Network Analysis:

The Hidden Structures behind the Webs We Weave

17-213 / 17-668

Affiliations and Overlapping Subgroups

Thursday, October 5, 2023

Patrick Park & Bogdan Vasilescu

Carnegie Mellon University
School of Computer Science

S3D
Software and Societal Systems Department

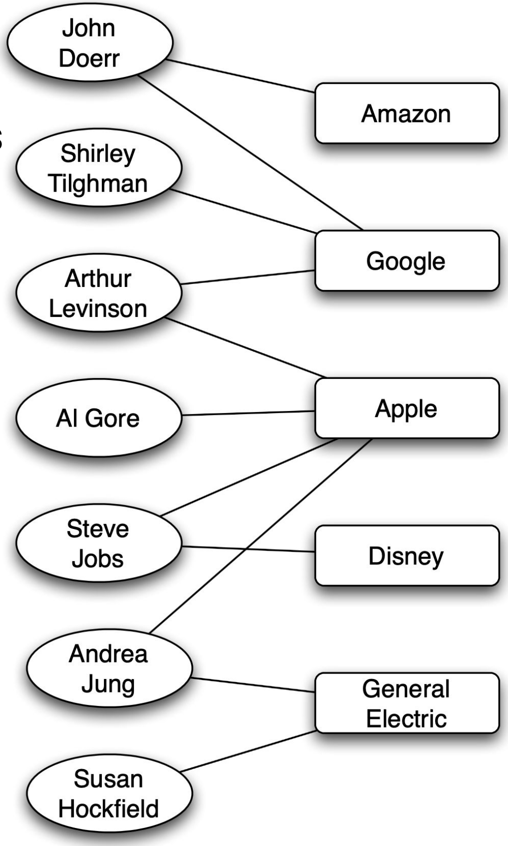
2-min Quiz, on Canvas

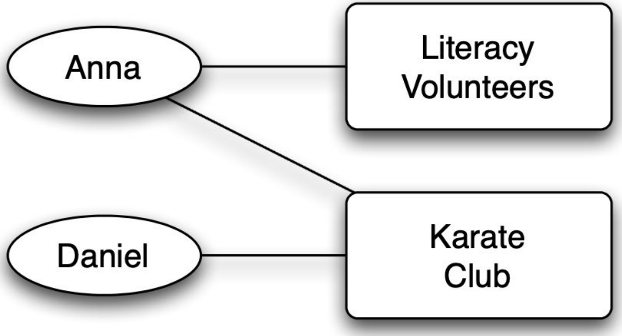
Quick Recap – Last Tuesday's Lecture

Affiliation networks are two-mode networks

We can represent the participation of a set of “actors” (people) in a set of “events” (activities) using a graph

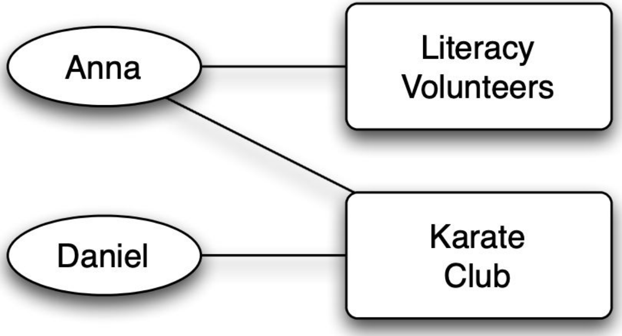
People on corporate boards of directors





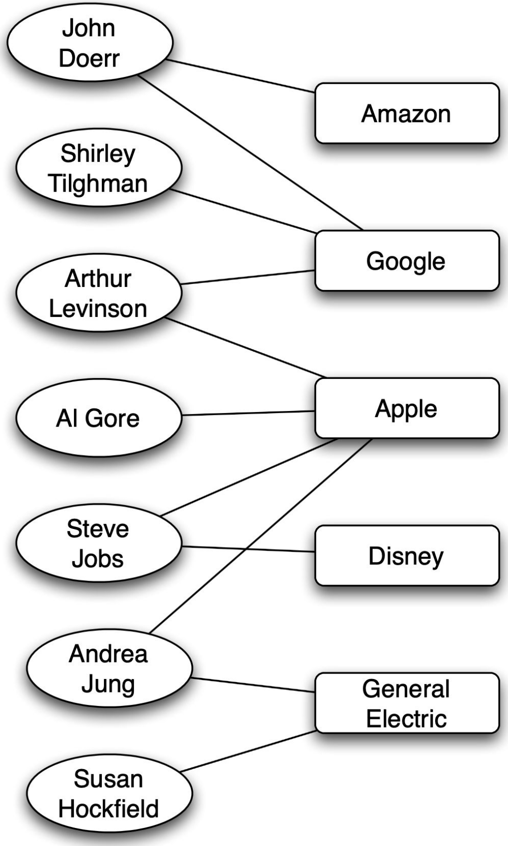
Individuals affiliated with groups or activities

Affiliation networks can reveal interesting relationships on both sides of the graph.



Actors linked by events

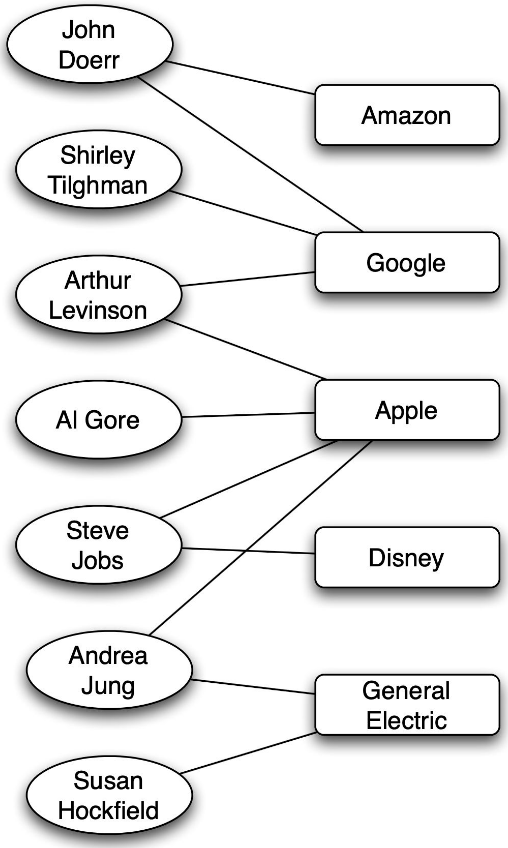
Events linked by authors



Affiliation networks can reveal interesting relationships on both sides of the graph.

