

- Community stationarity ζ (autocorrelation over time of community membership):

$$U(t, t+m) \equiv \frac{|G(t) \cap G(t+m)|}{|G(t) \cup G(t+m)|} \qquad \zeta \equiv \frac{\sum_{t=t_0}^{t'-1} U(t, t+1)}{t' - t_0 - 1}$$

- Node flexibility:
  - f_i = number of times node i changed communities divided by total number of possible changes
  - Flexibility $f = \langle f_i \rangle$

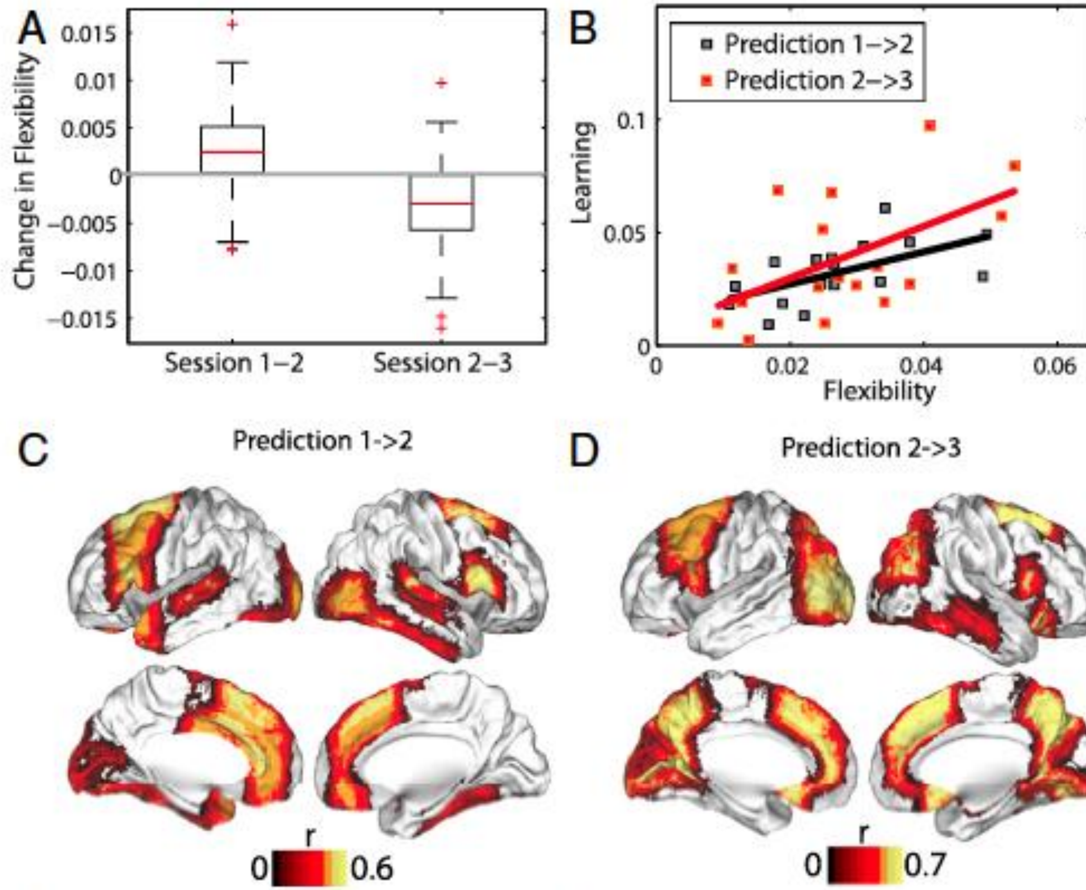


Fig. 4. Flexibility and learning. (*A*) Boxplots showing that the increase in flexibility from experimental session 1 to session 2 was significantly greater than zero (a one-sample *t*-test gives the result $t \approx 6.00$ with $p \approx 2 \times 10^{-8}$), and that the magnitude of the decrease in flexibility from session 2 to session 3 was significantly greater than zero ($t \approx 7.46$, $p \approx 2 \times 10^{-11}$). (*B*) Significant predictive correlations between flexibility in session 1 and learning in session 2 (black curve, $p \approx 0.001$) and between flexibility in session 2 and learning in session 3 (red curve, $p \approx 0.009$). Note that relationships between learning and network flexibility in the same experimental sessions (1 and 2) were not significant; we obtained $p > 0.13$ using permutation tests. (*C*) Brain regions whose flexibility in session 1 predicted learning in session 2 ($p < 0.05$, uncorrected for multiple