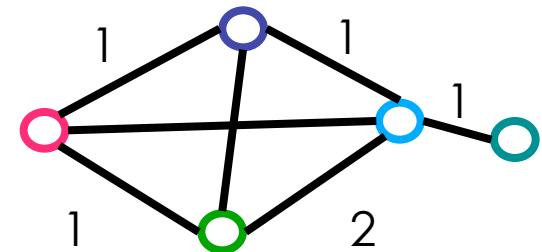
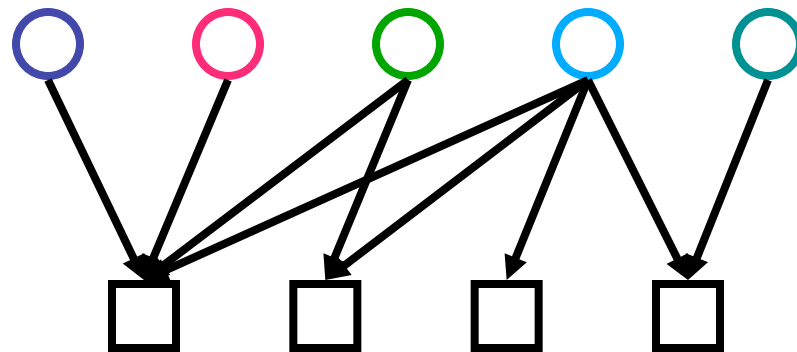


What makes a community?

- mutuality of ties
 - everybody in the group knows everybody else
- frequency of ties among members
 - everybody in the group has links to at least k others in the group
- closeness or reachability of subgroup members
 - individuals are separated by at most n hops
- relative frequency of ties among subgroup members compared to nonmembers