Two <u>companies</u> are implicitly linked by having the same person sit on both their boards
→ possible conduits for information and influence to flow between different companies.

Two <u>people</u> are implicitly linked by serving together on a board
→ patterns of social interaction among some of the most powerful members of society.



Three rationales for studying affiliation networks

Individuals' affiliations with events provides direct linkages between the actors and/or between the events.

Contact among individuals who participate in the same social events increases the likelihood of tie formation.

The interaction between actors and events forms a social system that should be studied as a whole.

Affiliation networks can be represented as a matrix

The affiliation matrix A = {a_{ij}} has rows as authors and columns as events.

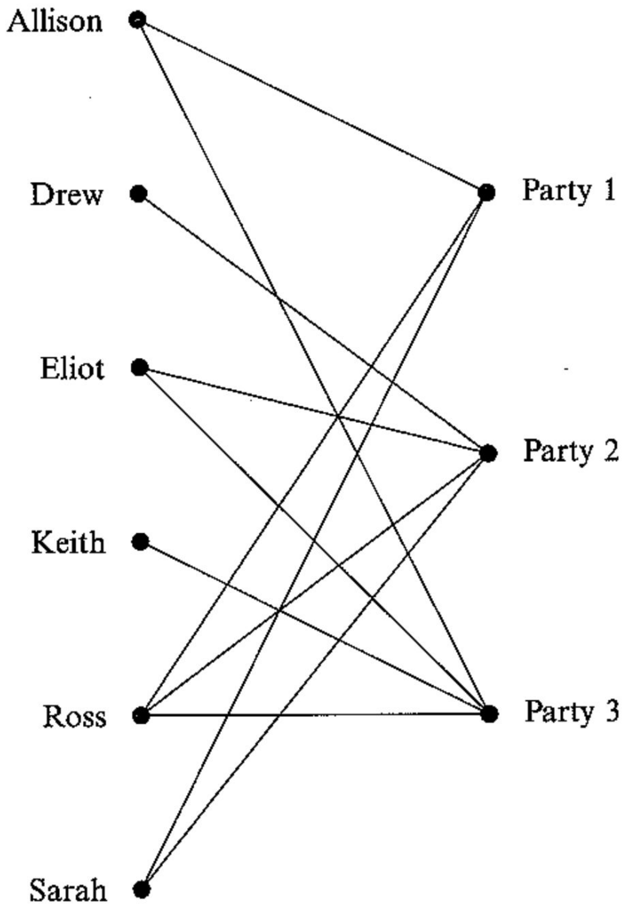
a_{ij} = 1 if actor i is affiliated with event j

Actor	Event Party 1	 Party 2	 Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0

(Wasserman & Faust)

… or as a bipartite graph

Note the context-dependent meaning of "degree."



(Wasserman & Faust)

The bipartite graph can also be represented as a sociomatrix

	Allison	Drew	Eliot	Keith	Ross	Sarah	Party 1	Party 2	Party 3
Allison	-	0	0	0	0	0	1	0	1
Drew	0	-	0	0	0	0	0	1	0
Eliot	0	0	-	0	0	0	0	1	1
Keith	0	0	0	-	0	0	0	0	1
Ross	0	0	0	0	-	0	1	1	1
Sarah	0	0	0	0	0	-	1	1	0
Party 1	1	0	0	0	1	1	-	0	0
Party 2	0	1	1	0	1	1	0	-	0
Party 3	1	0	1	1	1	0	0	0	-

(Wasserman & Faust)

The bipartite graph can also be represented as a sociomatrix

A:

	Allison	Drew	Eliot	Keith	Ross	Sarah	Party 1	Party 2	Party 3
Allison	-	0	0	0	0	0	1	0	1
Drew	0	-	0	0	0	0	0	1	0
Eliot	0	0	-	0	0	0	0	1	1
Keith	0	0	0	-	0	0	0	0	1
Ross	0	0	0	0	-	0	1	1	1
Sarah	0	0	0	0	0	-	1	1	0
Party 1	1	0	0	0	1	1	-	0	0
Party 2	0	1	1	0	1	1	0	-	0
Party 3	1	0	1	1	1	0	0	0	-

(Wasserman & Faust)

We can summarize the co-membership frequencies

The product of A and A' (transpose)

$$\mathbf{X}^{\mathcal{N}} = \mathbf{AA}'.$$

	n_1	n_2	n_3	n_4	n_5	n_6
n_1	2	0	1	1	2	1
n_2	0	1	1	0	1	1
n_3	1	1	2	1	2	1
n_4	1	0	1	1	1	0
n_5	2	1	2	1	3	2
n_6	1	1	1	0	2	2

Recall A:

		Event	
Actor	Party 1	Party 2	Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0

Similarly, we can summarize event overlap frequencies

The product of A’ and A

$$\mathbf{X}^{\mathscr{M}} = \mathbf{A}'\mathbf{A}$$

	m_1	m_2	m_3
m_1	3	2	2
m_2	2	4	2
m_3	2	2	4

Recall A:

	Event		
Actor	Party 1	Party 2	Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0

A take on “community” in this context

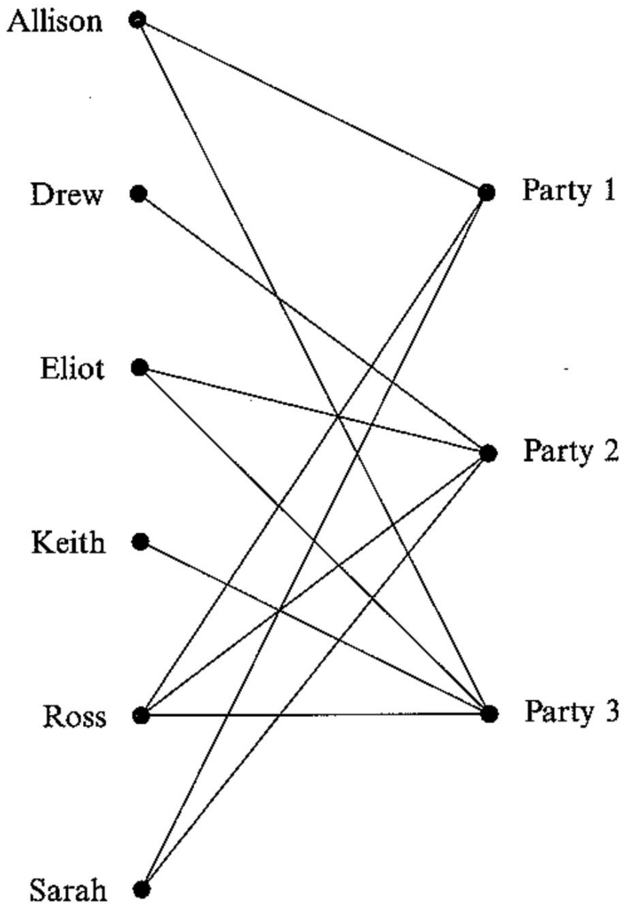
Community as clique at level c

For the co-membership relation for actors:

