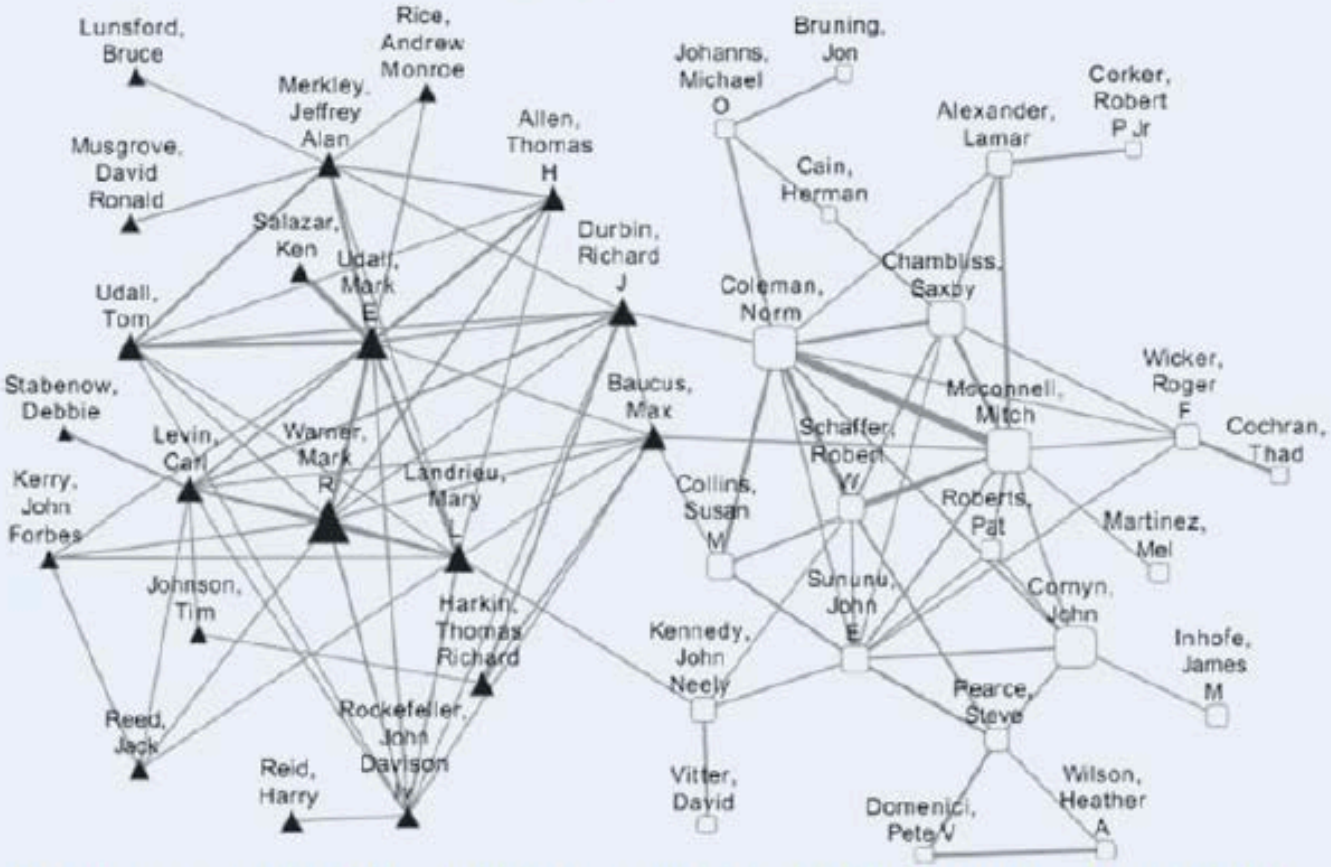
Detecting communities in time-evolving networks: Application to voting patterns in the US Senate

Based on
"Sequential detection of temporal communities
by estrangement confinement"
Kawadia, Sreenivasan, *Sci. Rep.* **1**:794 (2015)

Rensselaer

What are the networks that pervade the political arena?

Unipartite Projection of 2008 Co-Contribution Network of Senate Candidates' Campaign Committees



Note: Cut-off is at $50,000 for both nodes and links; thickness of edges and size of nodes corresponds to amount of co-contribution. Republicans are represented by squares and Democrats by triangles. Adapted from Onnela and Lazer (2009).

Onnela and Lazer (2009)

What are the networks that pervade the political arena?