comparisons). Regions that also passed false-positive correction were the left anterior fusiform cortex and the right inferior frontal gyrus, thalamus, and nucleus accumbens. (*D*) Brain regions whose flexibility in session 2 predicted learning in session 3 ($p < 0.05$, uncorrected for multiple comparisons). Regions that also passed false-positive correction for multiple comparisons were the left intracalcarine cortex, paracingulate gyrus, precuneus, and lingual gyrus and the right superior frontal gyrus and precuneus cortex. In (*C*) and (*D*), colors indicate the Spearman correlation coefficient *r* between flexibility and learning.

Details, Details, Details…

- Checked robustness of findings with respect to…
  - Length of time window
  - Strength of inter-layer edges
  - Ensemble of partitions
  - Definition of flexibility
    - Number of community changes (e.g., 1-2-1-2 is three changes) versus number of distinct communities (e.g., 1-2-1-2 is two distinct groups)
  - …
- ➔Over 10,000 CPU-days of computation
  - Multislice community detection code is now about 50 times faster (with same algorithm).
    - (Should be ready to be send to UCSB soon!)

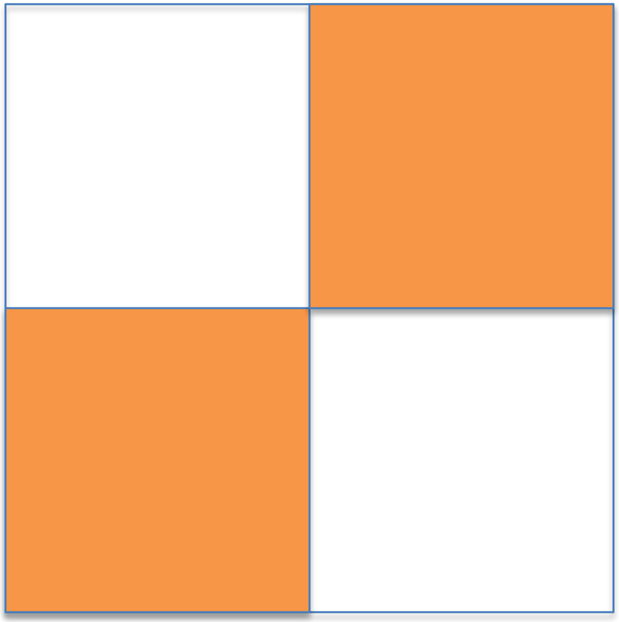
Stepping Back

(and some speculation)

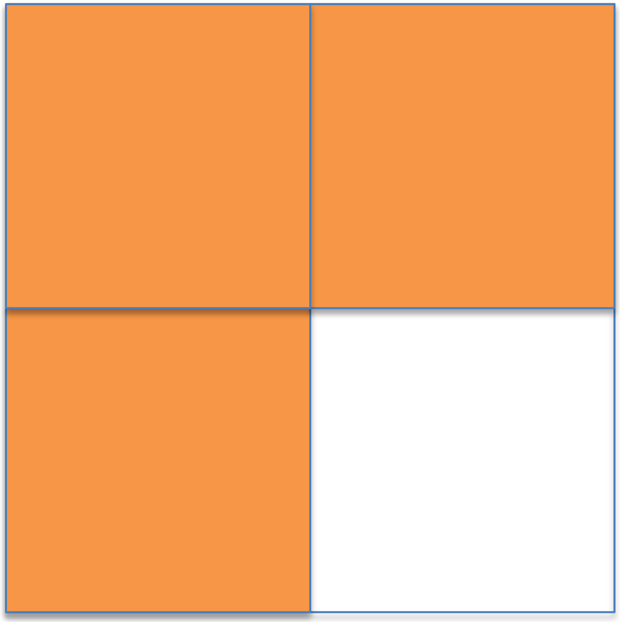
- "Community Structure" is only one type of *mesoscopic structure*.
- It's just by far the best-developed one.