- Subgraph in which all *pairs* of actors share memberships in at least c events

For the overlap relation for events:

- Subgraph in which all *pairs* of events share at least c members



(Wasserman & Faust)

Back to triadic closure

The projection of two-mode networks creates a number of issues

1. Tie formation

Each tie in a prototypical one-mode network is assumed to be created separately, e.g., a standard phone call creates a communication tie from one person to another.

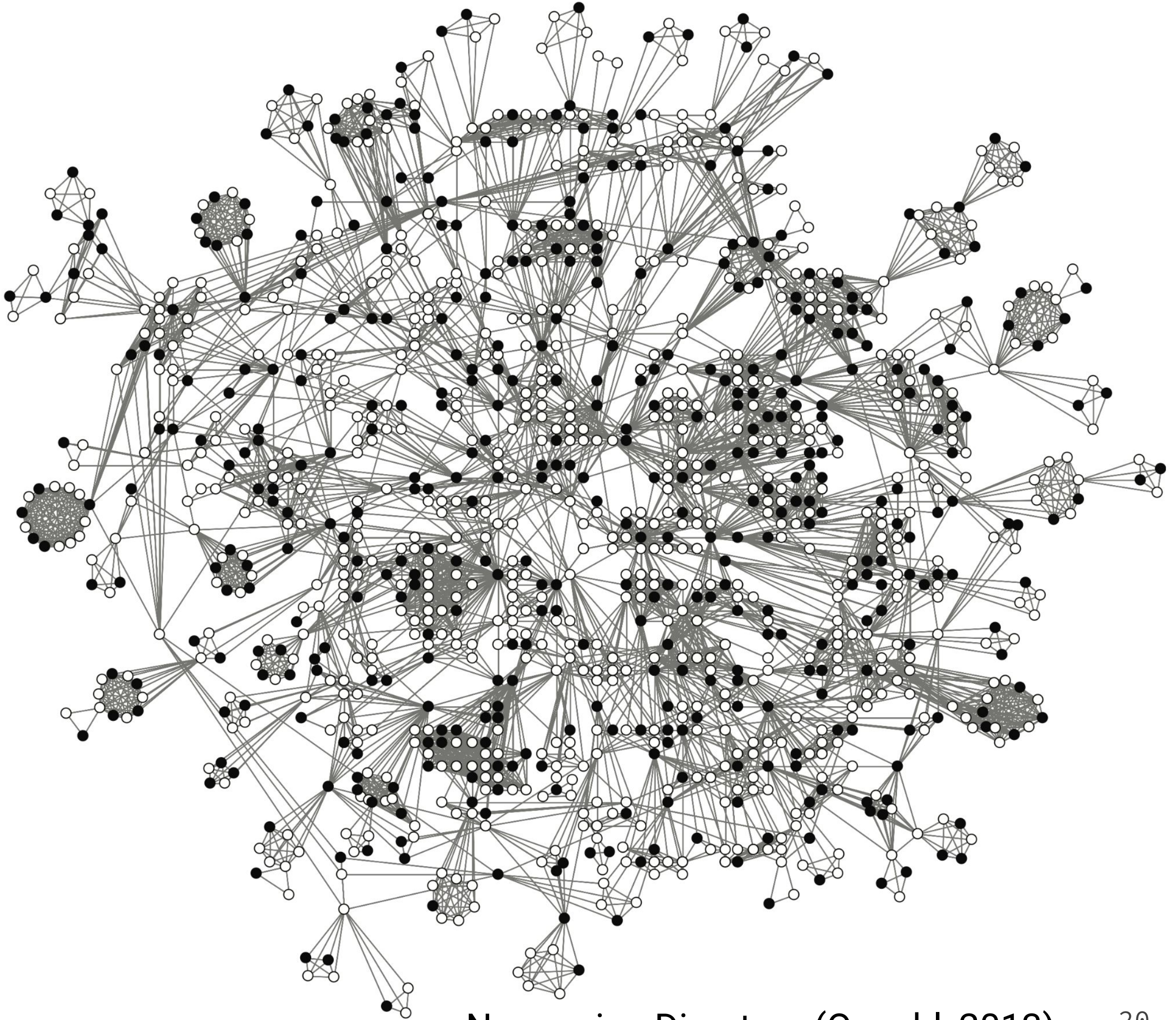
This is not the case in projected two-mode networks, e.g., a director forms ties with all the other Directors on a board when they joins that board.

→ How to use random networks to detect a baseline level? How to compare observed measures with those found in corresponding random networks?

2. Connectedness

A projected two-mode network tends to have more and larger fully-connected cliques than prototypical one-mode networks.

→ Impacts network measures based on triangles, e.g., clustering coefficients.



Norwegian Directors (Opsahl, 2013)

Clustering coefficients for one-mode networks

Global clustering coefficient: fraction of triplets or 2-paths (i.e., three nodes connected by two ties) that are closed by the presence of a tie between the first and the third node.