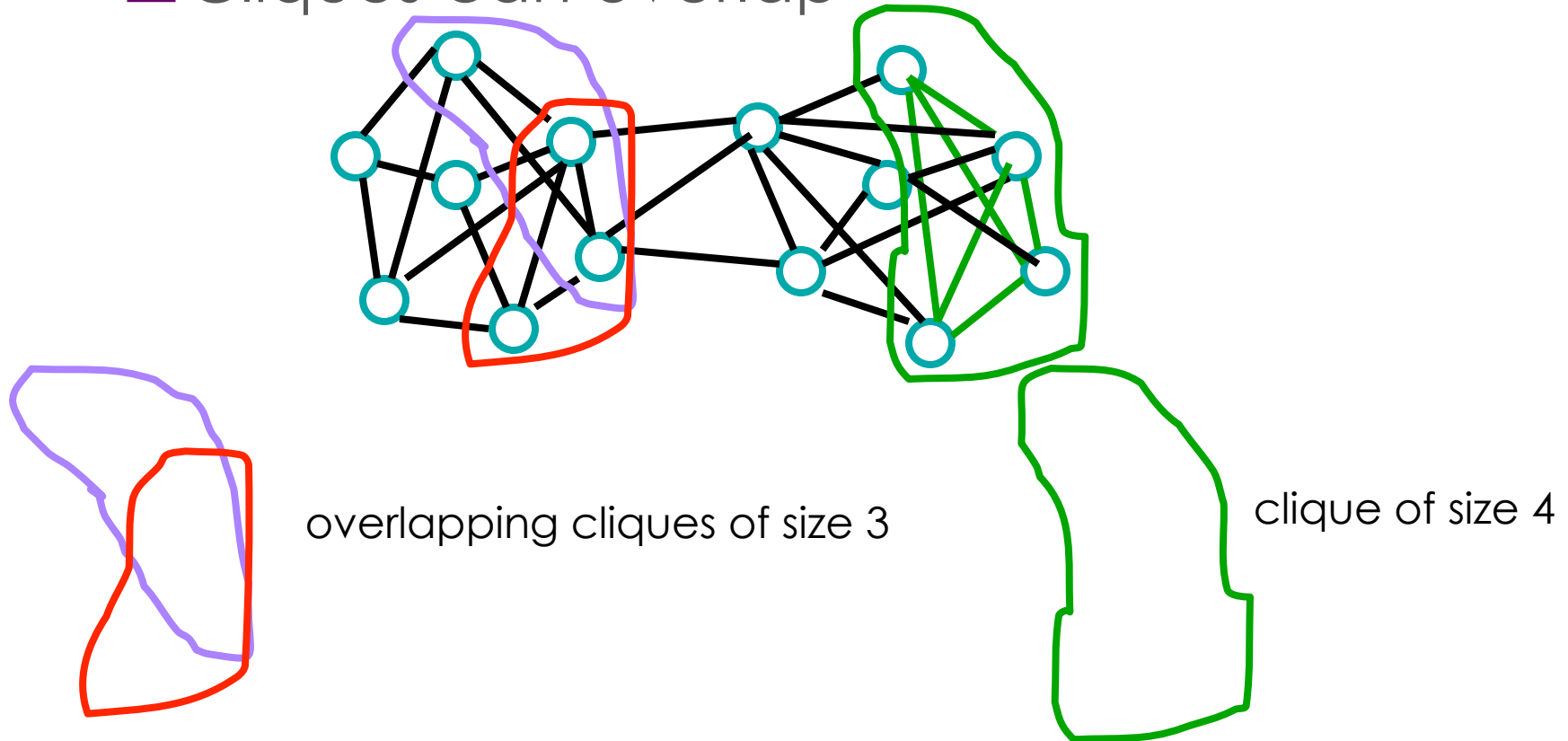
Affiliation networks

- otherwise known as
 - membership network
 - e.g. board of directors
 - hypernetwork or hypergraph
 - bipartite graphs
 - interlocks



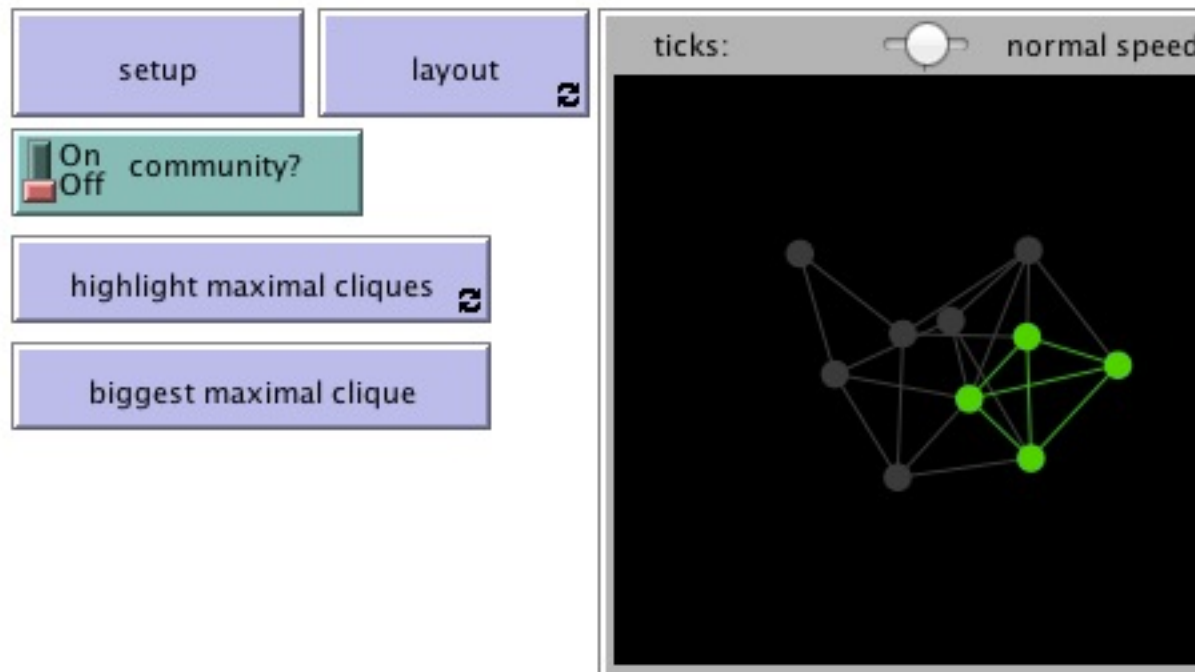
Cliques

- Every member of the group has links to every other member
- Cliques can overlap