Other "Block Models"

- E.g., bipartiteness



- E.g., core-periphery structure
  - Some ideas (especially in social science literature) but currently no method to find this systematically
  - M. P. Rombach, MAP, J. H. Fowler, & P. J. Mucha, in preparation



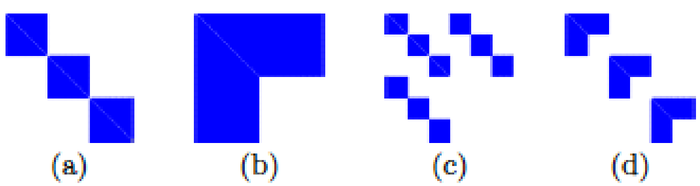
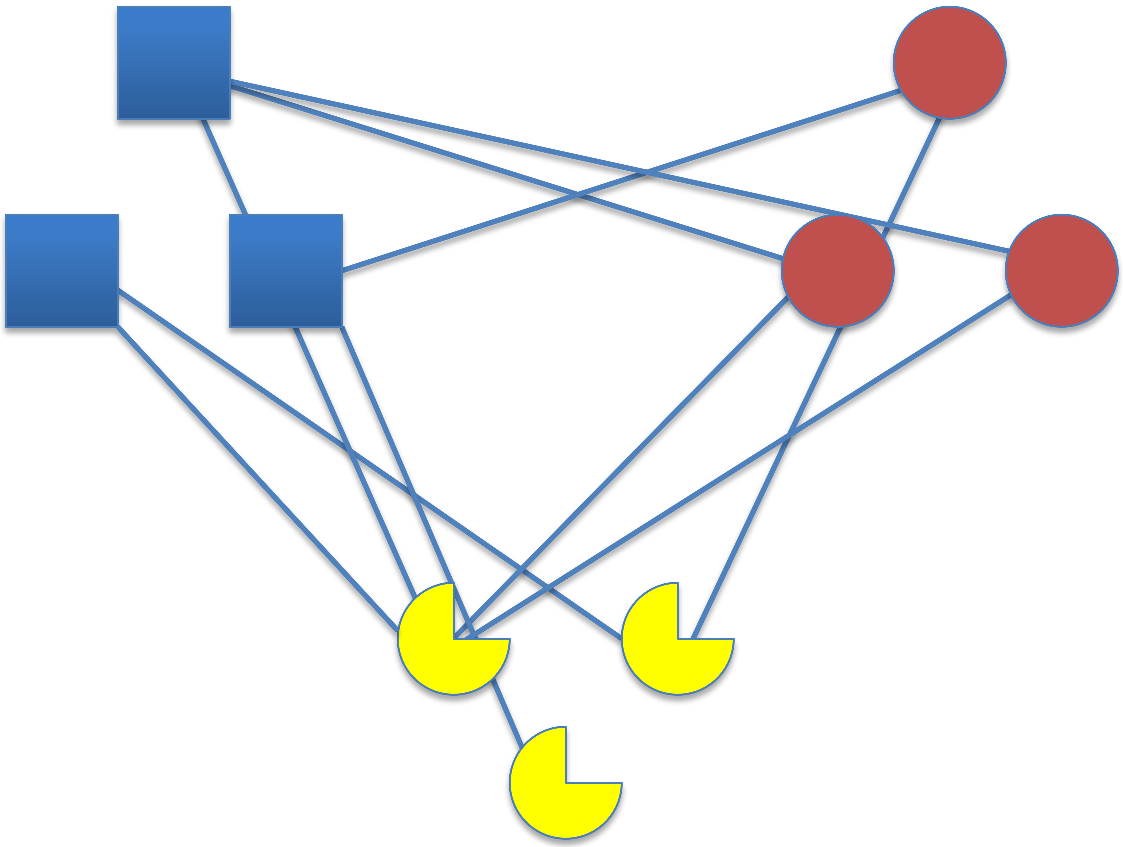


Fig. 2. Different network models. (a) community structure, (b) core-periphery structure, (c) global core-periphery / local community structure and (d) global community / local core-periphery structure.