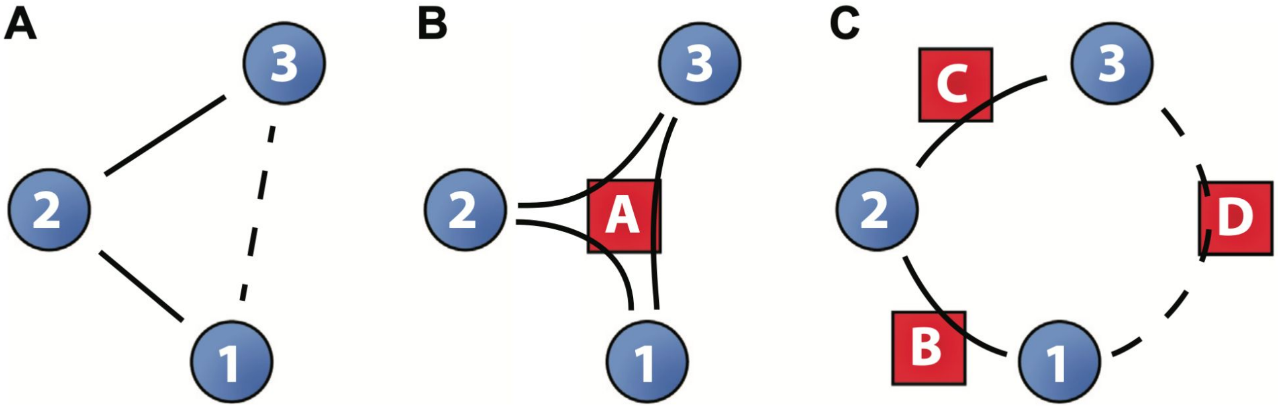
$$C = \frac{3 \times triangles}{triplets} = \frac{closed\ triplets}{triplets} = \frac{\tau_\Delta}{\tau}$$

Local clustering coefficient: the fraction of ties among a node's contacts over the possible number of ties between them.

$$C(i) = \frac{number\ of\ actual\ ties\ among\ node\ i's\ contacts}{number\ of\ possible\ ties\ among\ node\ i's\ contacts} = \frac{\tau_{i,\Delta}}{\tau}$$

A triangle in a projected two-mode network can be formed by two possible configurations

(Opsahl, 2013)



B: three primary nodes are connected to a common node, node A. Since this node creates the 2-path and closes it as well, all these 2-paths are closed by definition.

C: the three primary nodes become part of a 2-path when projected, but this 2-path is not closed by definition.

The clustering coefficient in one-mode classical random networks greatly underestimates the baseline-level of clustering in projected two-mode networks

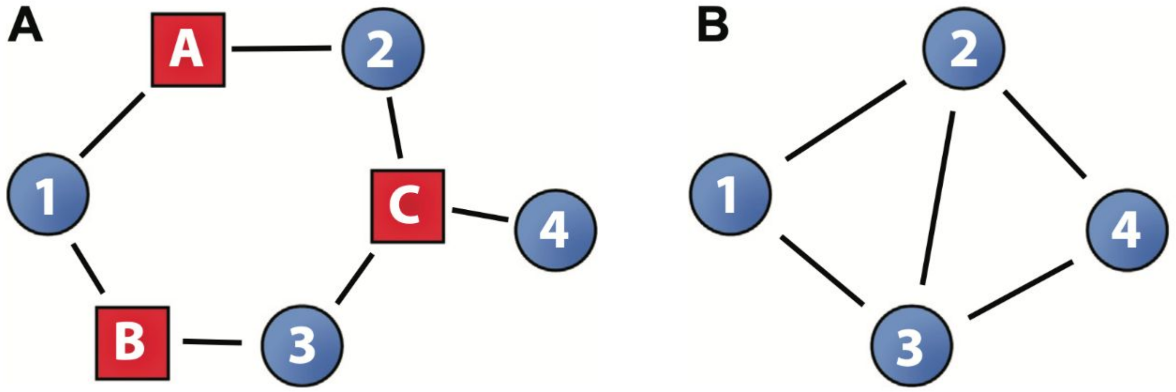
Opsahl randomized the two-mode structure of a scientific collaboration network while maintaining the degree distributions (i.e., randomly assigning the ties in the two-mode network while keeping each author's number of co-authored papers, and each paper's number of authors) before projecting it onto a one-mode network and calculating the global clustering coefficient.

Across 1000 projected random two-mode networks, the average global clustering coefficient was 0.1236, which is over 350 times larger than the coefficient in corresponding one-mode classical random networks.

Key idea: measure closure among three nodes from the primary node set instead of only two primary nodes

Example

A ⟶ B: one-mode projection of round blue nodes

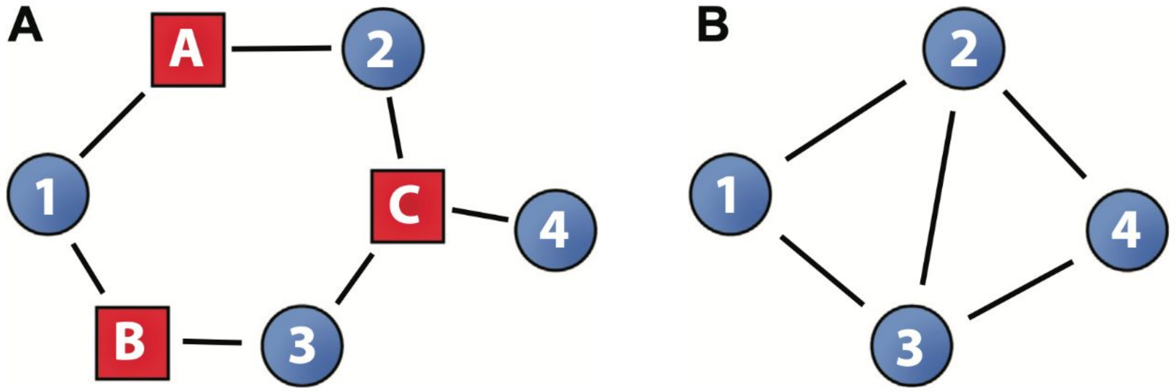


A: there are five 4-paths, three of which are closed:

- 1-A-2-C-3 (closed by node B)
- 1-A-2-C-4
- 1-B-3-C-2 (closed by node A)
- 1-B-3-C-4
- 2-A-1-B-3 (closed by node C)

