Structural analysis of signed networks by optimally clustering them using integer programming models

Samin Aref (University of Toronto)
Joint work with Mark Wilson, Andrew Mason, Zachary Neal, Ly Dinh, Shadi Rezapour, and Jana Diesner

Workshop on Signed Relations and Structural Balance in Complex Systems
ETH Zurich, 2024-05-17

Other network problem that I've been working on more recently

Home › Computational Science – ICCS 2023 › Conference paper

Heuristic Modularity Maximization Algorithms for Community Detection Rarely Return an Optimal Partition or Anything Similar

Conference paper | Open Access | First Online: 26 June 2023

pp 612–626 | Cite this conference paper

Bayan globally maximizes modularity for networks with up to 3000 edges on ordinary computers

```
%pip install bayanpy
```

```
import networkx as nx
import bayanpy
```

```
G = nx.barbell_graph(5,2)
```

```
bayanpy.bayan(G)
```

Paper 1: doi.org/10.1007/978-3-031-36027-5_48
Paper 2: doi.org/10.1016/j.jocs.2024.102283
Paper 3: arxiv.org/pdf/2310.10898
GitHub Repo: github.com/saref/bayan
Project website: bayanproject.github.io

Try Bayan on your smartphone ->

Google Colab examples:
tinyurl.com/bayancolab

UNIVERSITY OF TORONTO

In *balanced* signed networks:

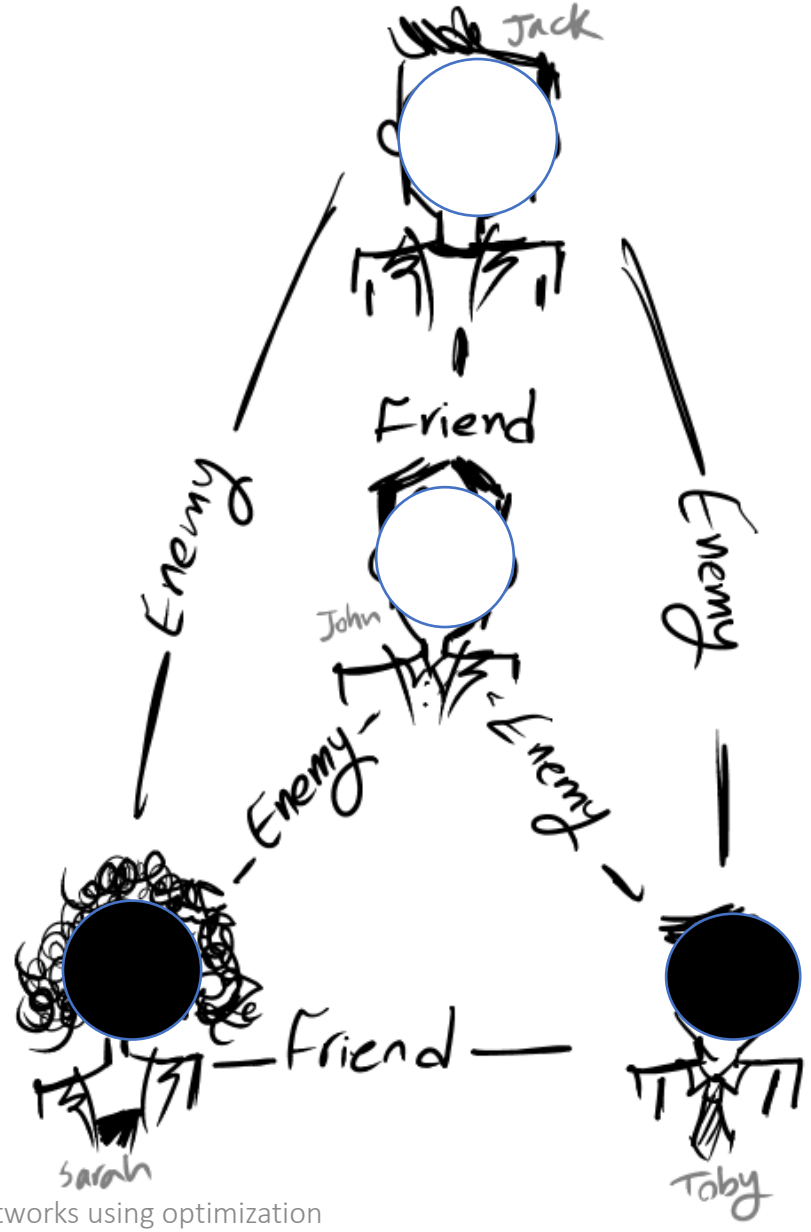
Enemy of an enemy = friend
Friend of a friend = friend
Enemy of a friend = enemy
Friend of an enemy = enemy

UNIVERSITY OF TORONTO

Balanced

In *balanced* signed networks:

Enemy of an enemy = friend ✓
Friend of a friend = friend ✓
Enemy of a friend = enemy ✓
Friend of an enemy = enemy ✓