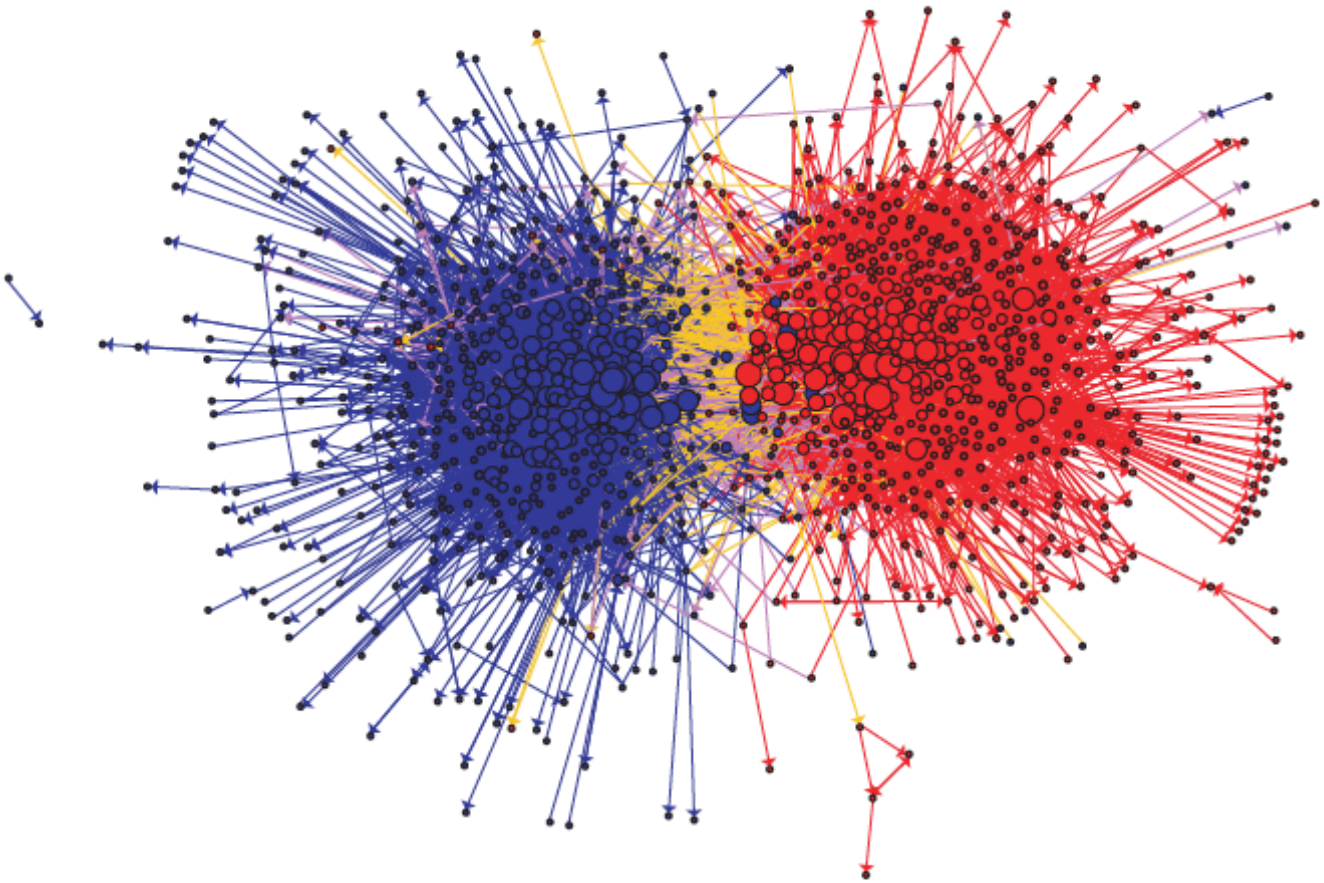


Figure 1: Community structure of political blogs (expanded set), shown using utilizing the GUESS visualization and analysis tool[2]. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

Adamic and Glance, Proceedings of the 3rd international workshop on Link discovery, 36 (2005)

What are the networks that pervade the political arena?

A nice review of prior research on networks in politics:

SYMPOSIUM

Networks in Political Science: Back to the Future

David Lazer, *Northeastern University and Harvard University*

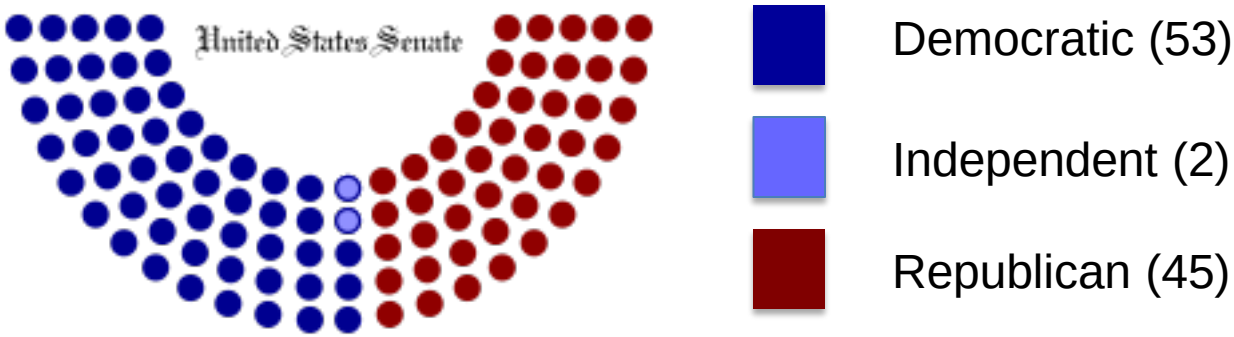
PS: Political Science & Politics, Volume 44, Issue 01, January 2011, pp 61-68

The United States Senate

- The United States Senate is a legislative chamber in the bicameral legislature of the United States of America, and together with the U.S. House of Representatives makes up the U.S. Congress.

- Each U.S. state is represented by two senators, regardless of population.

- Senators serve terms of six years each; the terms are staggered so that approximately one-third of the seats are up for election every two years.



Roll Call Voting Network

- Each senator's vote on a particular bill (*yea* or *nay*) is recorded. This kind of vote is called a roll call vote. Distinct from a voice vote where only total vote tallies are recorded.

- Roll call voting data for all senate sessions available at: http://voteview.com/DWNL.htm

- Using this data, we can construct for each senate session, a network among senators which indicates the similarity between their voting records for that session.

- Weight of an edge/link between two senators:

$$\frac{\text{Number of times they voted similarly}}{\text{Number of bills they both voted on}}$$

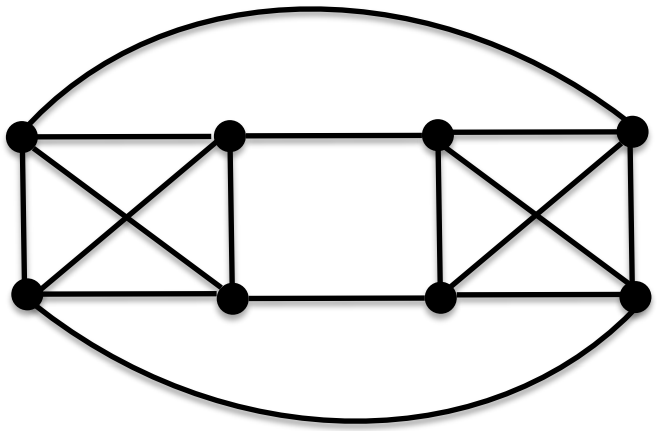
Q: Can we detect groups in the senate characterized by similar voting behavior?

Formal method to find *communities* in networks

What is a good way to define a community/group?

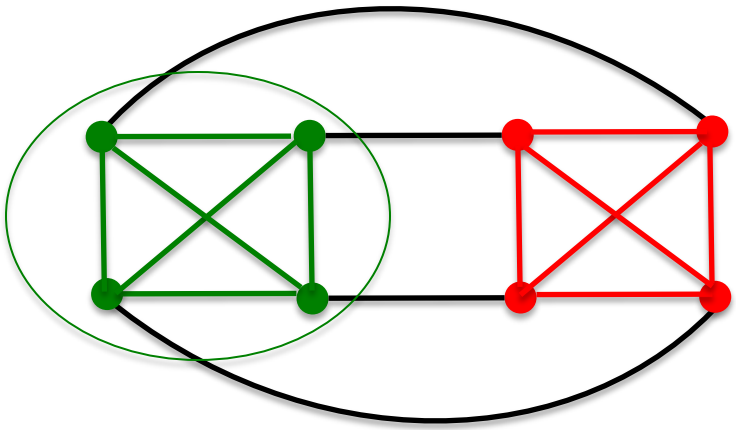
A subset of nodes whose induced subgraph has more links than expected by chance.

Example:



4-regular graph: every node has 4 edges

Let's try some candidate "partitions" of this graph into communities.