Cliques betray community structure

- Go to <http://www.ladamic.com/netlearn/nw/Cliques.html>
- Try the ER vs. community structure setup (they are the same as for the opinion formation model)



Quiz question

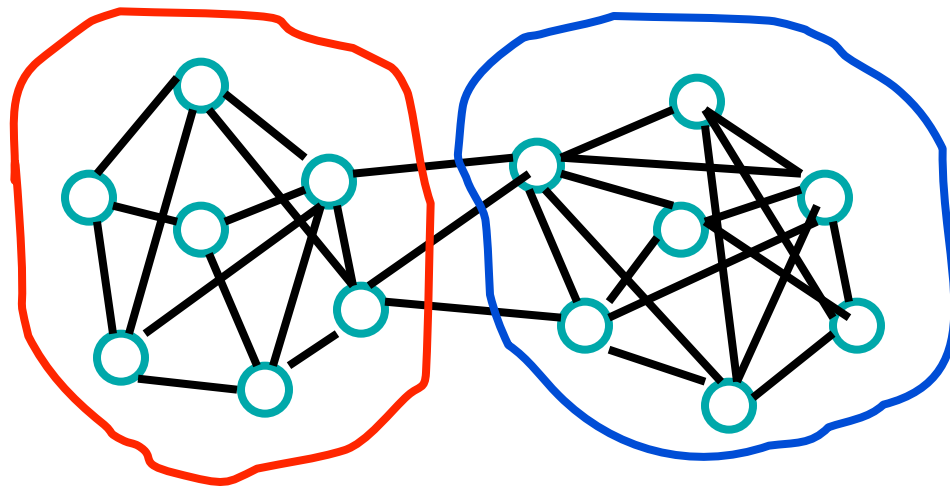
- Which has a larger maximal clique?
 - network with community structure
 - the equivalent ER random graph

Meaningfulness of cliques

- ❑ Not robust
 - ❑ one missing link can disqualify a clique
 - ❑ Not interesting
 - ❑ everybody is connected to everybody else
 - ❑ no core-periphery structure
 - ❑ no centrality measures apply
 - ❑ How cliques overlap can be more interesting than that they exist
-