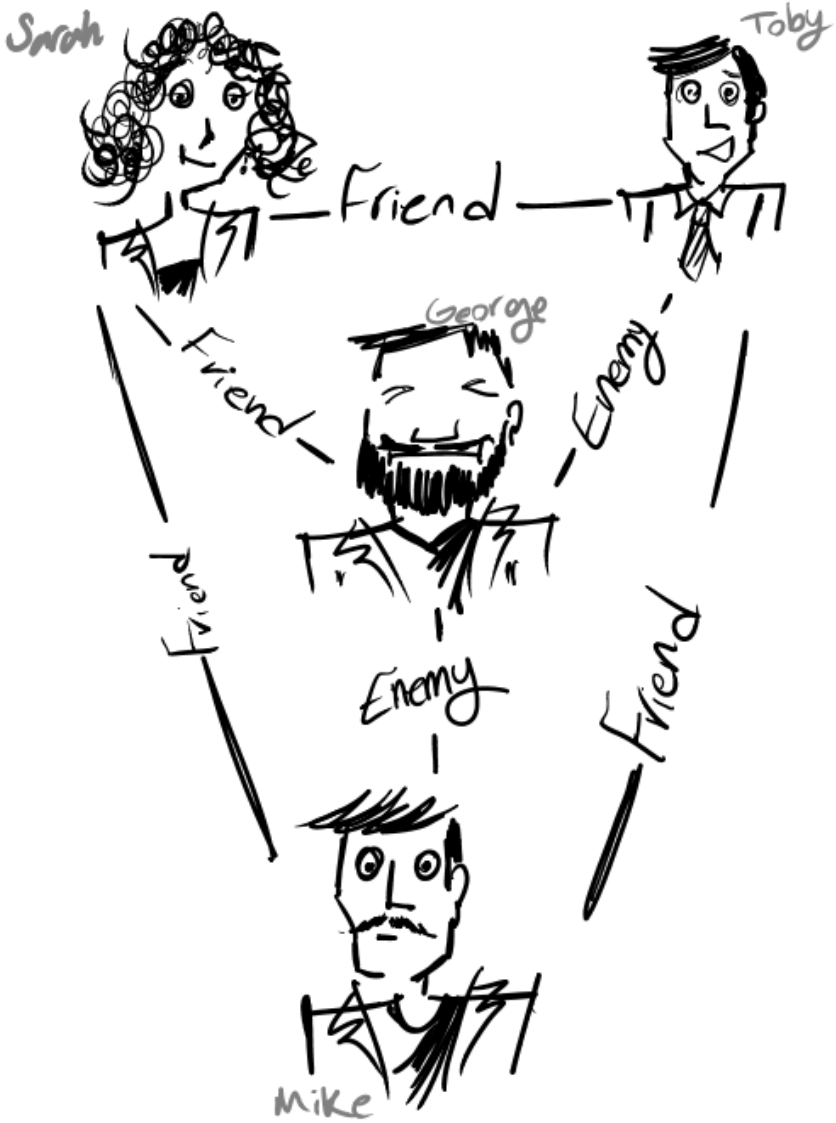
Unbalanced

Enemy of an enemy = friend ✓

Friend of a friend = friend ✗

Enemy of a friend = enemy ✗

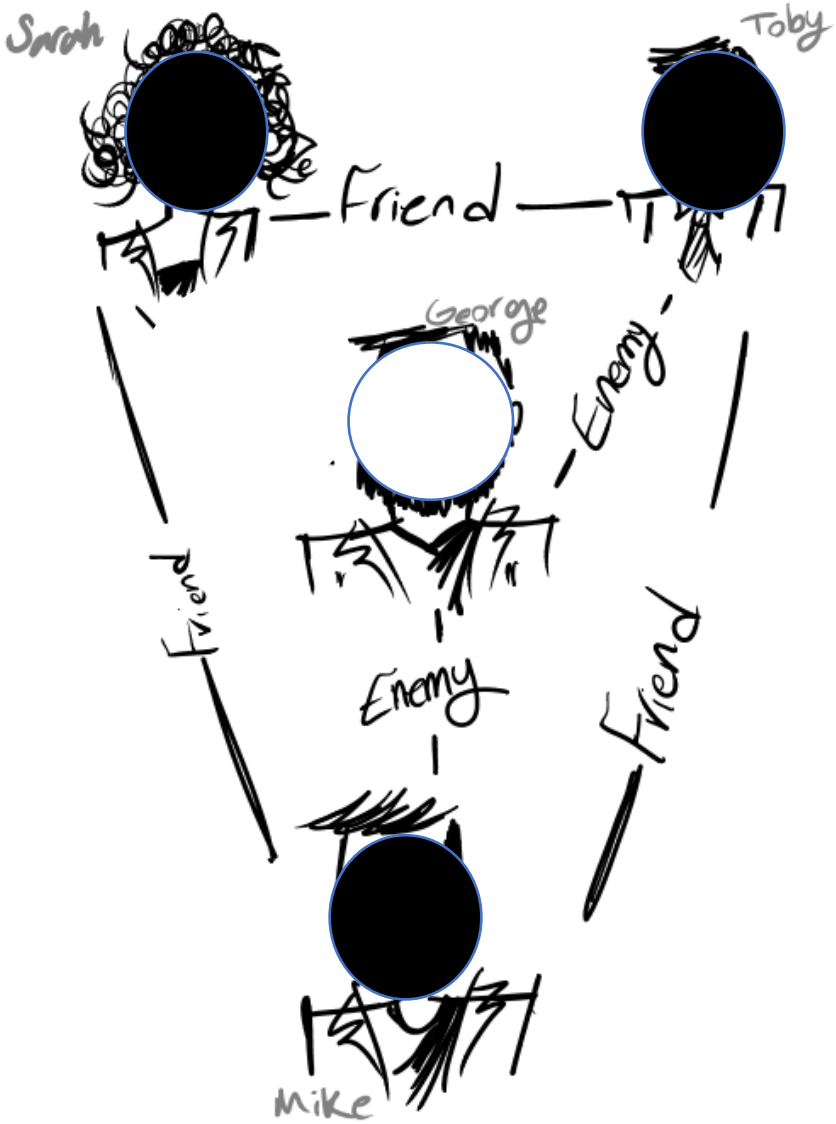
Friend of an enemy = enemy ✗



Balanced subgraph

It is 1 edge away from balance

- The *frustration index* is 1
- The *line index of balance* is 1



Excerpts from Cartwright and Harary 1956

Thus we define an *s-graph* (containing any number of points) as *balanced* if all of its cycles are positive.

Figure 5 illustrates this definition for four s-graphs containing four points. In each of these s-graphs there are seven cycles: *AB*, *BC*, *CA*; *AB*, *BD*, *DA*; *BC*, *CD*, *DB*; *AC*, *CD*, *DA*; *AB*, *BC*, *CD*, *DA*; *AB*, *BD*, *DC*, *CA*; and *BC*, *CA*, *AD*, *DB*. It will be seen that in s-graphs *a* and *b* all seven cycles are positive, and these s-graphs are therefore balanced. In

2. *Units containing more than three entities.* Nearly all theorizing about balance has referred to units of three entities. While Horowitz, Lyons, and Perlmutter studied units with four entities, they did not *define* balance for such cases. It would seem desirable to be able to speak of the balance of even larger units.

Although these authors indicate the possibility of treating the *P-O-Q-X* unit in terms of balance, they do not develop a formal definition of a balanced configuration consisting of four elements. They seem to imply that the *P-O-Q-X* unit will be balanced if the *P-O-X* and the *P-Q-X* units are both balanced. They do not consider the relation between *Q* and *O*, nor the logically possible components of which it could be a part. Their analysis is

UNIVERSITY OF TORONTO

Undirected signed graph $G = (V, E, \sigma)$

$$\sigma : E \rightarrow \{-1, +1\}$$

$$|V| = n$$

$$|E| = m = m^{+} + m^{-}$$

$$|E^{+}| = m^{+}$$

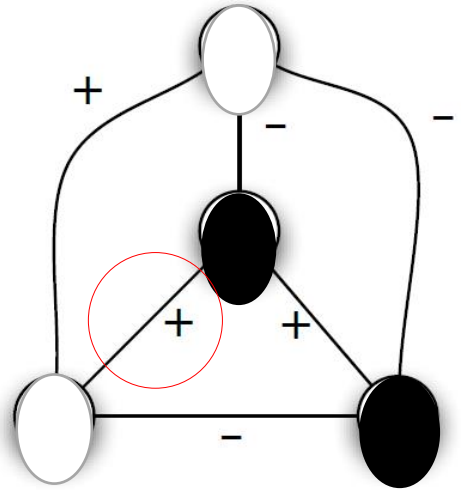
$$|E^{-}| = m^{-}$$

Defining frustration based on coloring nodes

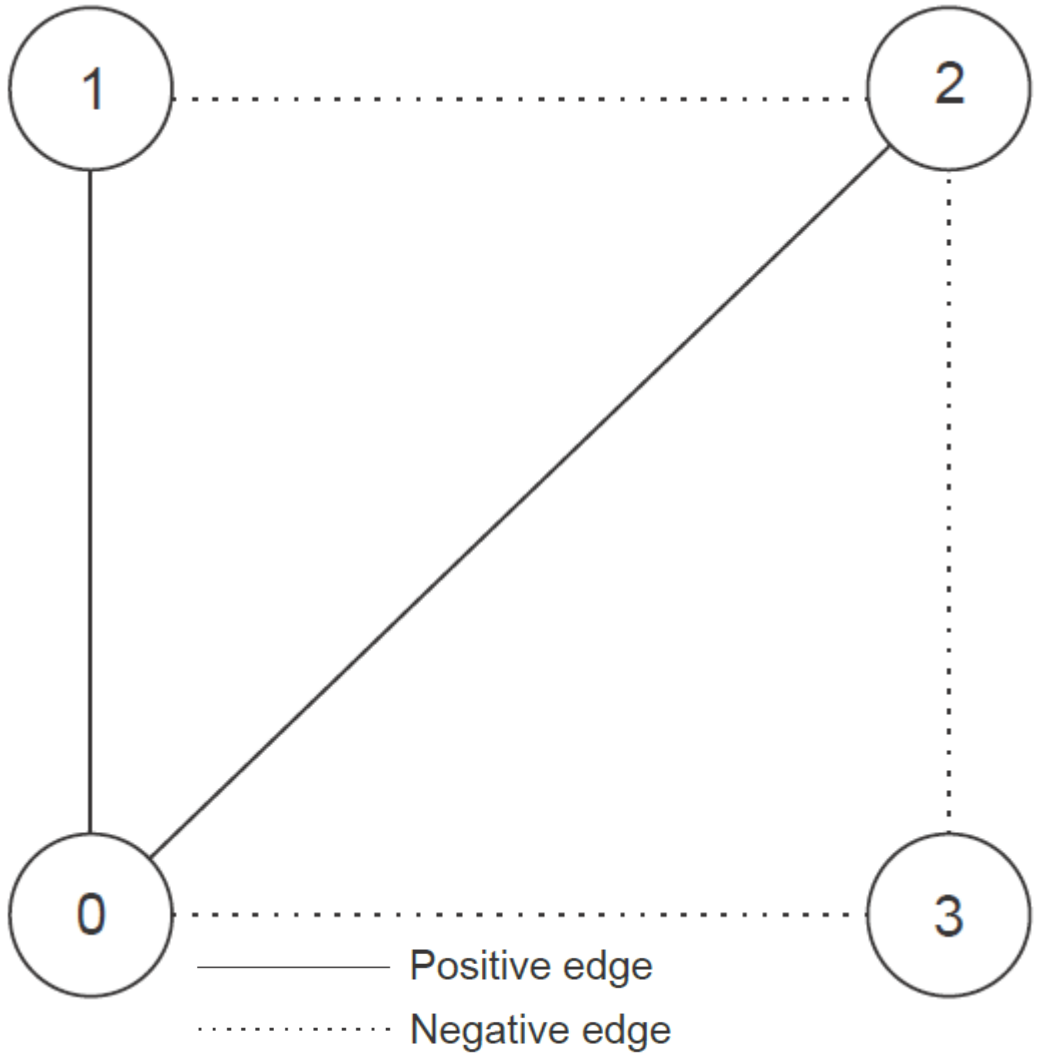
Frustrated edges:

-positive edges with different endpoint colors

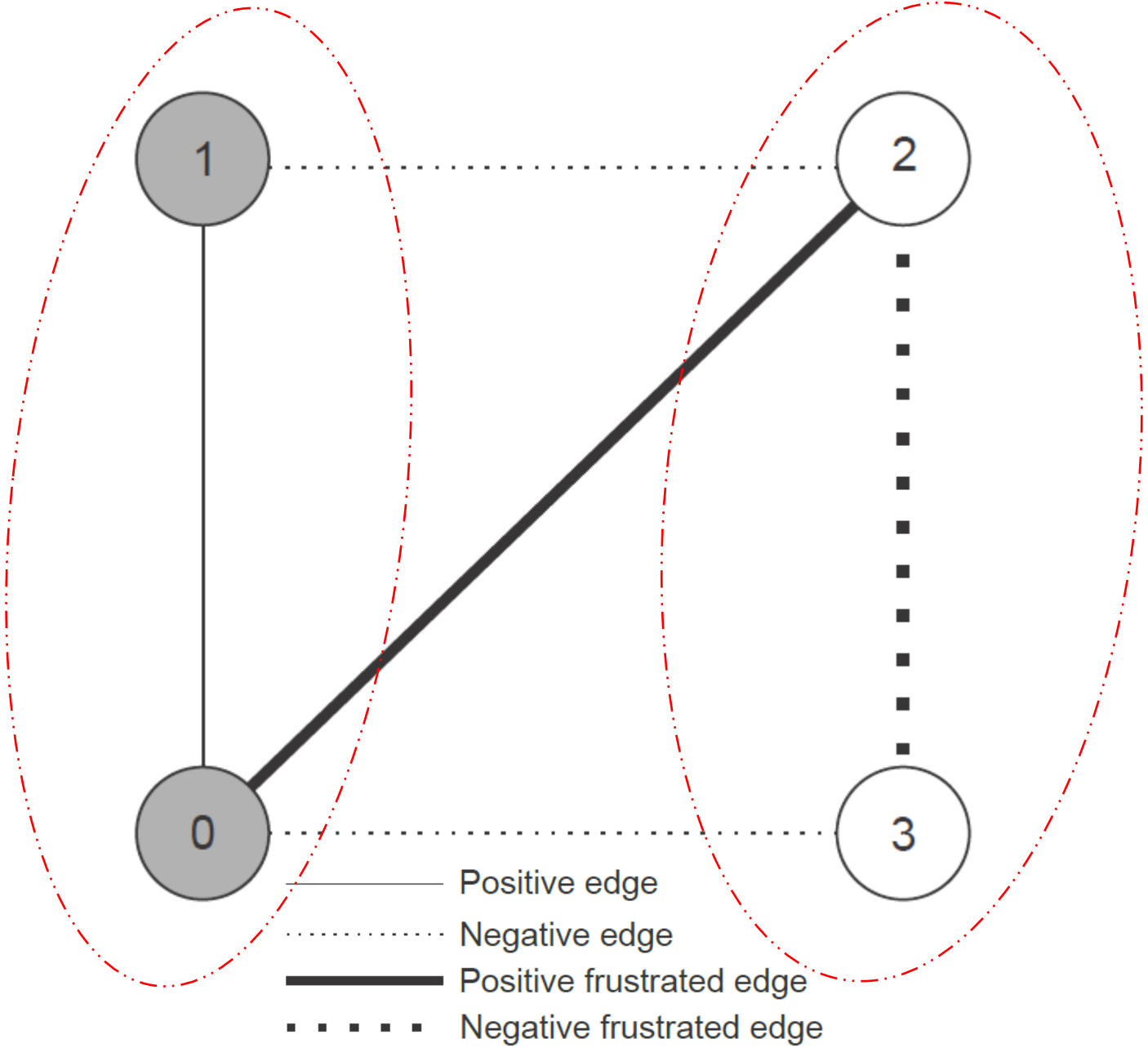
-negative edges with the same endpoint colors



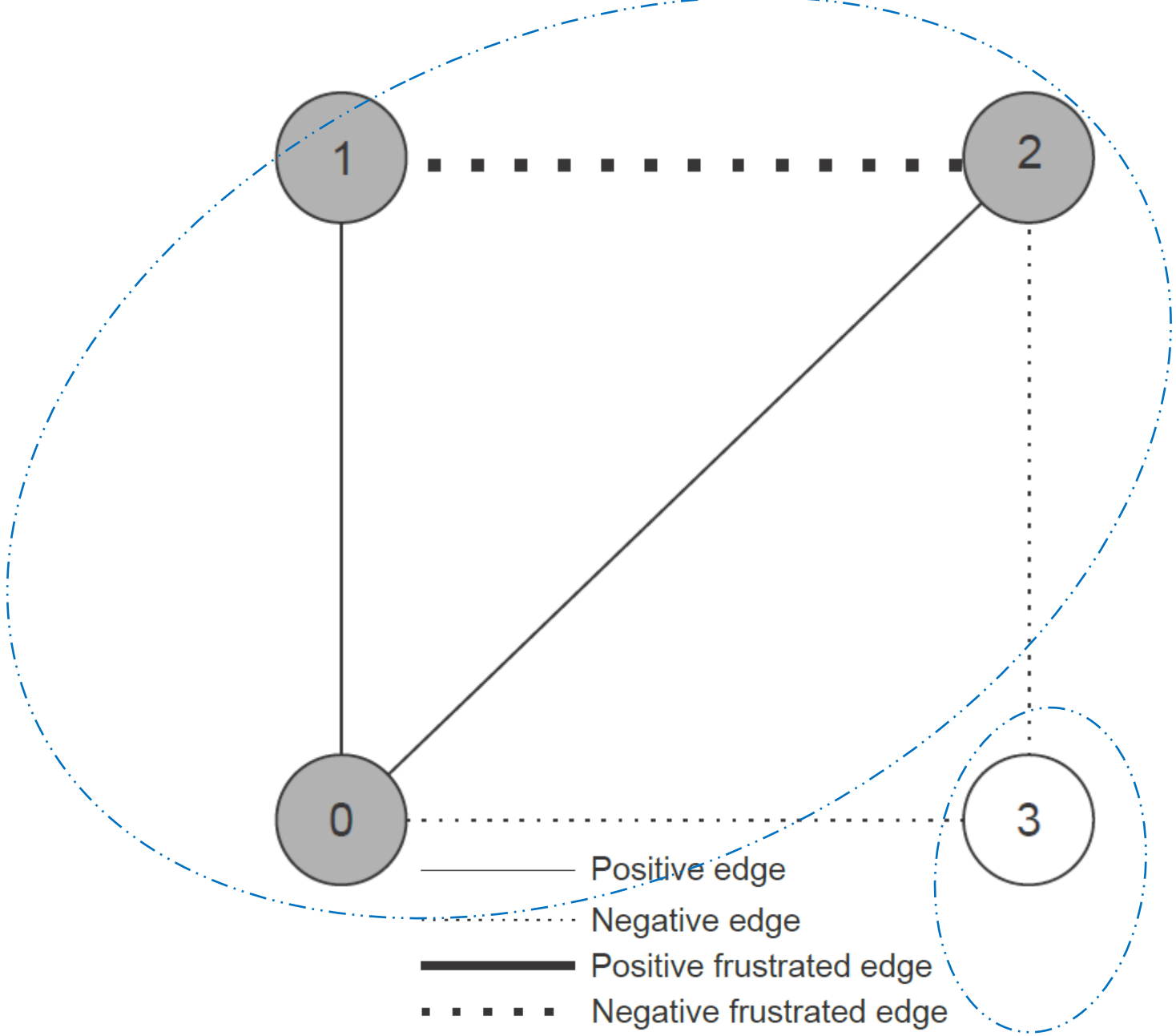
Figures: D. Easley and J. Kleinberg, Networks, crowds, and markets: Reasoning about a highly connected world. pp 122 Cambridge University Press, 2010



UNIVERSITY OF TORONTO



UNIVERSITY OF TORONTO



UNIVERSITY OF TORONTO

Minimizing the number of frustrated edges

x_i : binary variable representing color of node i

f_{ij} : binary variable representing the frustration of the signed edge (i, j) based on endpoint colors

$$f_{ij} = x_i + x_j - 2x_i x_j \; \forall (i,j) \in E^+$$

$$f_{ij} = 1 - (x_i + x_j - 2x_i x_j) \; \forall (i,j) \in E^-$$

$$\min_{x_i : i \in V} Z = \sum_{(i,j) \in E^+} x_i + x_j - 2x_i x_j + \sum_{(i,j) \in E^-} 1 - (x_i + x_j - 2x_i x_j)$$