Total number of links: $\frac{1}{2}\sum_{uv} A_{uv}$ = 6

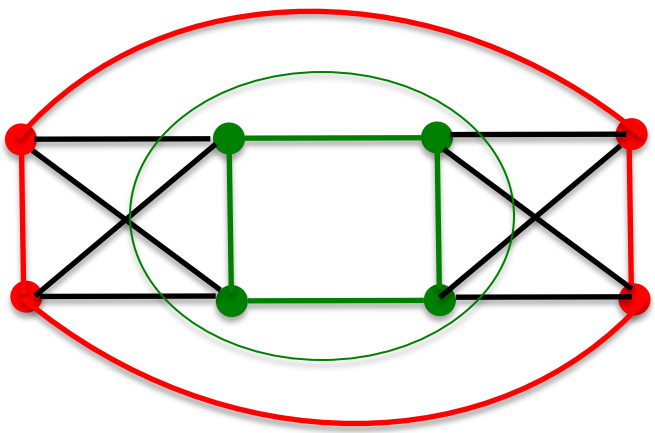
Number of links expected by chance:

$$\frac{1}{2}\sum_{uv}\frac{k_u k_v}{2M} = 3$$

Difference = 3

↑

More plausible community partition



Total number of links: $\frac{1}{2}\sum_{uv} A_{uv}$ = 4

Number of links expected by chance:

$$\frac{1}{2}\sum_{uv}\frac{k_u k_v}{2M} = 3$$

Difference = 1

Formal method to find *communities* in networks

More formally:

For a specific partition of nodes, sum over all communities that the partition creates:

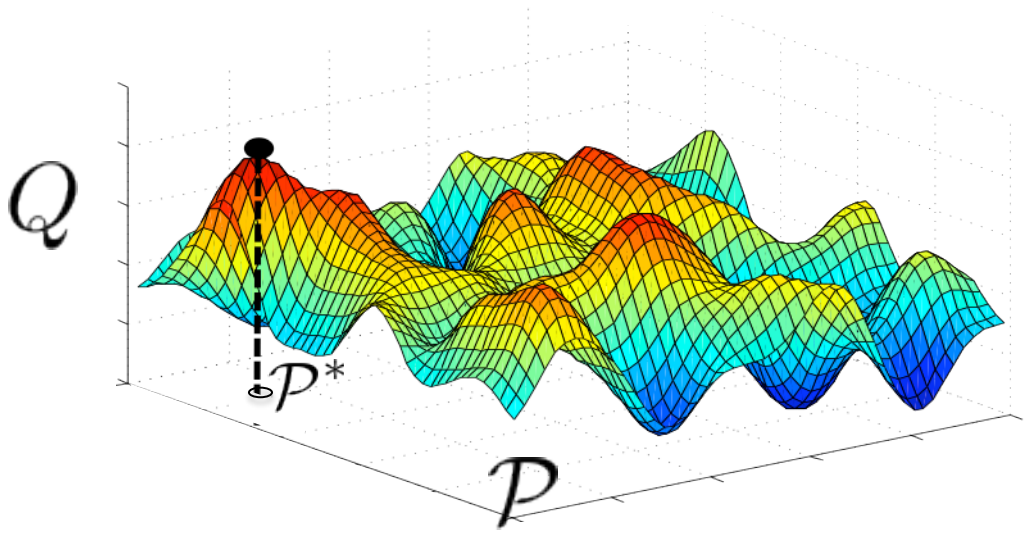
$$Q = \frac{1}{2M} \sum_{u,v} (A_{uv} - \frac{k_u k_v}{2M}) \delta(l_u, l_v)$$

Q : *Modularity* of the given partition.

One way to find communities: Find the partition that maximizes Q

Formal method to find *communities* in networks

One way to find communities: Find the partition that maximizes Q



Partition space

Formal method to find *communities* in networks

Modularity maximization using the Label Propagation Algorithm (LPAm)

M. J. Barber and J. W. Clark, *Phys. Rev. E* 80, 026129

Steps:

- *Initialize* each node's community label to the respective node's index.
- *do*:

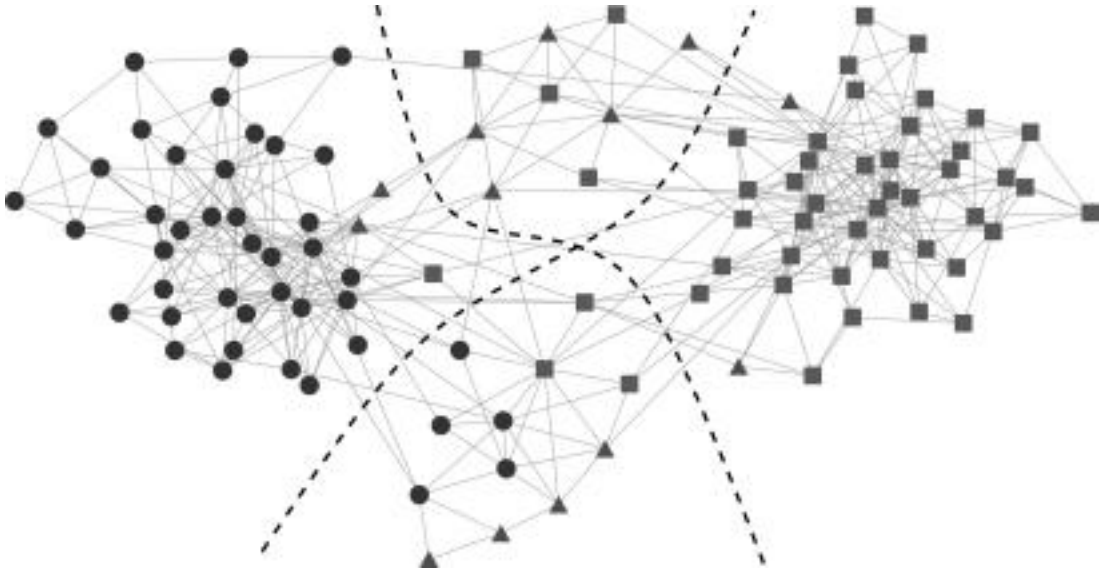
  for each node u update its label using the following rule:

$$l_u = \arg\max_l \left(N_{ul} - \frac{k_u K_l}{2M} + \frac{k_u^2}{2M} \delta(l_u, l) \right)$$

  while: at least one node label has changed.