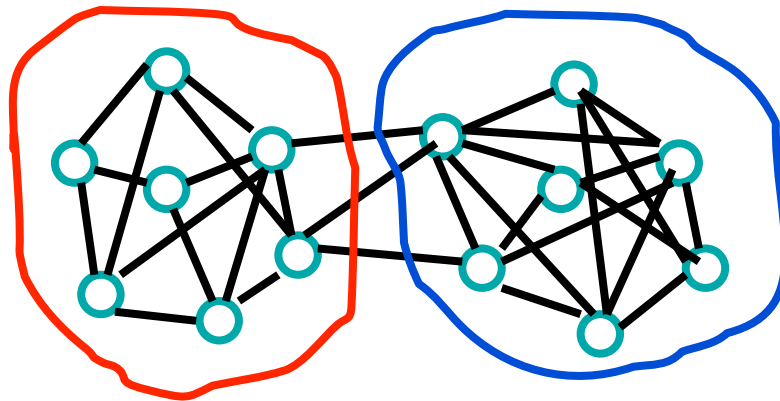
k-cores: similar idea, less stringent

- Each node within a group is connected to k other nodes in the group



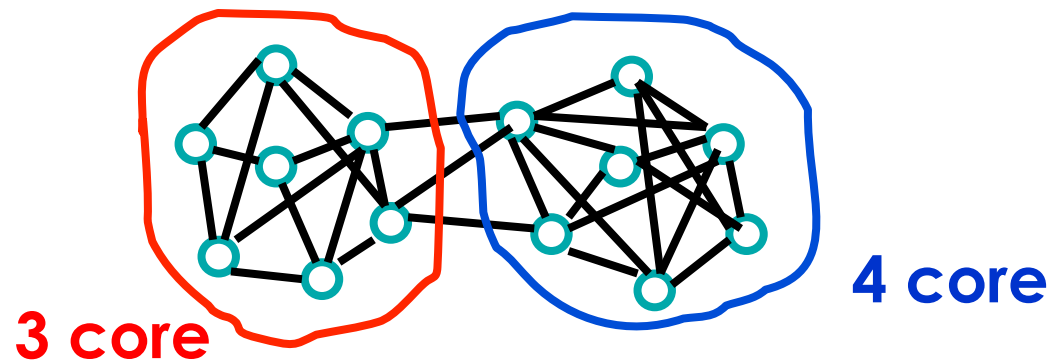
Quiz Question

- What is the “k” for the core circled in red?
- What is the “k” for the core circled in blue?

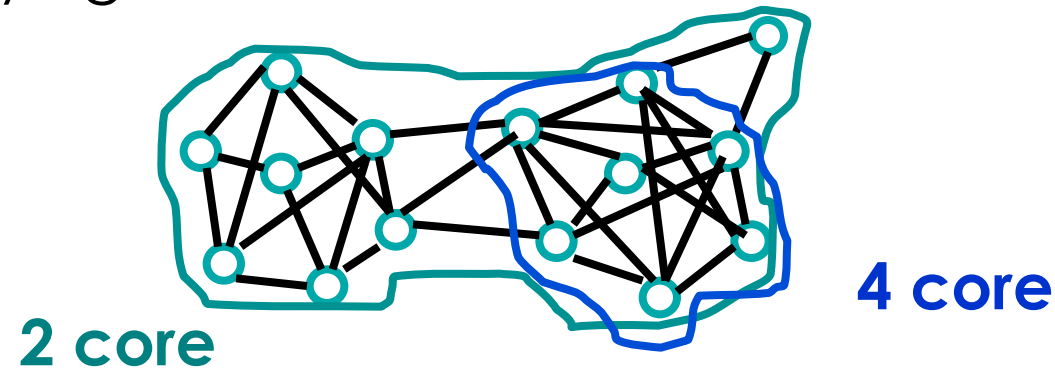


k-cores

- Each node within a group is connected to k other nodes in the group

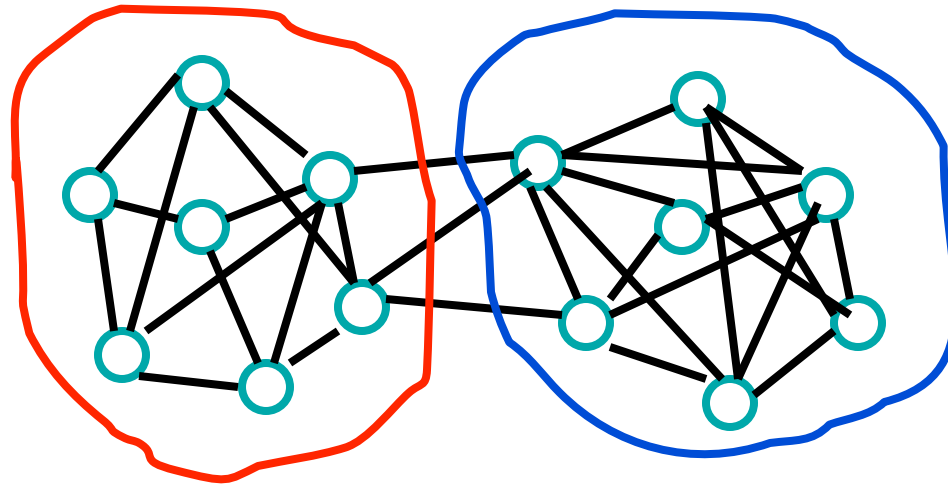


- but even this is too stringent of a requirement for identifying natural communities