Example

A → B: one-mode projection of round blue nodes



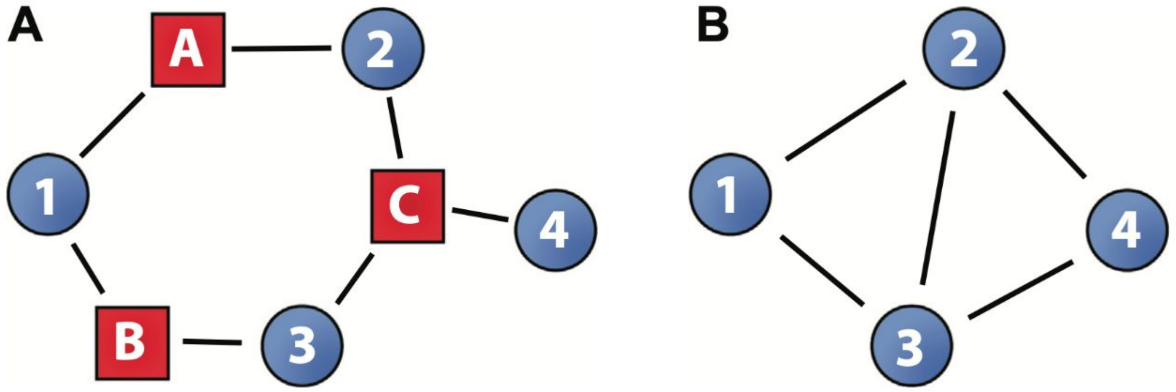
These 4-paths represent five 2-paths in the one-mode projection (panel B):
1-2-3 (closed); 1-2-4; 1-3-2 (closed); 1-3-4; 2-1-3 (closed)

However, in the one-mode projection, there are an additional three 2-paths:
2-3-4 (closed); 2-4-3 (closed); 3-2-4 (closed)

These three are created among node 2, node 3, and node 4 as these nodes are all connected to node C in the two-mode network.

Example

A → B: one-mode projection of round blue nodes



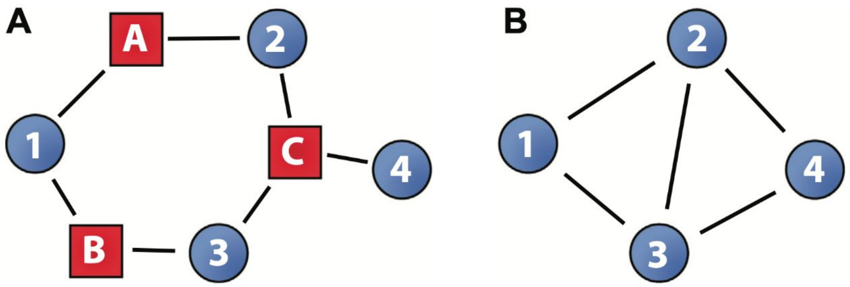
The clustering coefficient of the two-mode network (panel A) is 0.6, while the clustering coefficient of the one-mode projection (panel B) is 0.75.

Improved clustering coefficient definition

$$C^* = \frac{closed\,4paths}{4paths} = \frac{\tau^*_\Delta}{\tau^*}$$

where τ^* is the number of 4-paths, and τ^*_Δ is the number of these 4-paths that are closed by being part of at least one 6-cycle (i.e., a loop of six ties with five nodes).

C* = ⅗ = 0.6 in our example



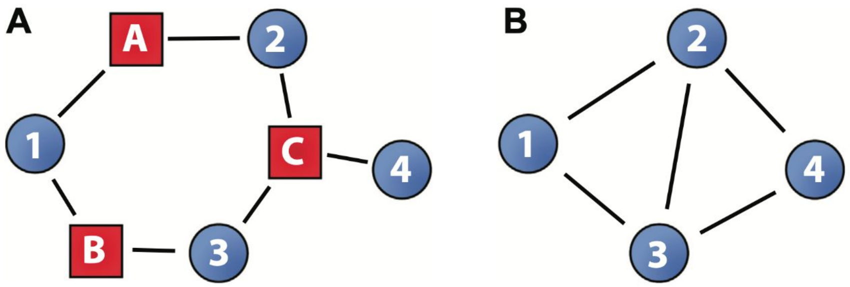
The local clustering coefficient for two-mode networks can be defined similarly

$$C^*(i) = \frac{closed\ 4paths\ centered\ on\ node\ i}{4paths\ centered\ on\ node\ i} = \frac{\tau^*_{i,\Delta}}{\tau^*_i}$$

where τ^*_i is the number of 4-paths centered on the focal node i, and $\tau^*_{i,\Delta}$ is the subset of these in which the first and the last nodes of the path share a common node that is not part of the 4-path (i.e., part of at least one 6-cycle).

Node 3 is in the center of two 4-paths, where node 1 can be seen as the first node, and nodes 2 and 4 as the last ones.

→ C*(3) = ½



C(3) = ⅔

Closure more broadly

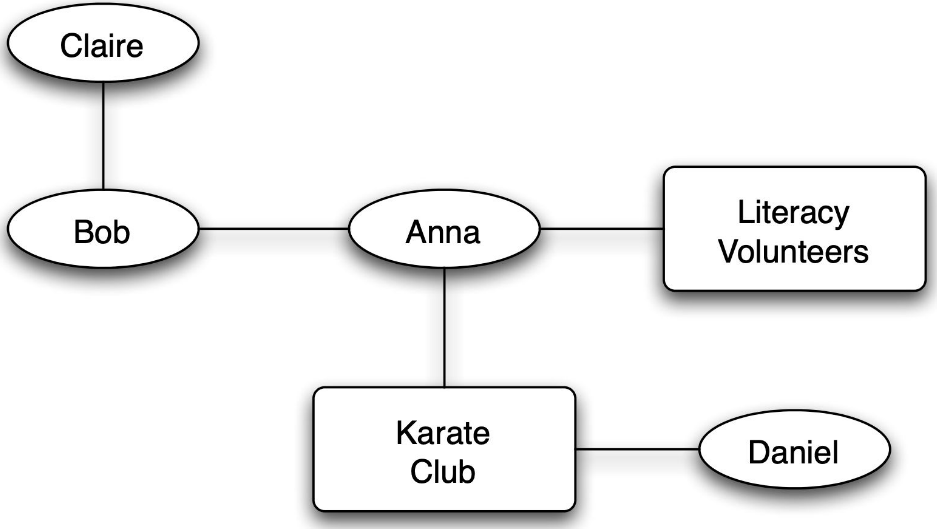
Let's slightly extend the notion of an affiliation network

As before, nodes for people and foci.