$$\text{s.t.} \quad x_i \in \{0, 1\} \quad \forall i \in V$$

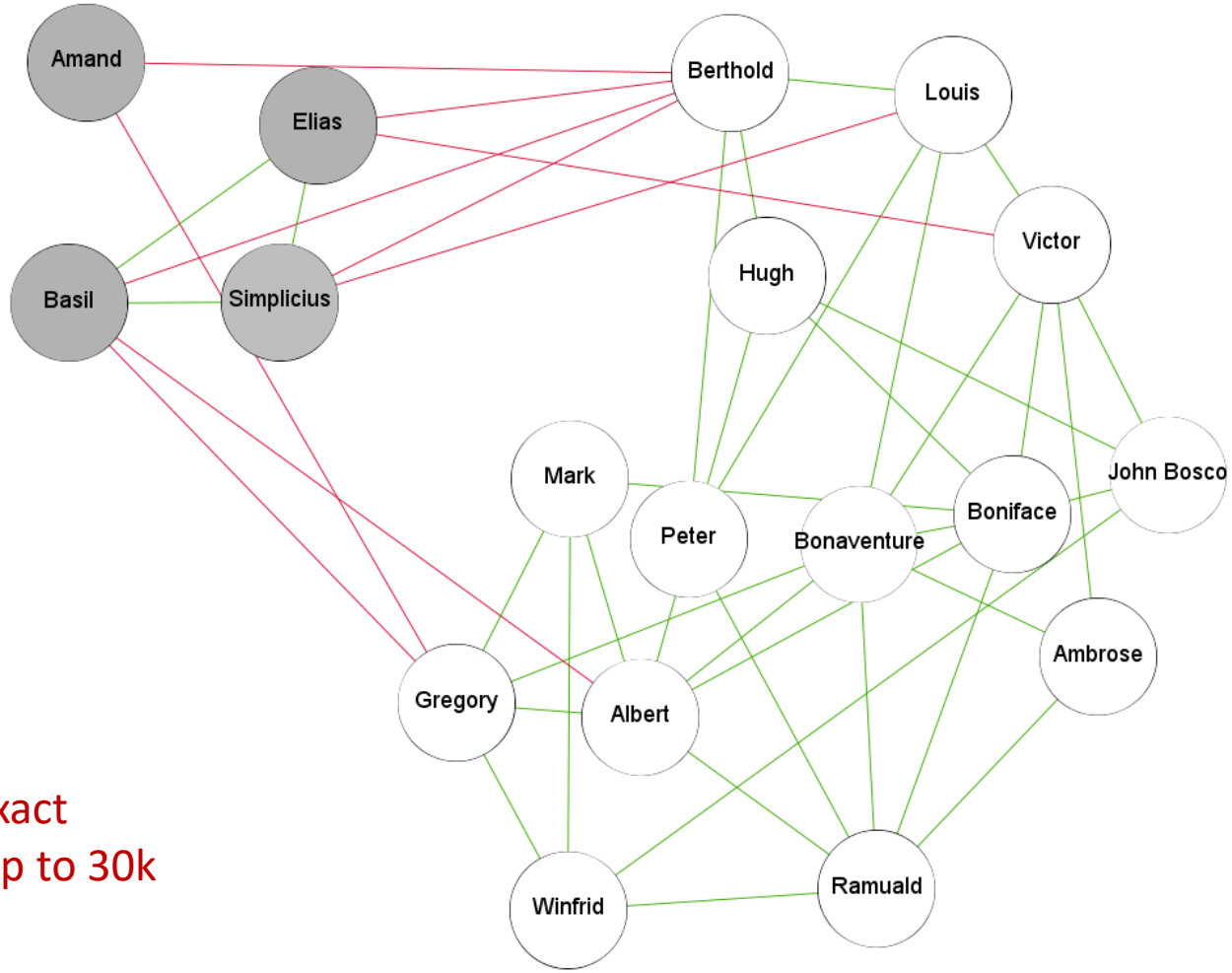
UNIVERSITY OF TORONTO

Sampson's monastery interactions

Social context

Nodes: people (monks)
Edges: top choices for like and dislike

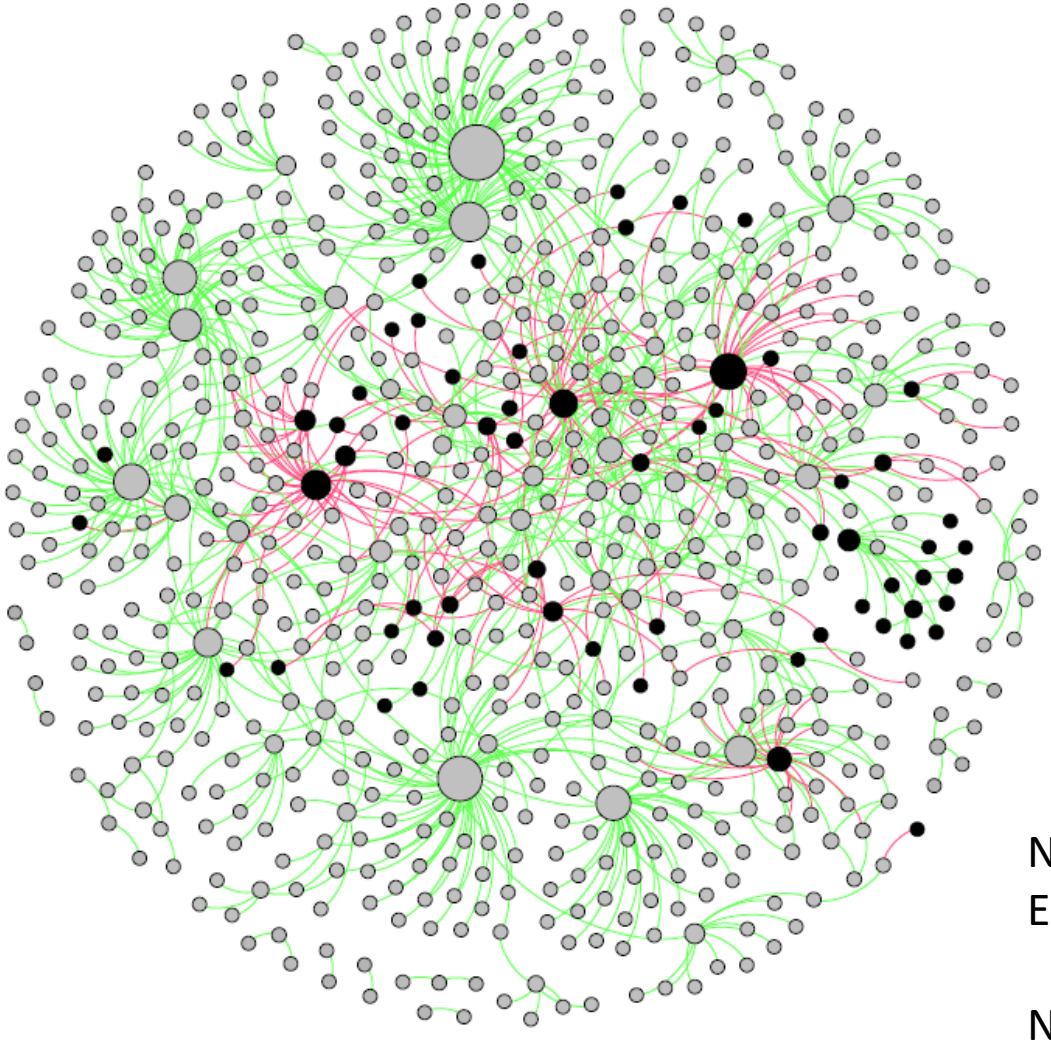
Negative edges are shown in red



See github.com/saref/frustration-index-XOR for exact calculation of frustration index in networks with up to 30k edges (strong balance)

Baker's yeast

Biological context



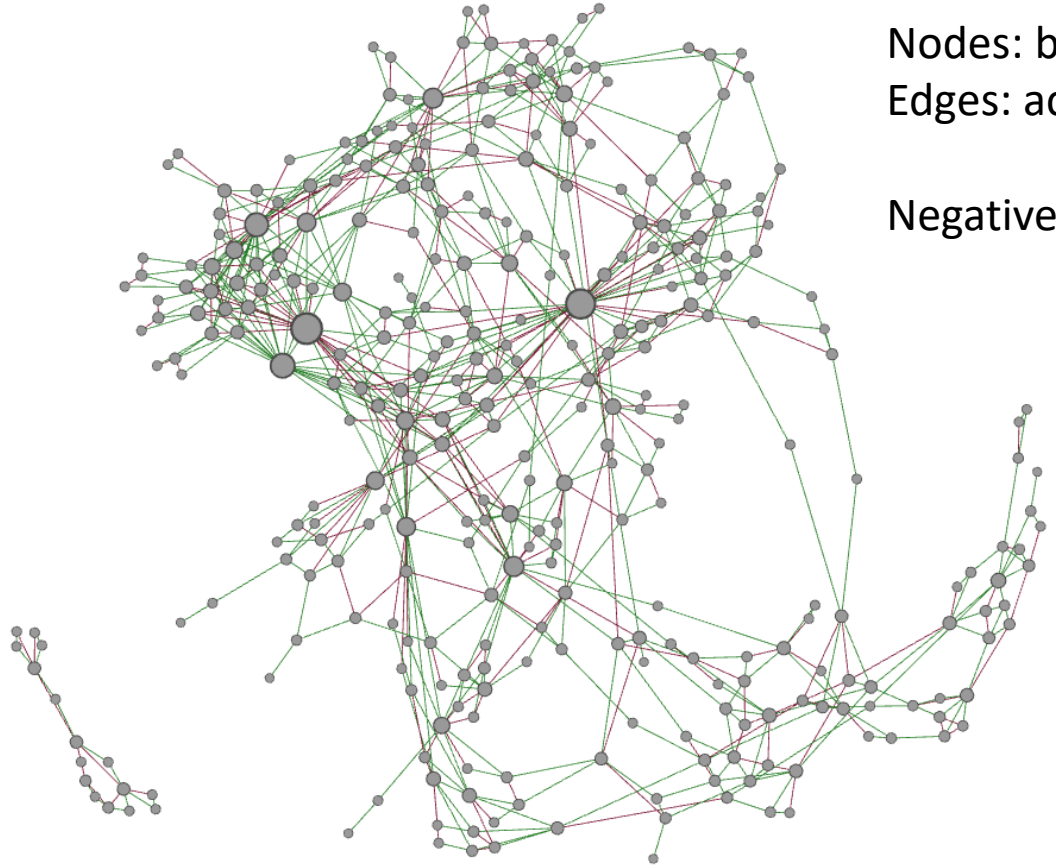
n	m	m^-
690	1080	220

Nodes: biological molecules
Edges: activation or inhibition relations

Negative edges are shown in red

Signed network of a protein (EGFR)