Formal method to find *communities* in networks

Communities found by modularity maximization on the co-purchasing network of political books



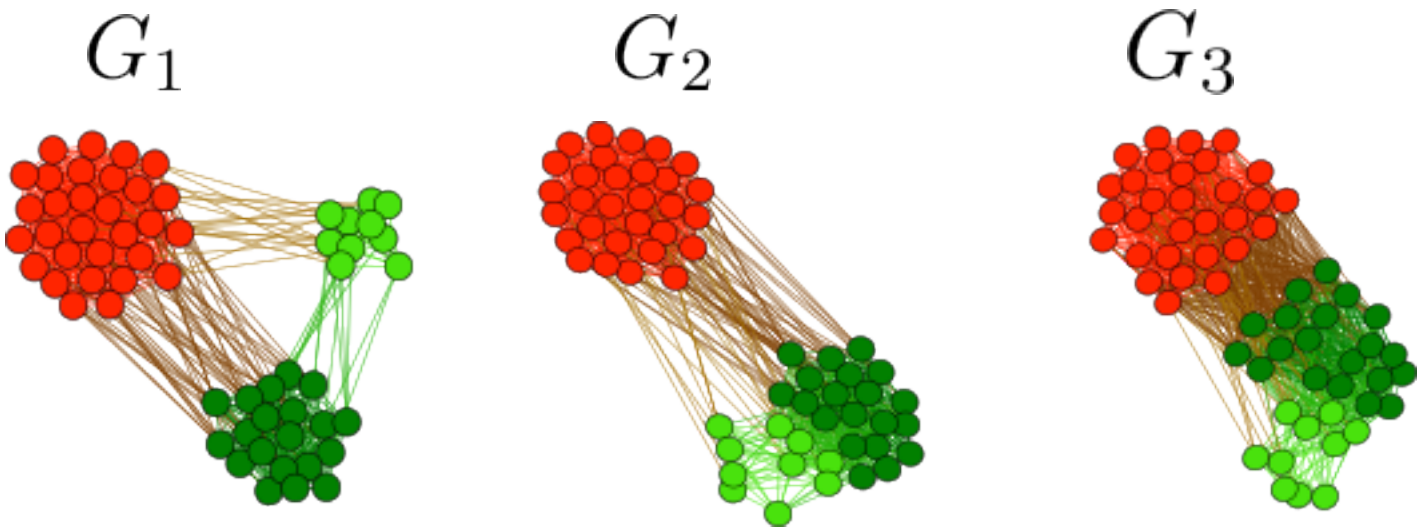
Liberal

Conservative

Centrist/Unaligned

Community detection in temporally evolving networks

Network structure changes in time => community structure could also be evolving



Simplest method:

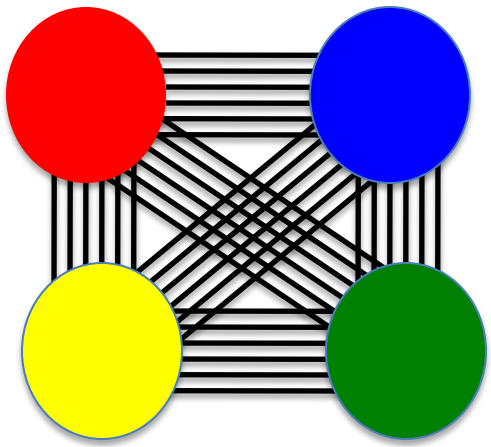
Apply the *modularity* hammer on each snapshot independently.

…there is a problem, though.

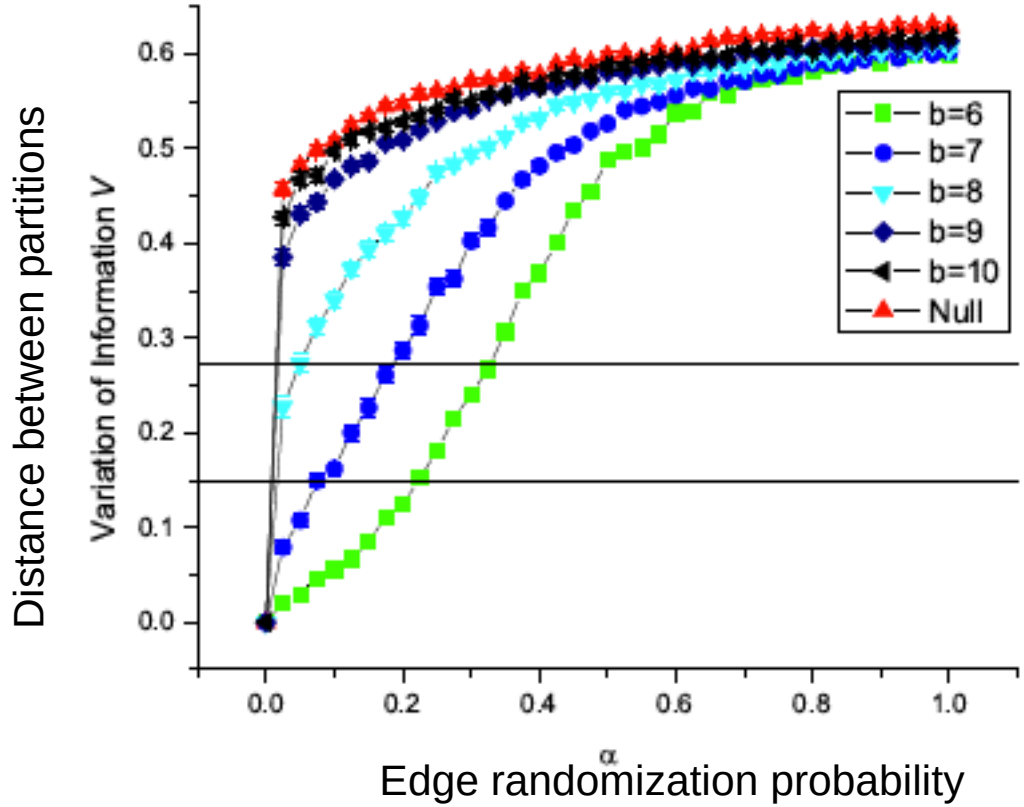
Community detection in temporally evolving networks

Small change in network structure can have a large effect on Q landscape

4 communities,
b edges btw. communities



Karrer, Levina and Newman,
Phys. Rev E 77 046119 (2008)