subgroups based on reachability and diameter

- n – cliques
 - maximal distance between any two nodes in subgroup is n



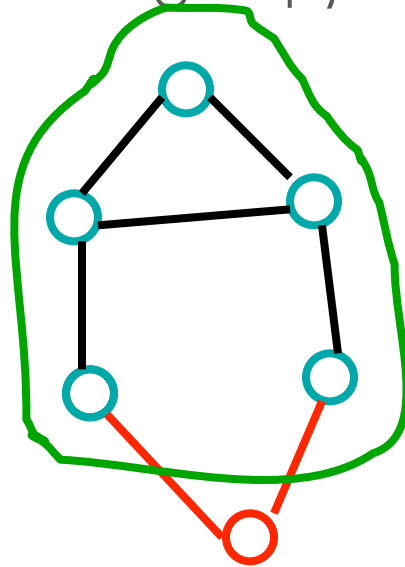
2-cliques

- theoretical justification
 - information flow through intermediaries

considerations with n-cliques

▣ problem

- ▣ diameter may be greater than n
- ▣ n -clique may be disconnected (paths go through nodes not in subgroup)



2 – clique
diameter = 3

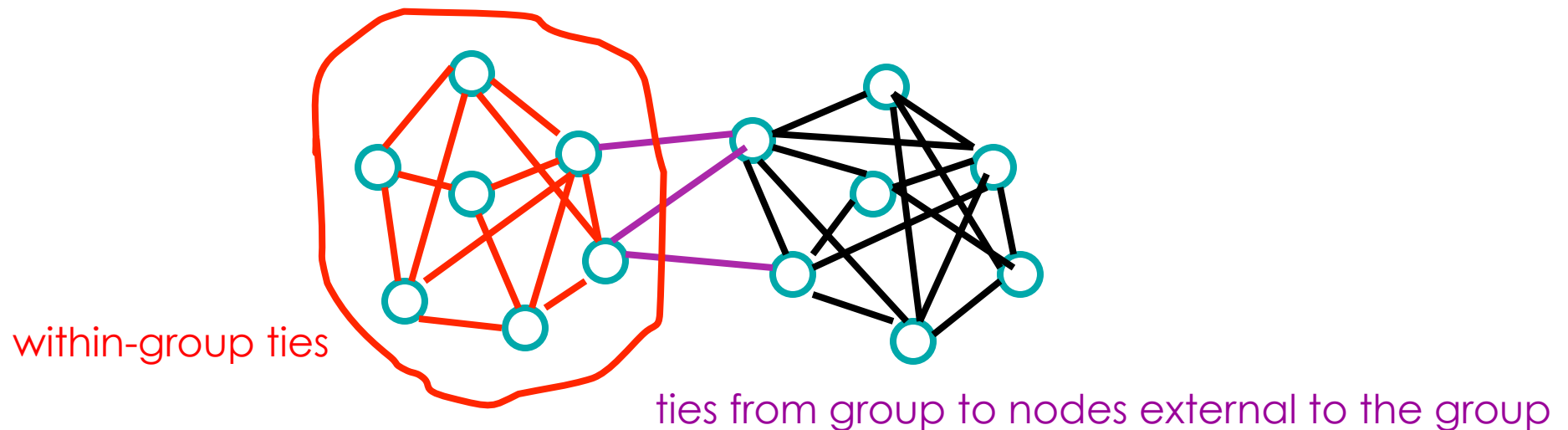
path outside the 2-clique

■ fix

- n -club: maximal subgraph of diameter 2

p-cliques: frequency of in group ties

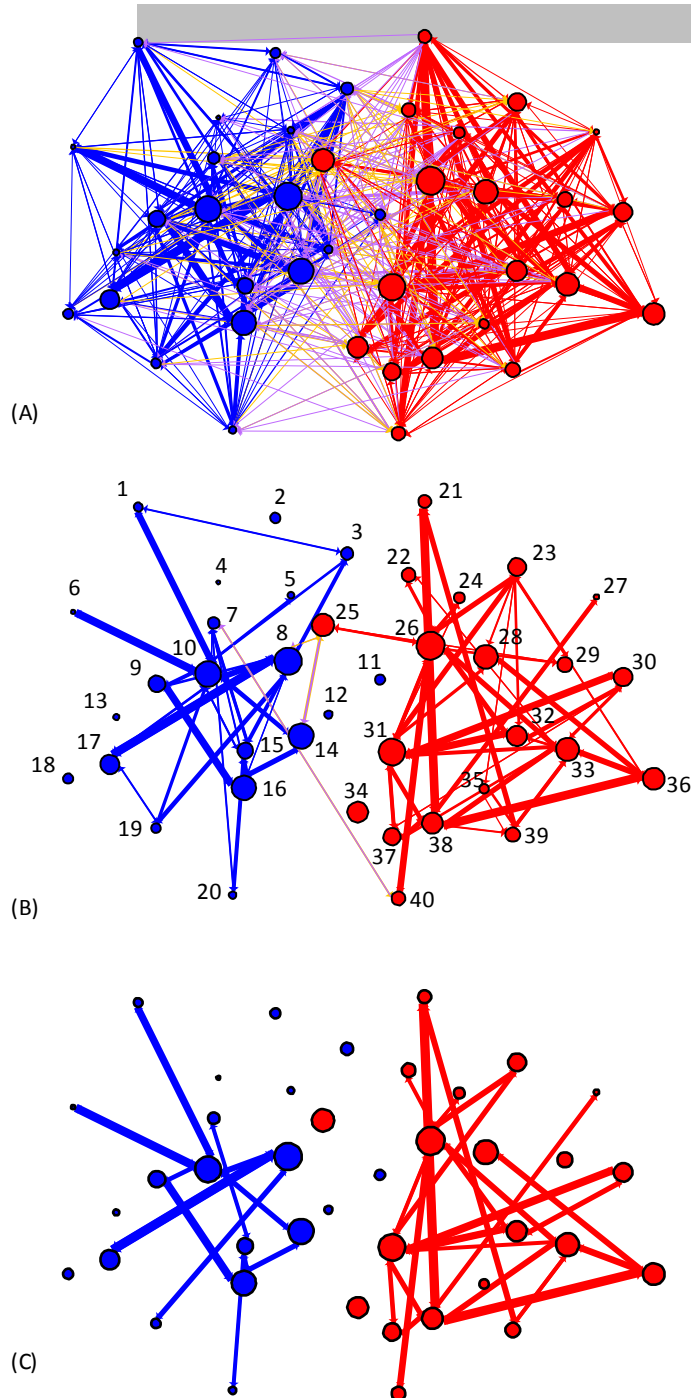
- partition the network into clusters where vertices have at least a proportion p (number between 0 and 1) of neighbors inside the cluster.



cohesion in directed & weighted networks

- something we've already learned how to do:
 - find strongly connected components

 - keep only a subset of ties before finding connected components
 - reciprocal ties
 - edge weight above a threshold
-



- 1 Digbys Blog