But, two kinds of edges:

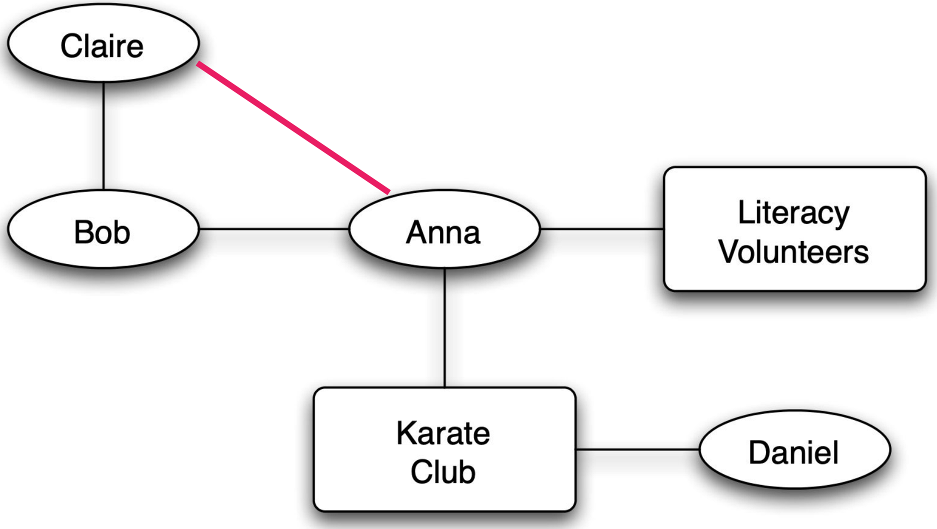
- Social (e.g., friendship)
- Affiliation (e.g., participation in activity)

Result: a "social-affiliation" network



Different mechanisms for link formation can now all be viewed as types of closure processes.

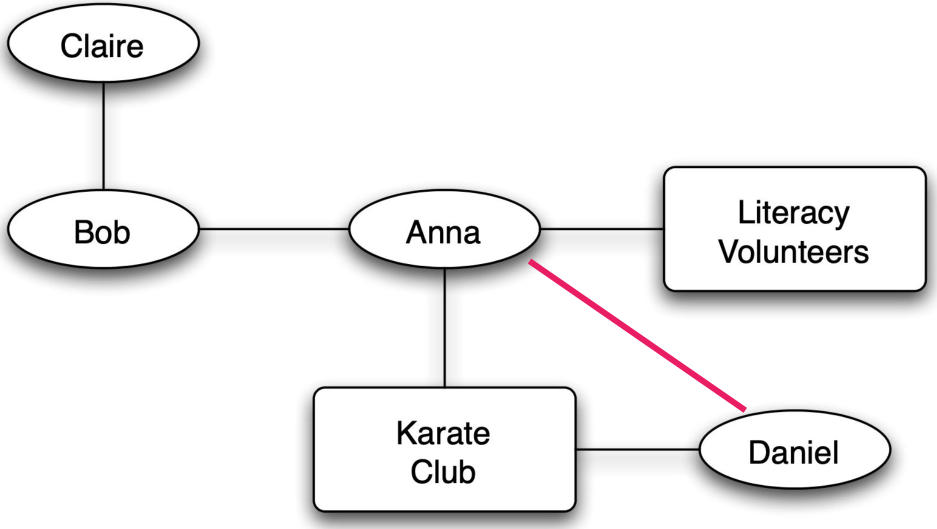
Triadic closure



Different mechanisms for link formation can now all be viewed as types of closure processes.

"Focal closure"

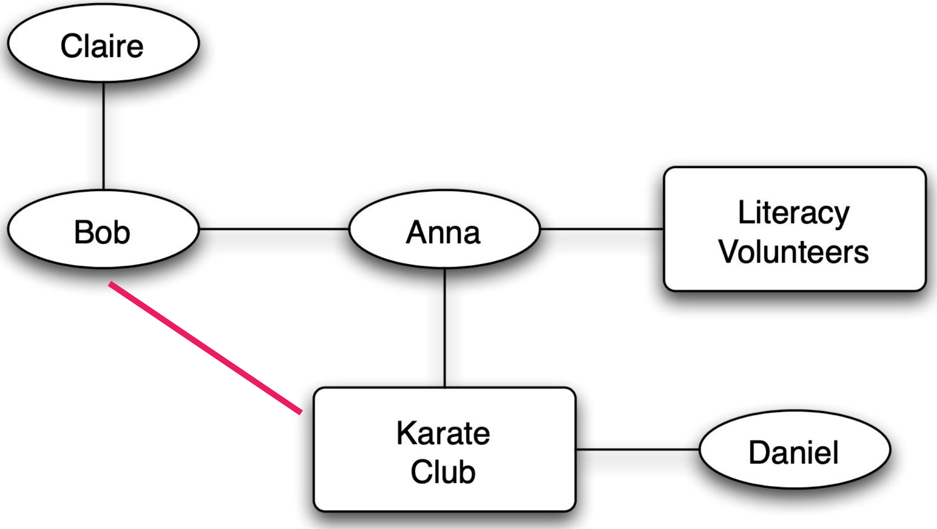
Selection mechanism: two people form a link when they have a focus in common.



Different mechanisms for link formation can now all be viewed as types of closure processes.

"Membership closure"

Social influence: Bob takes part in a focus that his friend Anna is already involved in.

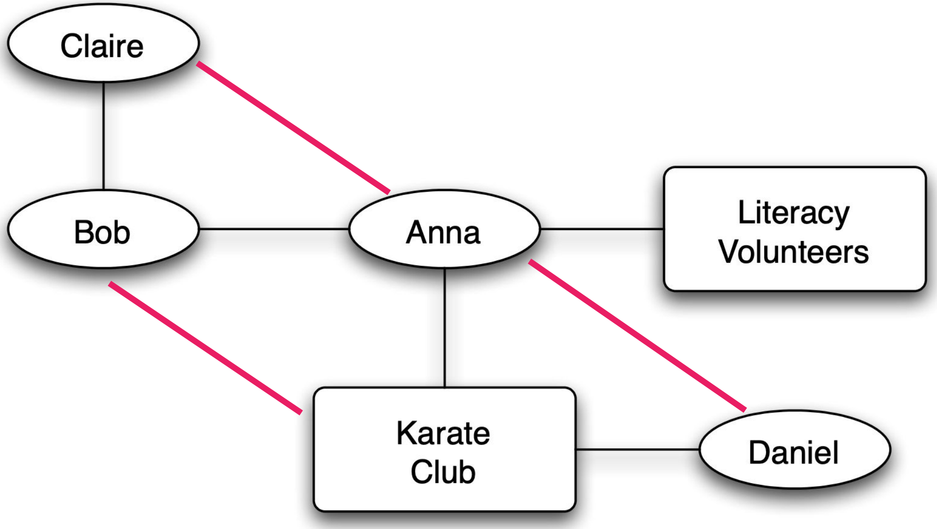


Different mechanisms for link formation can now all be viewed as types of closure processes.