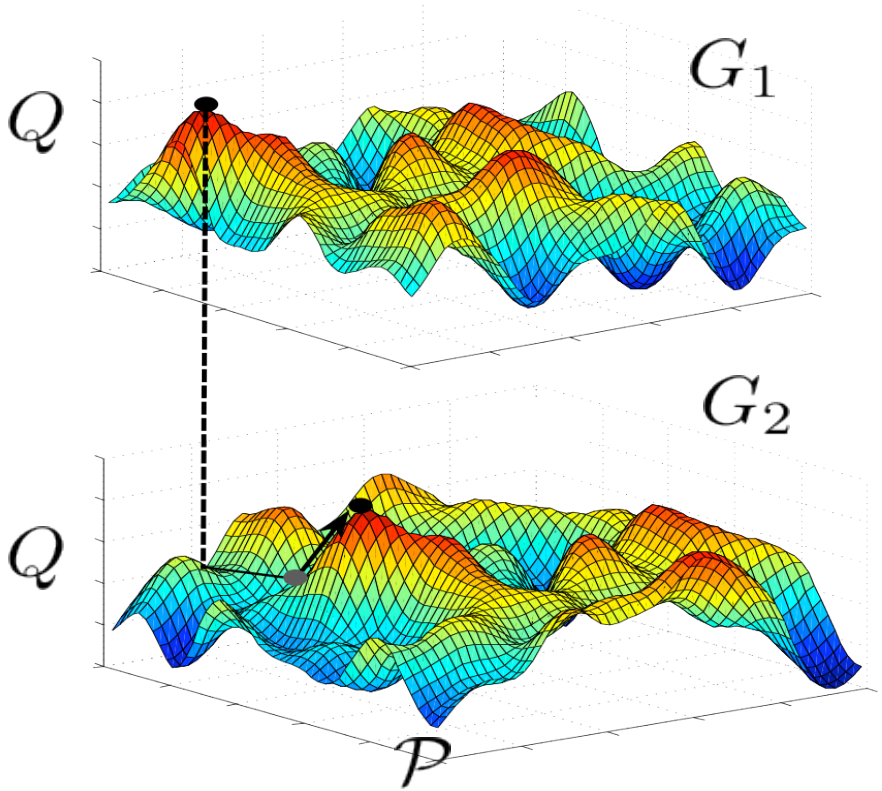


Undesirable consequence: there can be a big “jump” in community structure, even when the network structure change is small.

Community detection in temporally evolving networks

If network change is small we want to maximize Q but with the constraint that the new partition is *close* to the old one => smooth community evolution.

(If network change is large, allow partition to change more)



Partition space

Community detection in temporally evolving networks

We need a penalty function that:

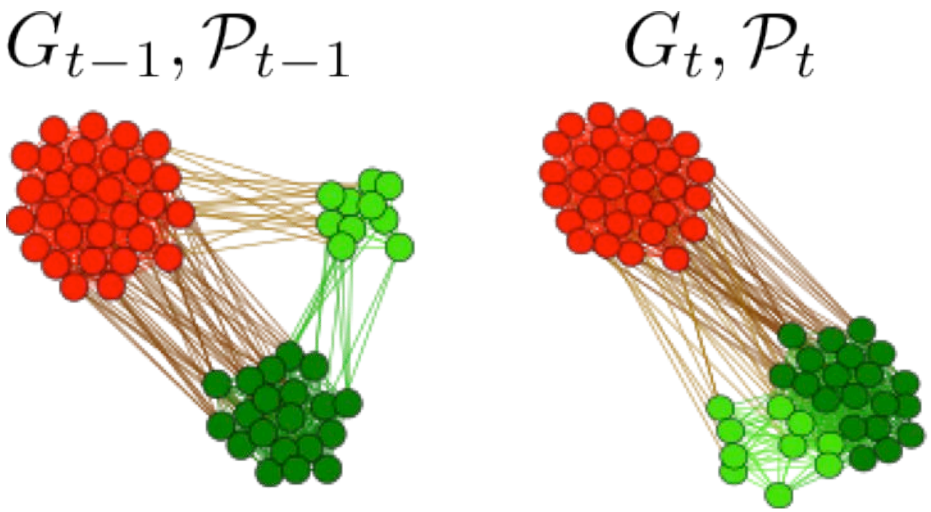
- Penalizes changes in partition structure with consideration of how much the network has changed.

- Penalizes changes in partition structure locally – different parts of the network can have different extents of partition change, depending on their respective link churns.

- Can be broken down into local contributions (just like modularity), so that it can be incorporated into the Label Propagation rule.

Community detection in temporally evolving networks

A candidate penalty function: **Estrangement**

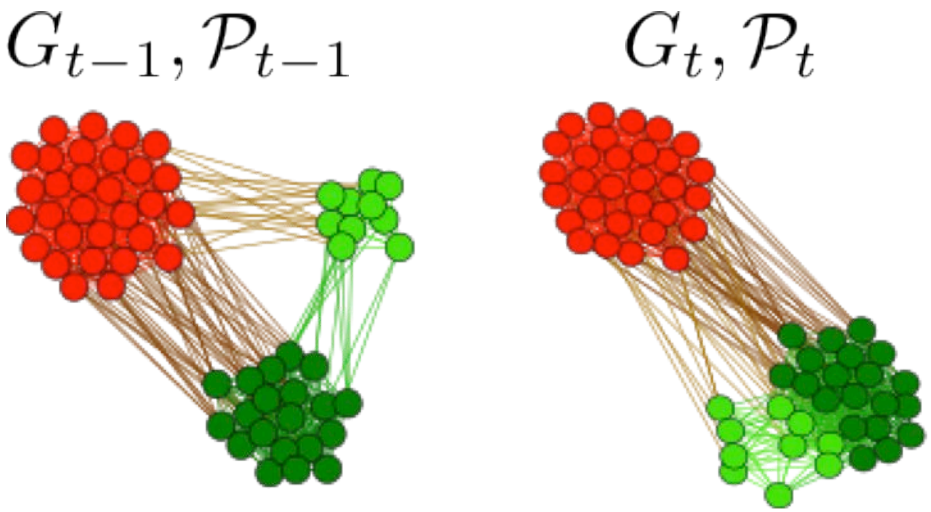


Definition 1:

Estranged link/edge: A link/edge present in both G_{t-1} and G_t , whose end points belonged to the same community in P_{t-1} but belong to different communities in P_t .