- 2 James Walcott
- 3 Pandagon
- 4 blog.johnkerry.com
- 5 Oliver Willis
- 6 America Blog
- 7 Crooked Timber
- 8 Daily Kos
- 9 American Prospect
- 10 Eschaton
- 11 Wonkette
- 12 Talk Left
- 13 Political Wire
- 14 Talking Points Memo
- 15 Matthew Yglesias
- 16 Washington Monthly
- 17 MyDD
- 18 Juan Cole
- 19 Left Coaster
- 20 Bradford DeLong
- 21 JawaReport
- 22 Voka Pundit
- 23 Roger L. Simon
- 24 Tim Blair
- 25 Andrew Sullivan
- 26 Instapundit
- 27 Blogs for Bush
- 28 Little Green Footballs
- 29 Belmont Club
- 30 Captain's Quarters
- 31 Powerline
- 32 Hugh Hewitt
- 33 INDC Journal
- 34 Real Clear Politics
- 35 Winds of Change
- 36 Allahpundit
- 37 Michelle Malkin
- 38 WizBang
- 39 Dean's World
- 40 Volokh

Example: political blogs (Aug 29th – Nov 15th, 2004)

- A) all citations between A-list blogs in 2 months preceding the 2004 election
- B) citations between A-list blogs with at least 5 citations in both directions
- C) edges further limited to those exceeding 25 combined citations

only 15% of the citations bridge communities