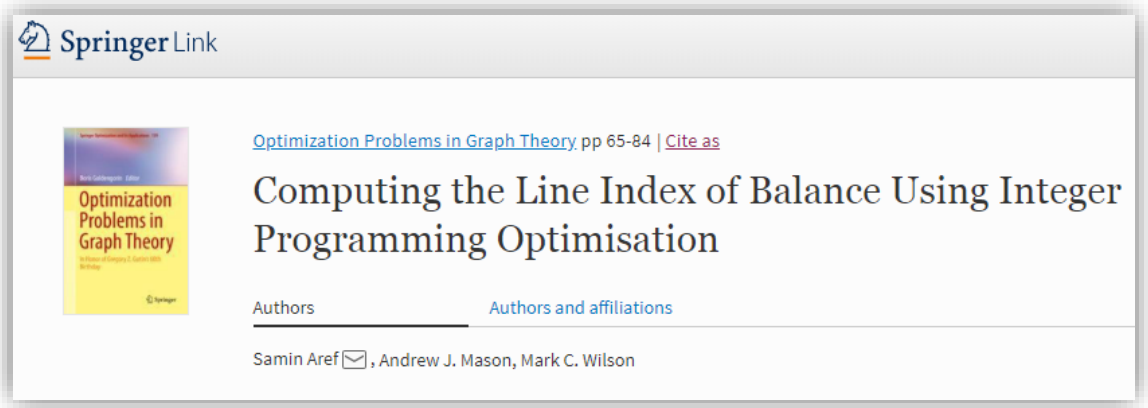
Biological context



Nodes: biological molecules
Edges: activation or inhibition relations

Negative edges are shown in red

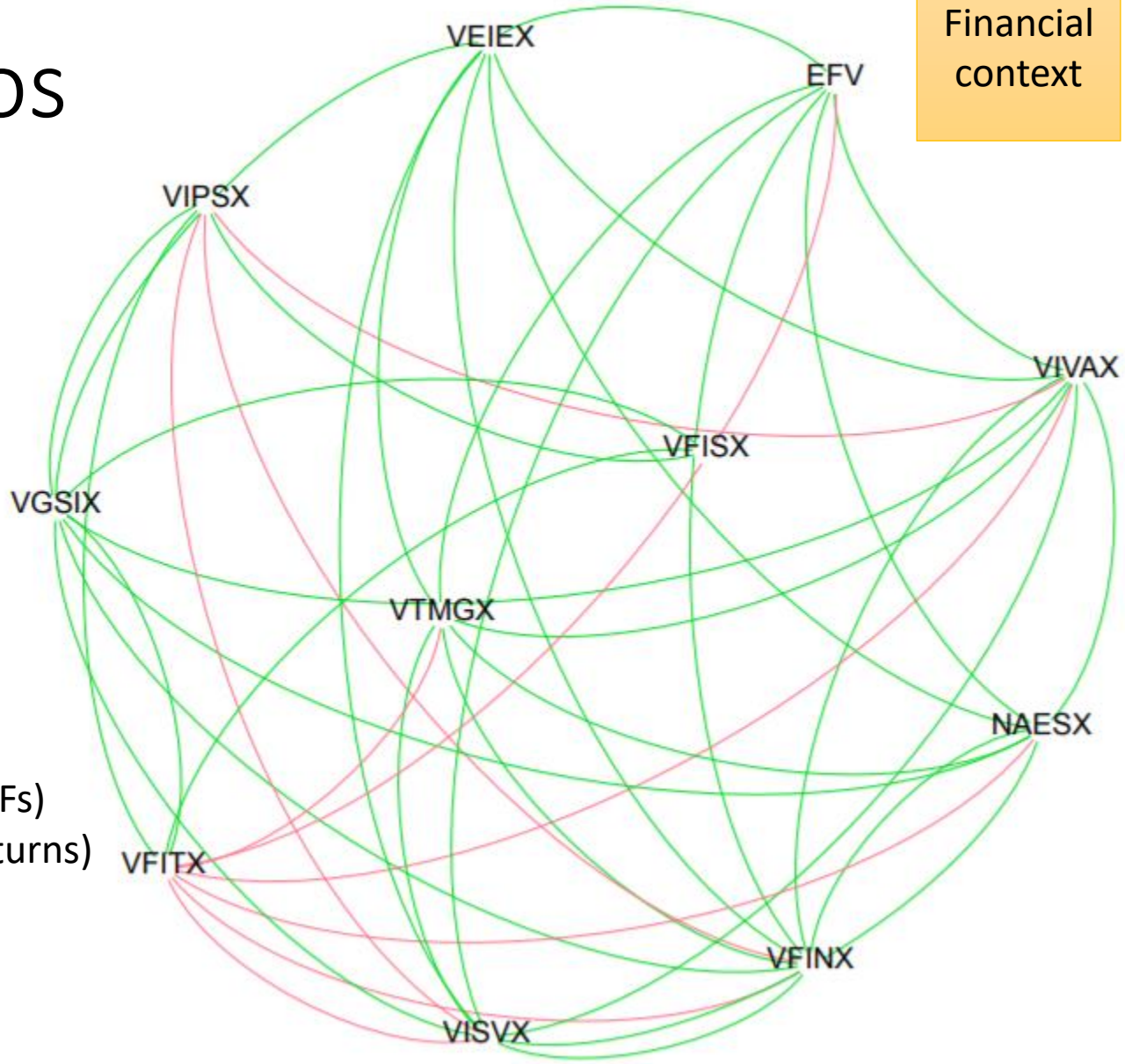
n	m	m^-
329	779	264



UNIVERSITY OF TORONTO

Financial portfolios

Financial context

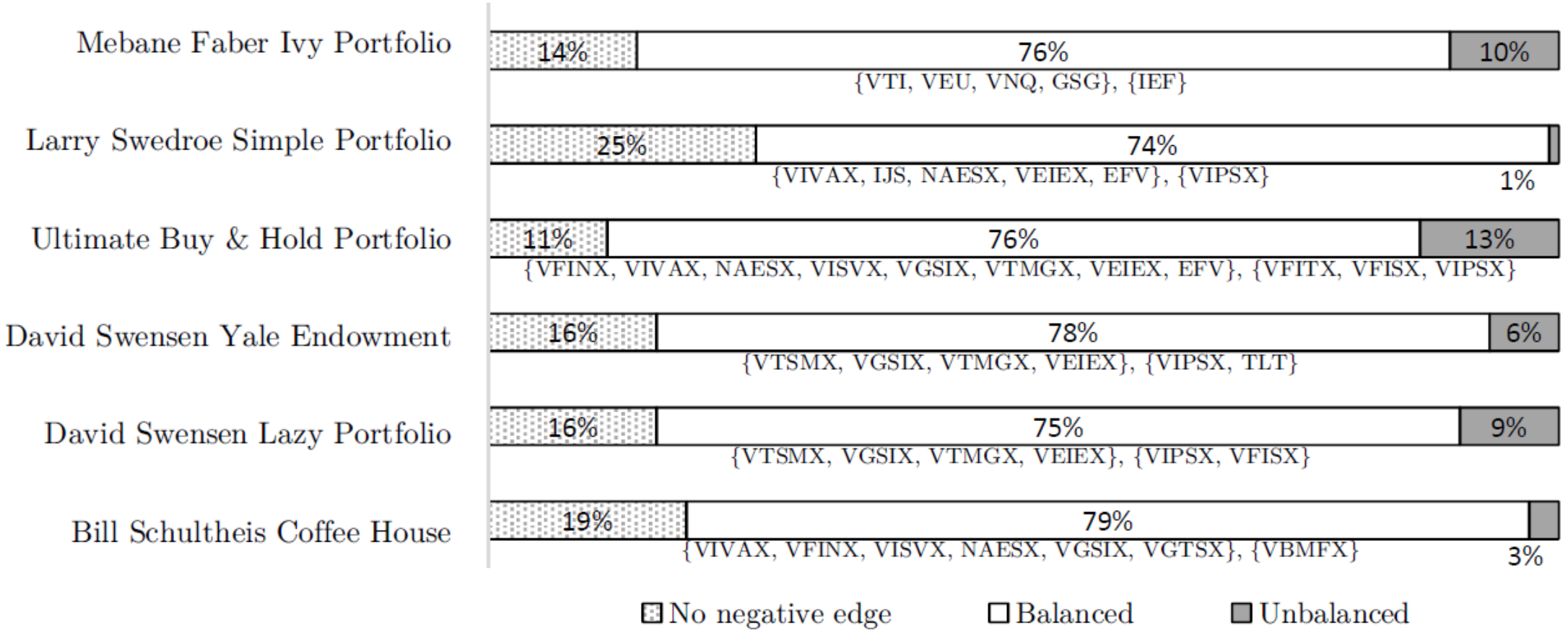


Nodes: securities (exchange traded funds - ETFs)
Edges: positive or negative correlations (of returns)
Network: portfolio

UNIVERSITY OF TORONTO

Financial portfolios
(within the 2008-2016 period)