Community detection in temporally evolving networks

A candidate penalty function: **Estrangement**

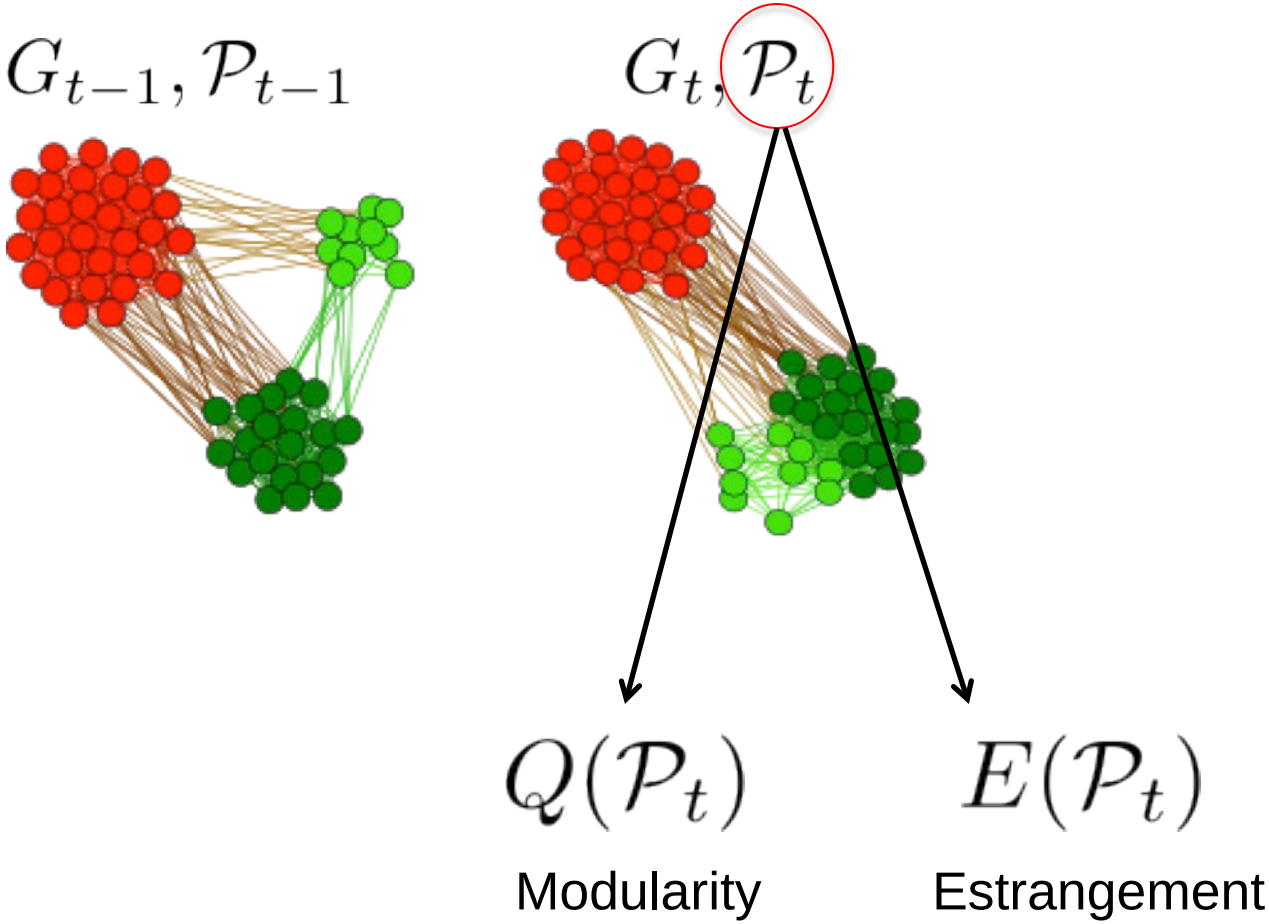


Definition 2:

Estrangement: $$\frac{\text{Total number of estranged edges between } G_{t-1} \text{ and } G_t}{\text{Total number of edges in } G_t}$$

Community detection in temporally evolving networks

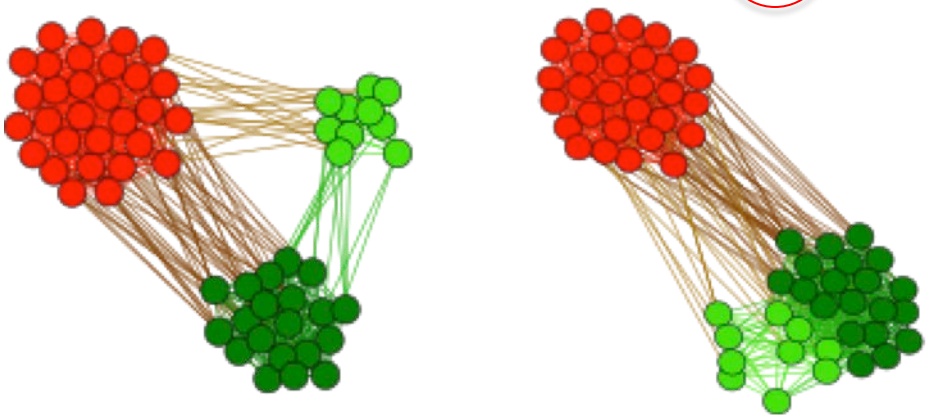
Two quantities associated with any given partition of graph at time t:



Community detection in temporally evolving networks

Choice of partition at time t becomes a constrained optimization problem:

$G_{t-1}, \mathcal{P}_{t-1}$ $\qquad$ $G_t, \mathcal{P}_t$



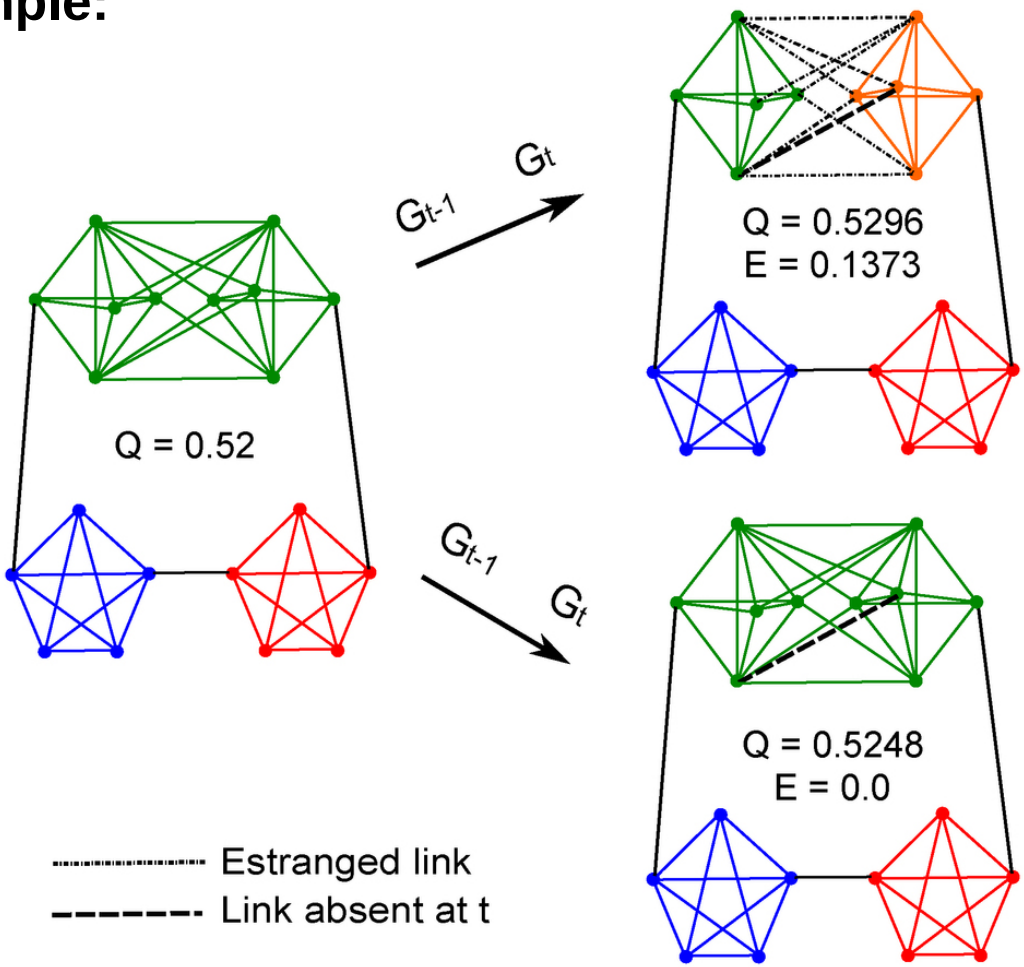
$$\underset{\mathcal{P}_t}{\text{maximize}}\ Q(\mathcal{P}_t)$$