(i) Bob introduces Anna to Claire.

(ii) Karate introduces Anna to Daniel.

(iii) Anna introduces Bob to Karate.



Summary

Affiliation networks can reveal interesting relationships on both sides of the bipartite graph.

We need to rethink many of our one-mode measures.

Examples of Two-Mode Networks

Political polarization and the structure of the board interlock network

- Longer geodesic, less cohesion

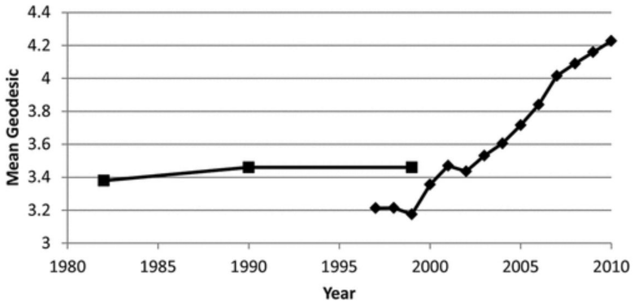


Fig. 1. Mean geodesic in main component of board interlock networks, 1982–2010. Data for 1982–99 are from Davis et al. (2003); 1997–2010, this study; study population differs across the sources.

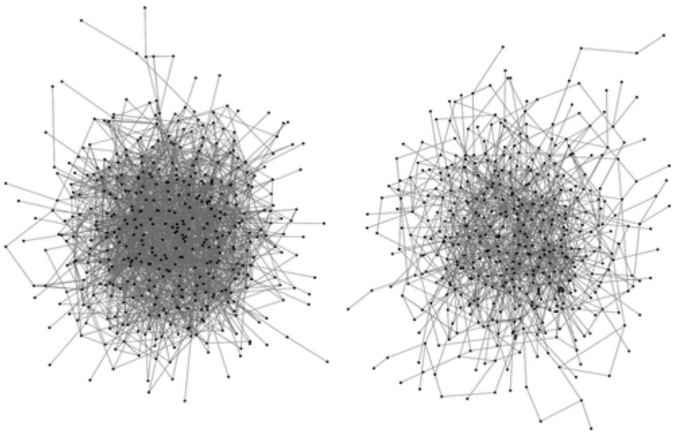


Fig. 2. S&P 500 interlock network main component, 1996 (*left*) and 2010 (*right*)

Chu and Davis 2016

Examples of Two-Mode Networks

Political polarization and the structure of the board interlock network

- Longer geodesic, less cohesion

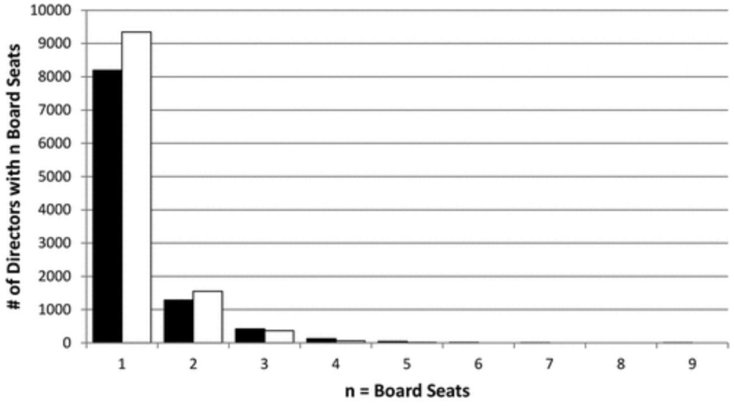


Fig. 3. Distribution of directors by number of S&P 1500 board seats, 2000 (*black bars*) and 2010 (*white bars*).

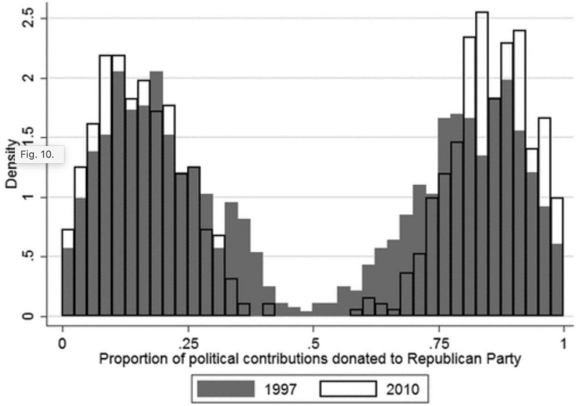


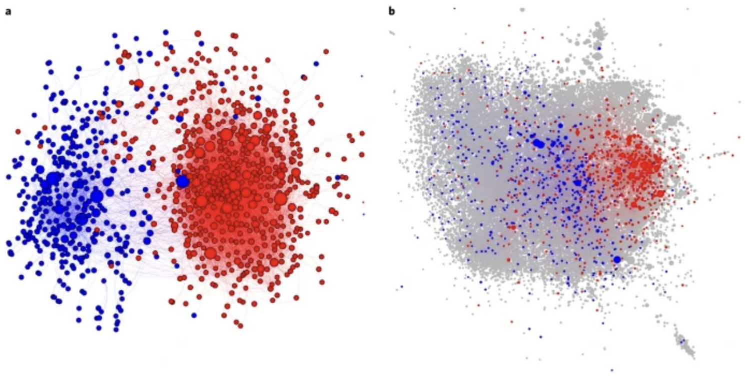
Fig. 10. Simulated distribution of number of executives by proportion of political contributions allocated to the Republican Party.