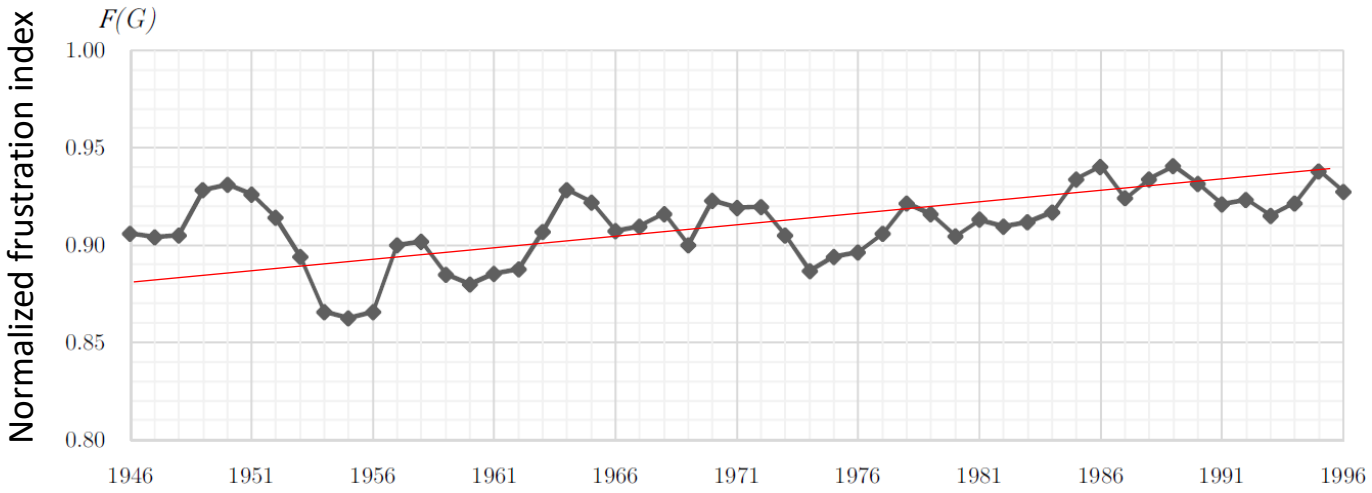
Financial context

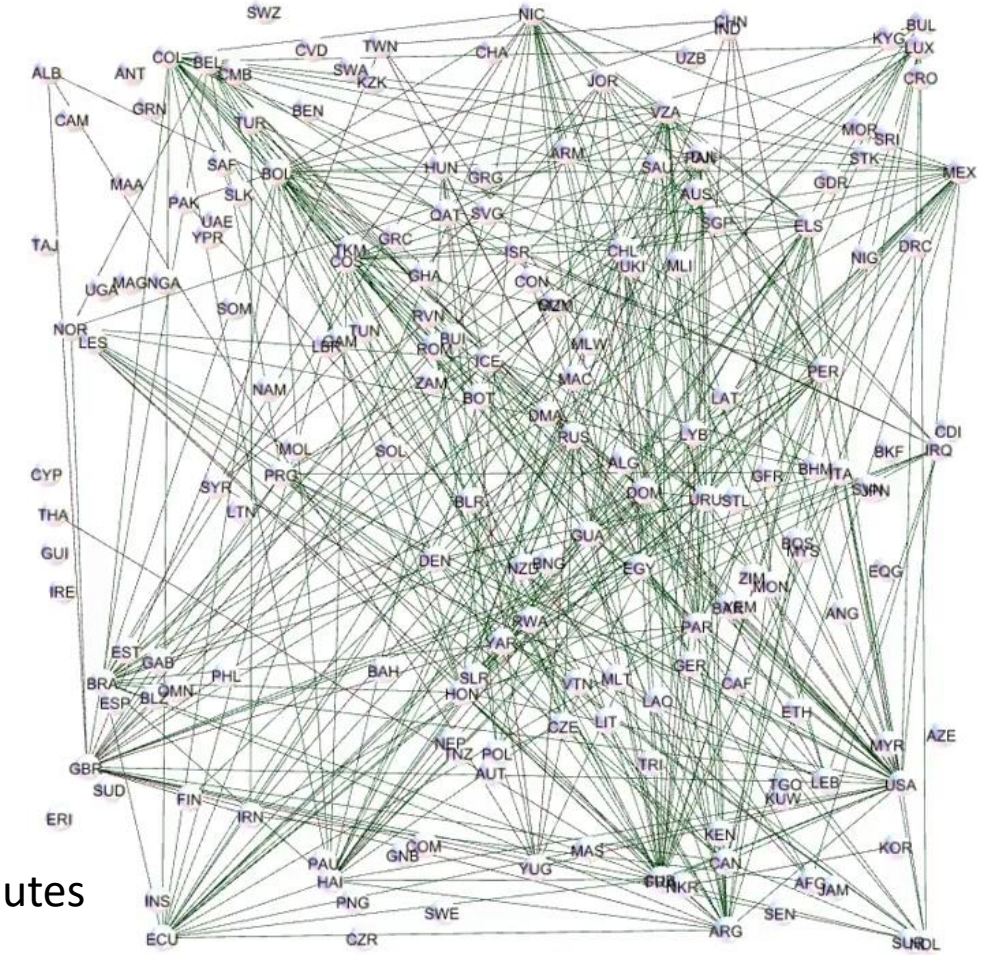


International relations

Signed international relations Correlates of War dataset (1946-1996)