$$\text{subject to } E(\mathcal{P}_t) \leq \delta$$

δ : Quantifies how much “jumpiness” is tolerated in community evolution

Community detection in temporally evolving networks

Example:



Higher modularity

Higher estrangement

Lower modularity

Zero estrangement

Community detection in temporally evolving networks

Constrained Modularity maximization using the Label Propagation Algorithm

Steps:

- *Initialize* each node's community label to the respective node's index.
- *do*:

  for each node u update its label using the following rule:

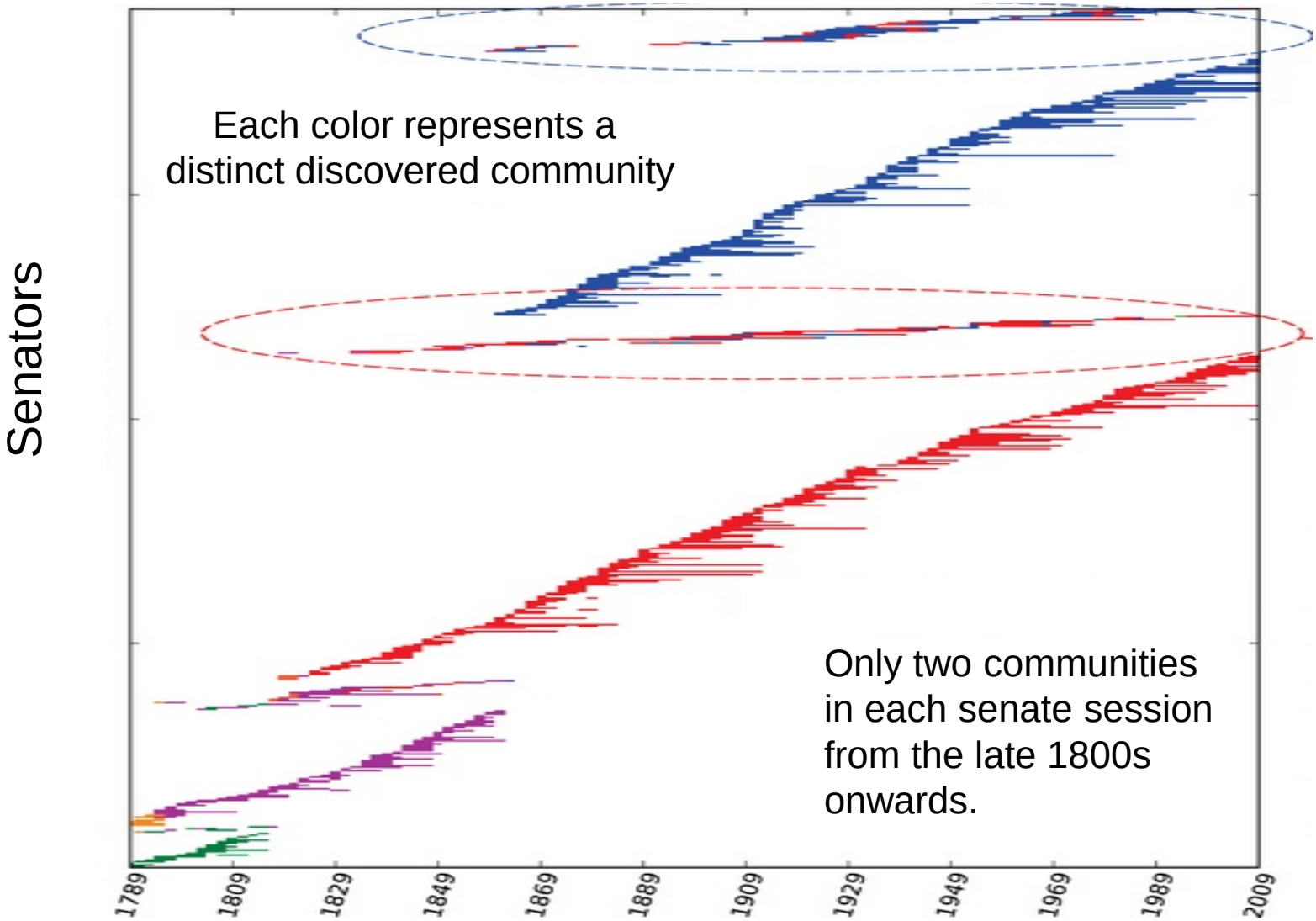
$$l_u = \arg\max_l \left(N_{ul} - \frac{k_u K_l}{2M} + \frac{k_u^2}{2M}\delta(l_u, l) + \lambda\, O_{ul} \right)$$

  while: at least one node label has changed.