Examples of Two-Mode Networks

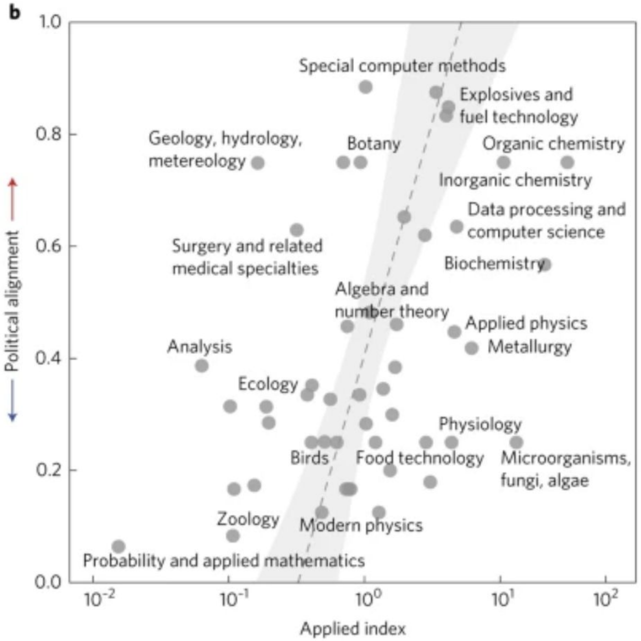
Is science politicized?: Partisan difference in the consumption of science

- Amazon book co-purchase data

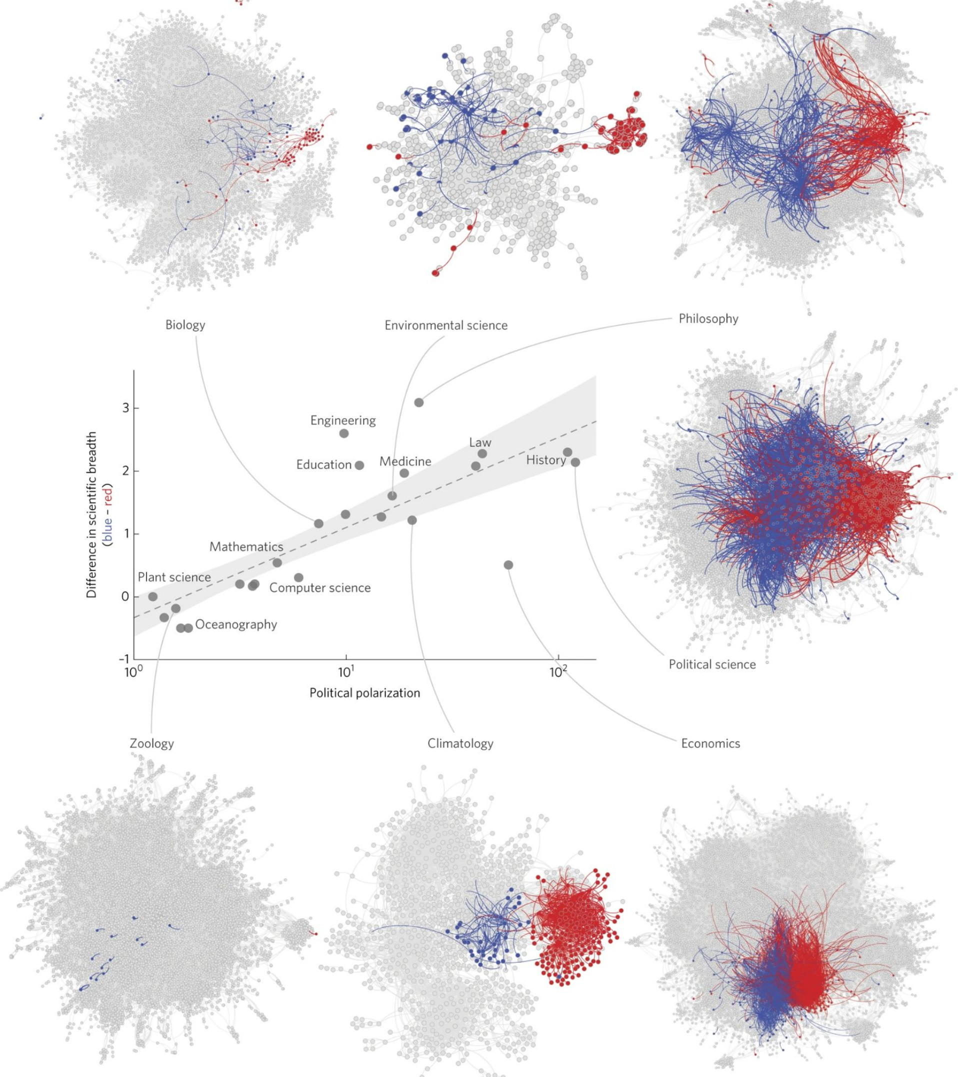
Figure 1: Visualization of the co-purchase network among liberal, conservative and scientific books.

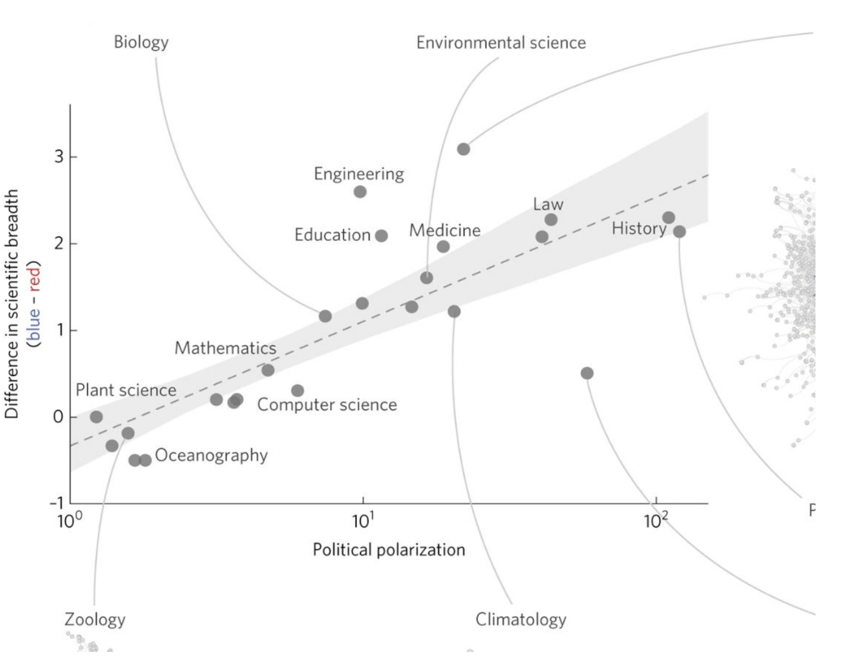


a, Links between 583 liberal (blue) and 673 conservative (red) books. **b**, Links between these books and science (grey) books. As shown in **a**, 97.2% of red books linked to other reds and 93.7% of blue books



Shi et al. 2017





Political books in science disciplines

Shi et al. 2017

Affiliation networks and Prediction

Affiliation networks are also widely used to predict and recommend products to online users

- based on similarities in people's connections to artifacts (affiliations)