YouTube link: TinyURL.com/CoWDynamics

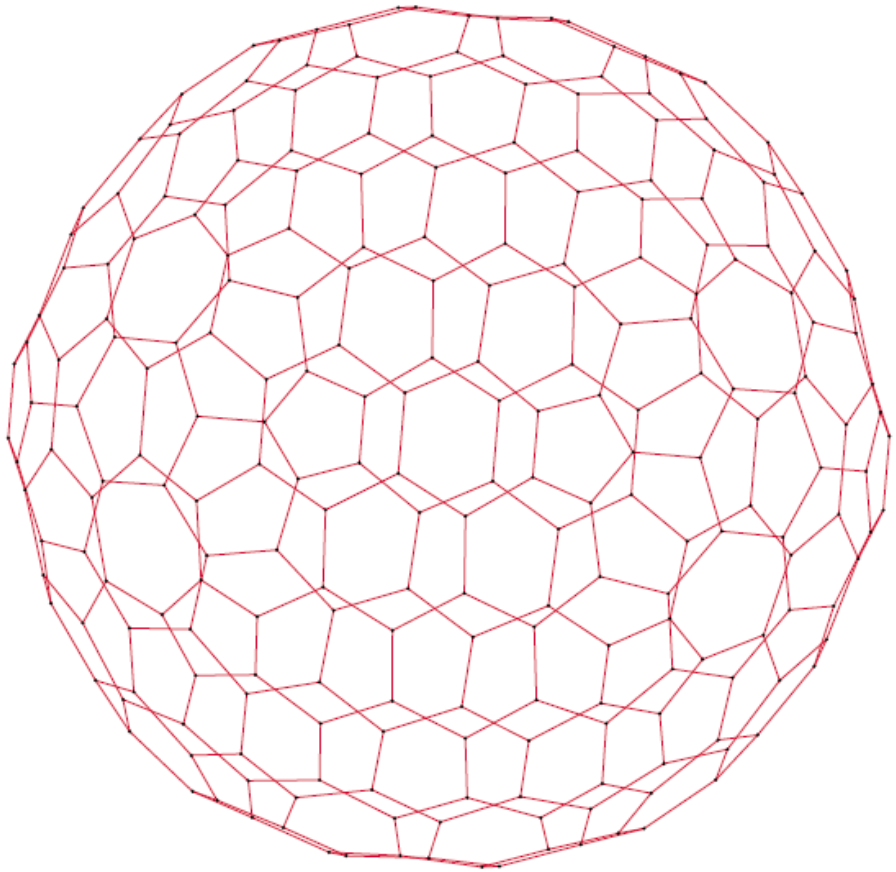


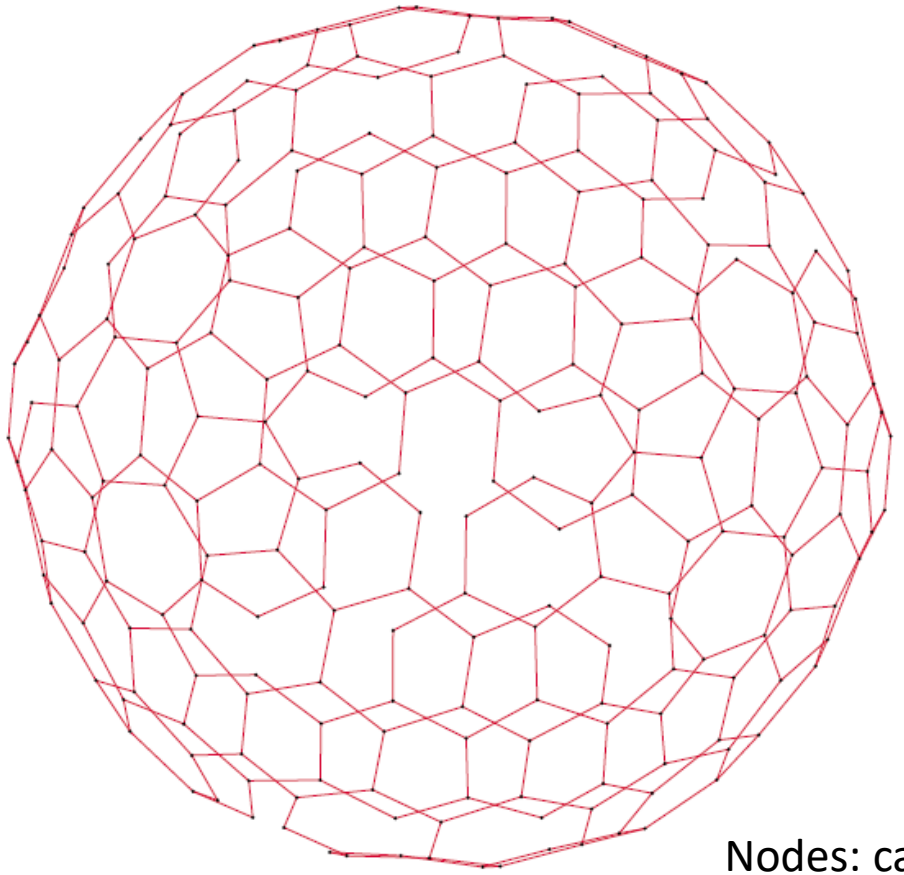


Nodes: Countries
Positive edges: Formal military alliances
Negative edges: Military conflicts, ideological conflicts, or border disputes

Bipartivity of fullerenes

Closely related problem from Chemistry





Nodes: carbon atoms
Edges: atomic bonds (considered negative)
Network: molecule

UNIVERSITY OF TORONTO

Ising spin glass models with ±1 interactions

Ground-state spin configuration and energy (H)