Structural/Regular Equivalence



Overlapping Community Structure

- Might want cohesive groups to overlap ("soft partitioning")
  - There are some methods that allow this.
  - Additionally, some hard-partitioning methods have accompanying computations of strength of attachment to communities.

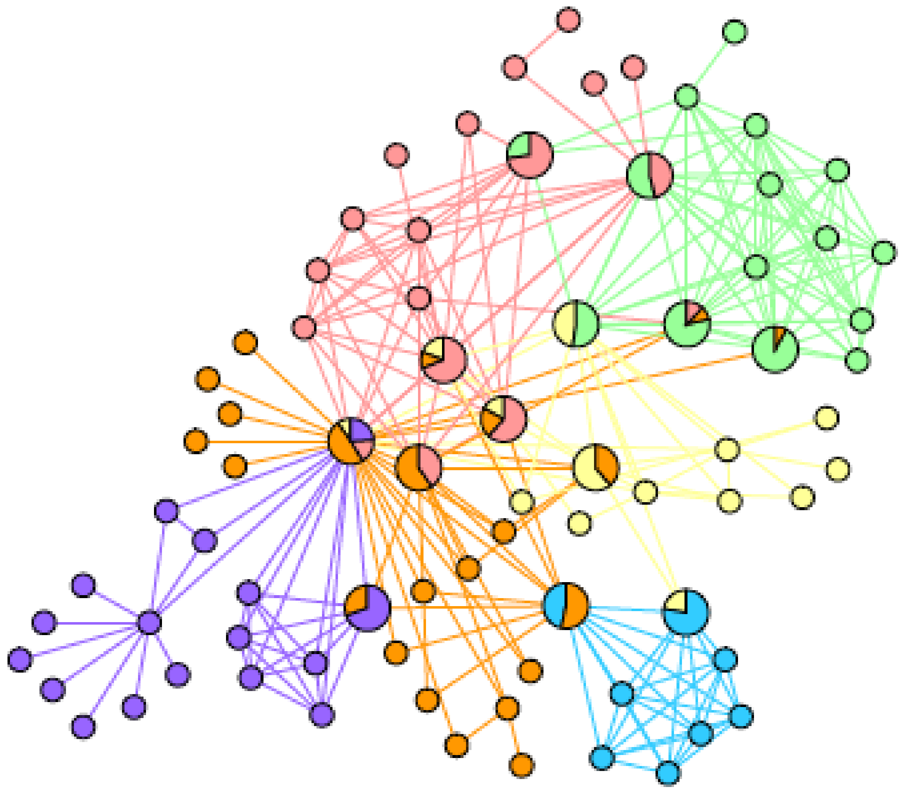


Image from Ball et al. (PRE, 2011)

More Big-Picture Ideas

- “Community detection” versus “community extraction”
  - Goal: Extract cohesive functional groups
- Defining cohesive groups via appropriate dynamics on networks
  - E.g., instead of using a random walk, use a different dynamical system (appropriate to the problem under study) and see if that can be mapped to a different quality function to optimize
- Spatial Networks: Networks are affected by the fact that they’re embedded in space
  - Review article: Barthelemy, 2011
  - E.g., two network diagnostics might not be strongly correlated in general, but they might be if one considers only networks embedded in $\mathbf{R}^3$

Going Beyond Networks

- Well, we should… But one needs to develop good concepts, develop good algorithms, etc.
- Hypergraphs
  - E.g. What in Hell is an "interaction" in protein-protein interaction networks?
- Tensors instead of Matrices
  - E.g., Multislice networks (Mucha et al, *Science*, 2010), SVDs on 'cubes' of data
- *More intricate structures entail more choices*
  - E.g. clustering coefficient already has many choices for weighted graphs, so there are multiple ways to generalize and the "right" ways should be domain-specific
  - E.g., parameter choices in multislice community detection
  - E.g., what properties of SVDs to preserve in higher-dim generalization?

Balanced Binary Trees