Lagrange multiplier

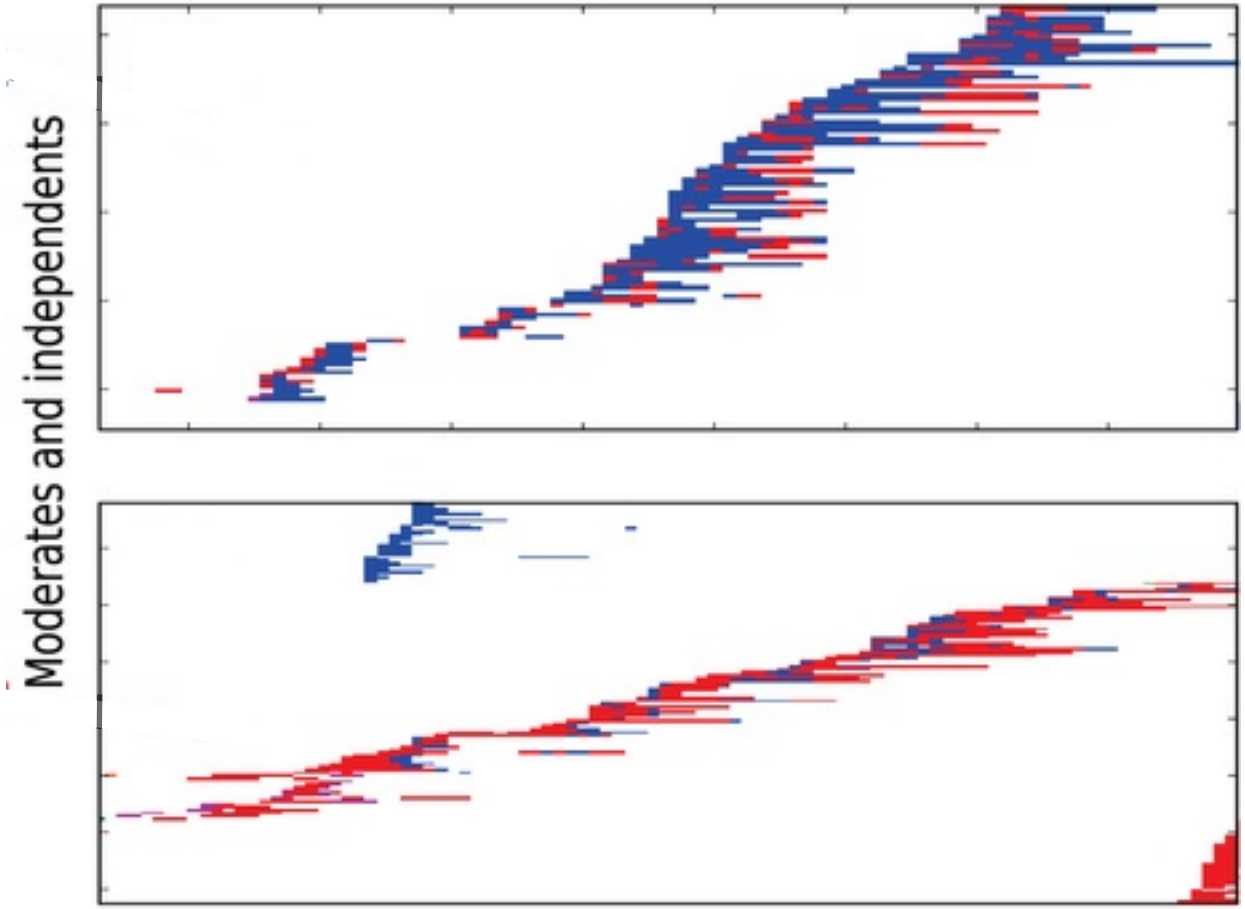
Contribution of node u to E if it accepted l as its community label

Estrangement constrained community detection on the US Senate co-voting network



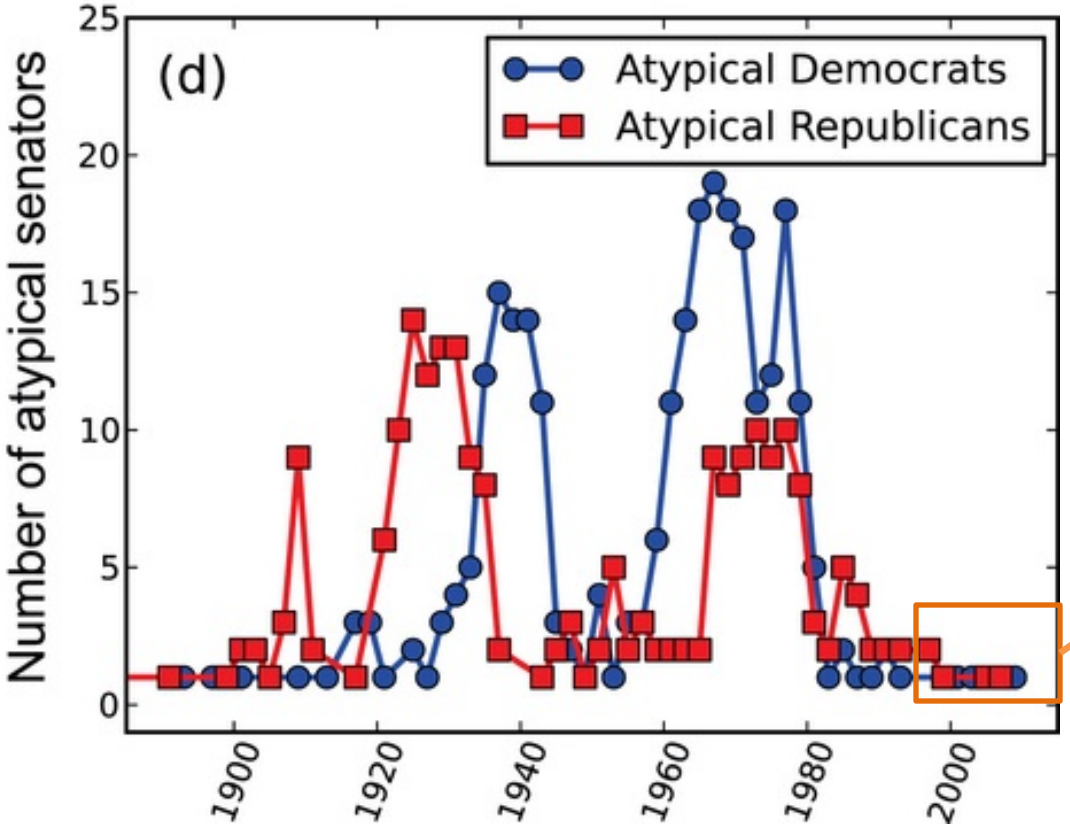
Estrangement constrained community detection on the US Senate co-voting network

Senators whose community membership switches often



Estrangement constrained community detection on the US Senate co-voting network

"Atypical" senators: A senator whose community label is different from the majority of his party colleagues.



Larger partisanship

Summary

- Community detection on an implicit or explicit network is a useful approach to segmenting populations based on behavior.

- Can be applied to other networks, for example:

  Politics: networks based on co-sponsorship of bills by senators, campaign co-contribution network for senate candidates etc.

  Recommendation systems: similarity networks among users based on products they buy, movies they watch etc.

- Methods and results presented here were published in:

  V. Kawadia and S. Sreenivasan, Scientific Reports 2, 794 (2012)
  https://www.nature.com/articles/srep00794