Nodes: spins
Edges: aligned or unaligned couplings


Journal of Complex Networks

Issues Advance articles Submit ▾ Purchase Alerts About ▾

Article Contents

Abstract

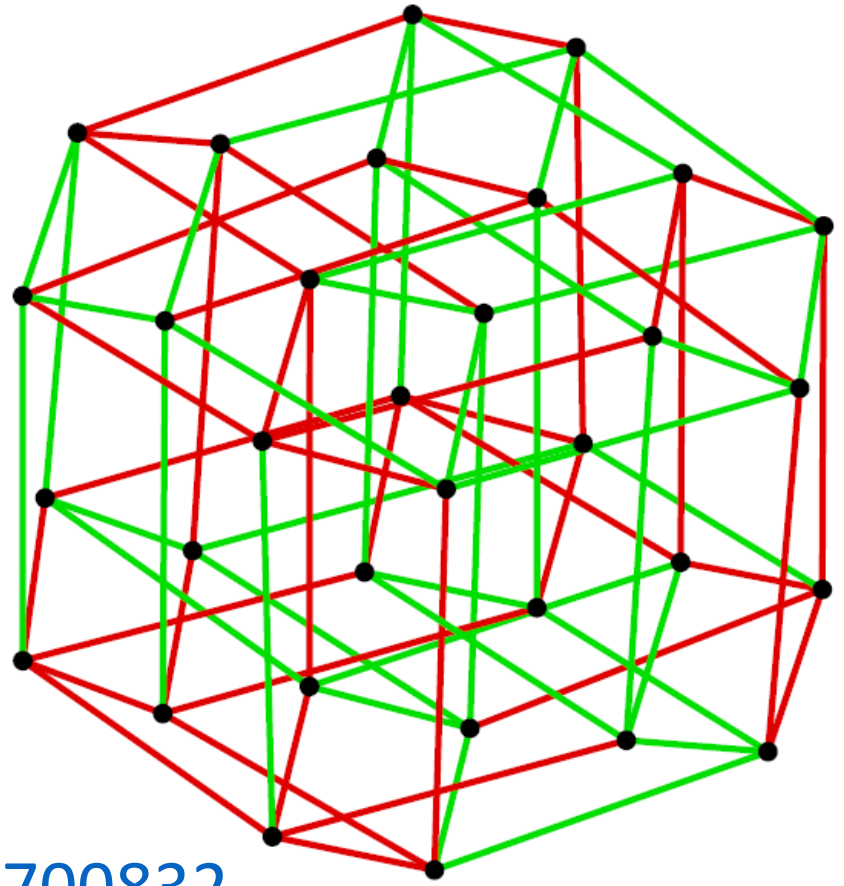
1. Introduction

Balance and frustration in signed networks

Samin Aref ✉, Mark C Wilson

Journal of Complex Networks, cny015, https://doi.org/10.1093/comnet/cny015

Published: 14 August 2018 **Article history** ▾

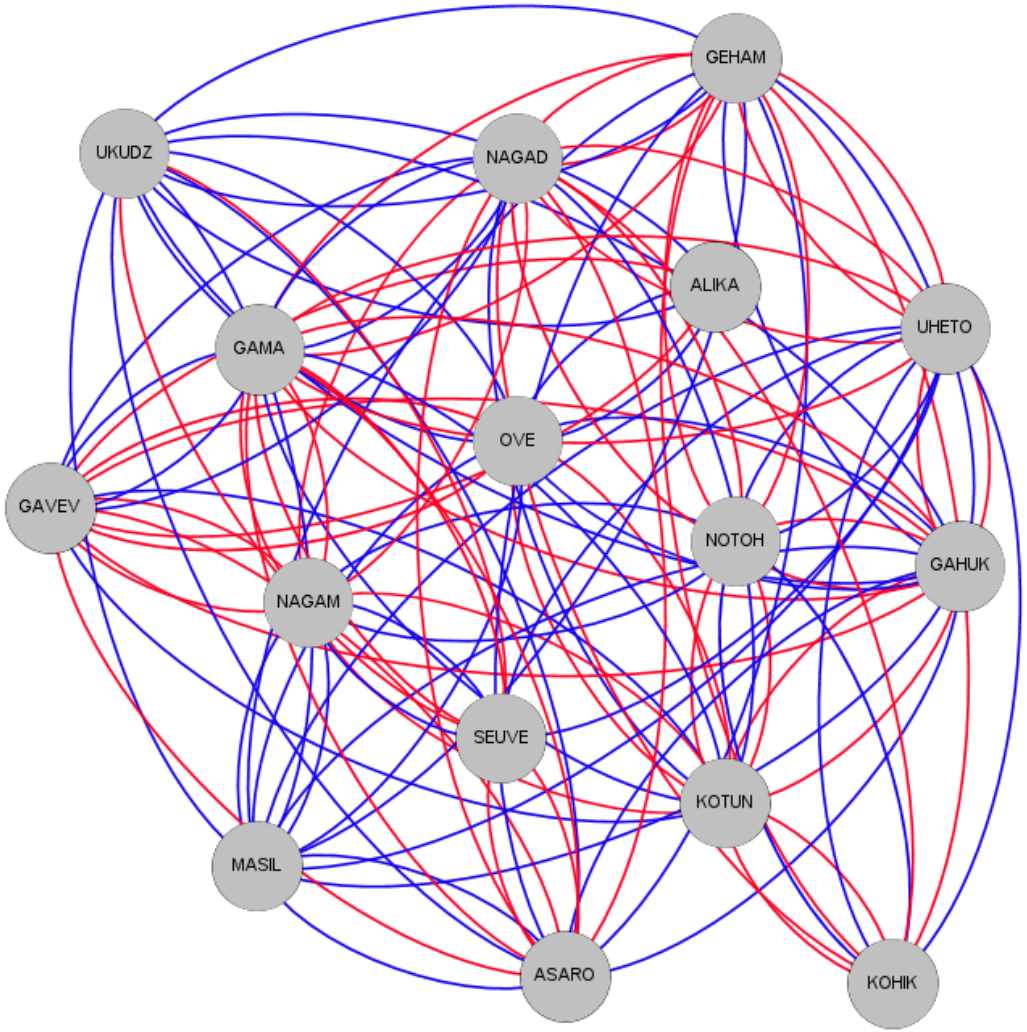


Data: https://doi.org/10.6084/m9.figshare.5700832

Paper 5: Multi-level assessment of balance

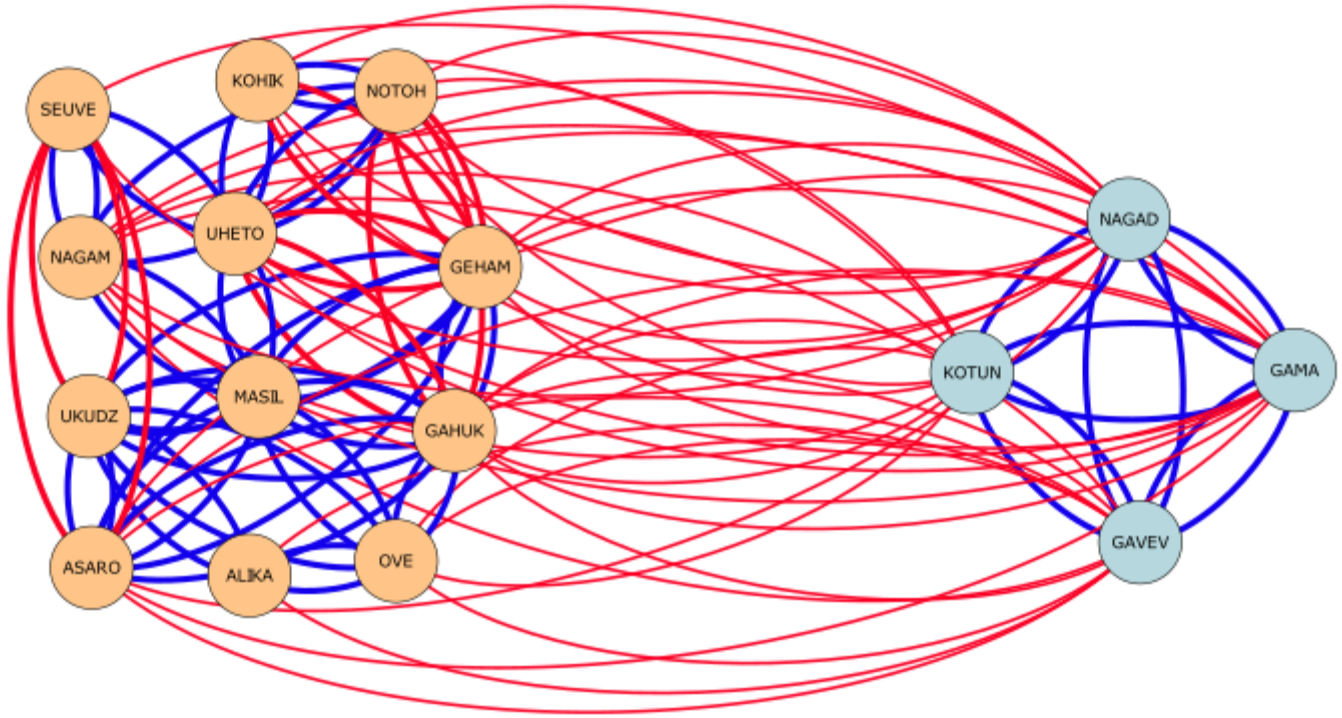
New Guinean Tribes

Direction of edges: clockwise curvature
Negative edges are shown in red



Nodes: Tribes
Edges: Alliance and enmity between tribes

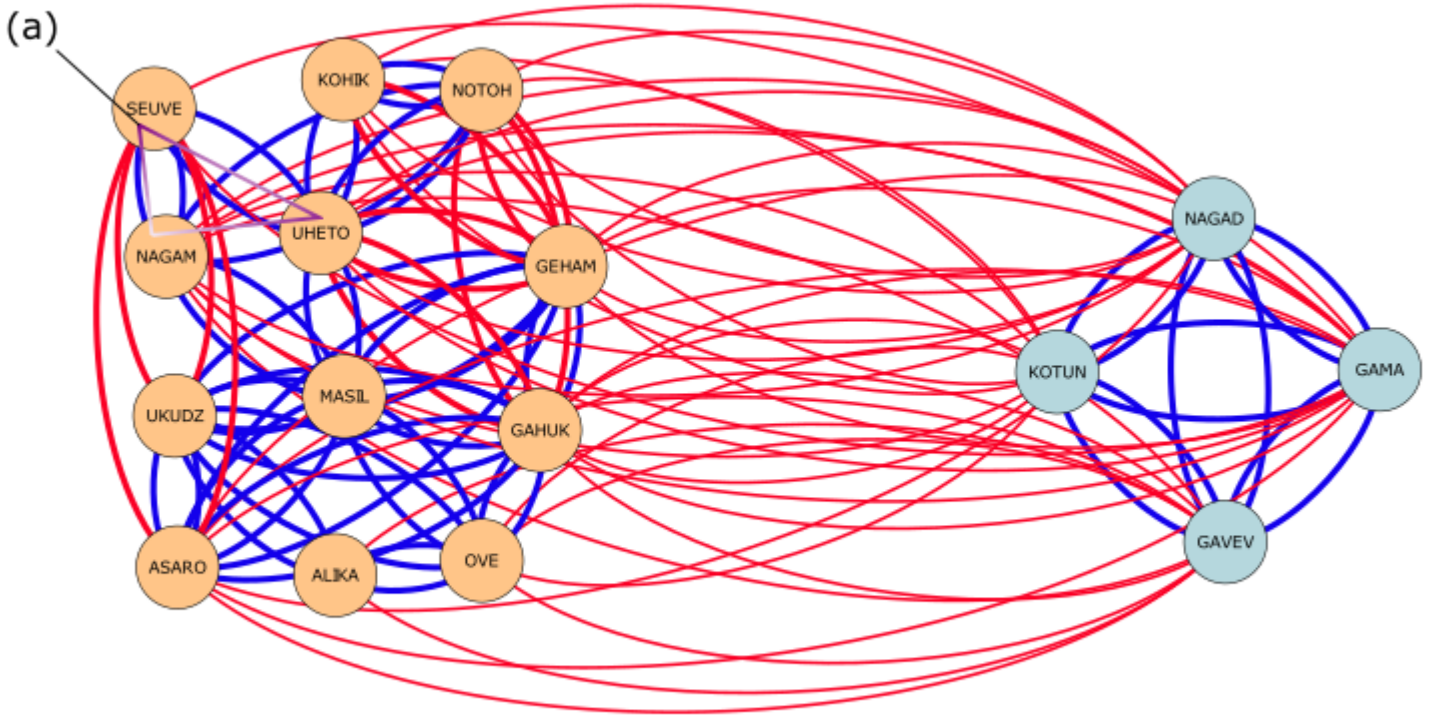
Optimal partition for a directed signed network



Nodes: Tribes

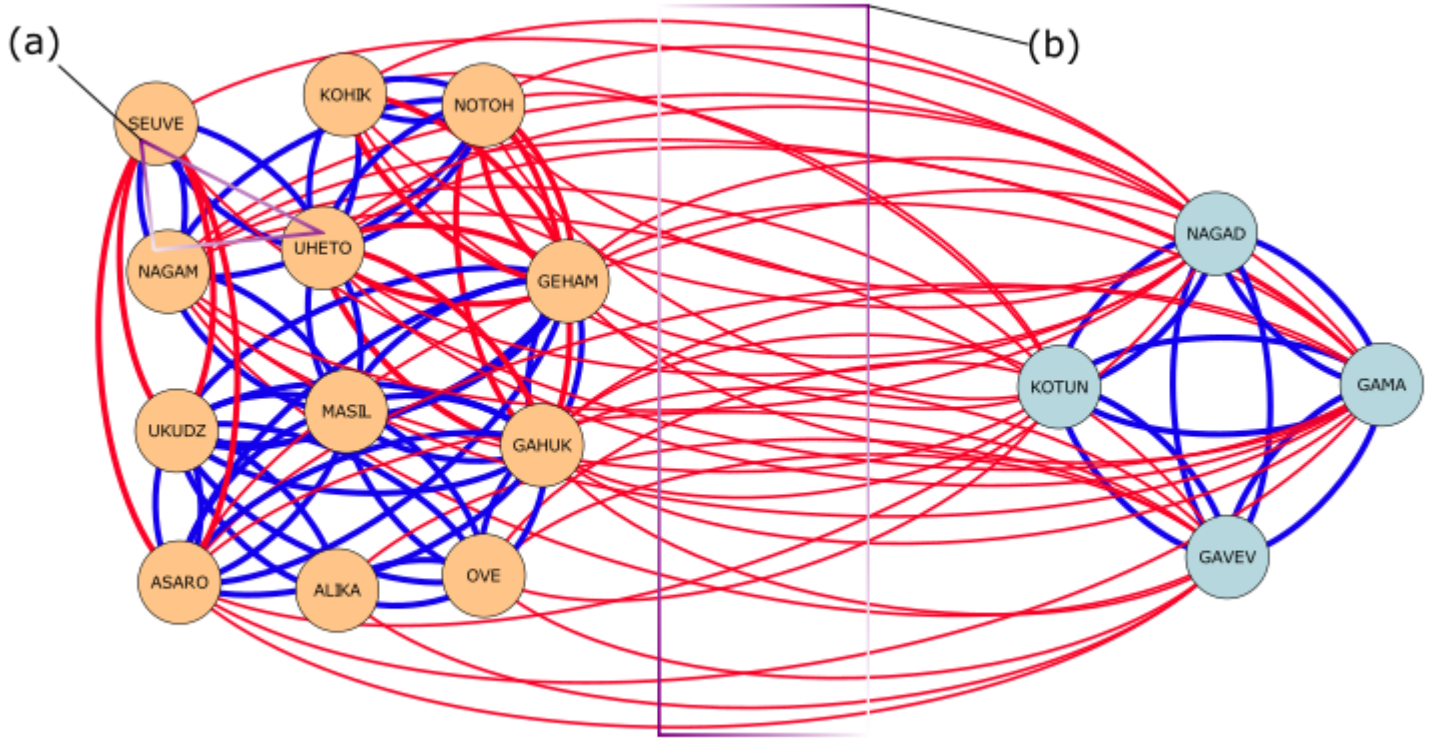
Edges: Alliance and enmity between tribes

a. Micro-level balance (triadic balance)



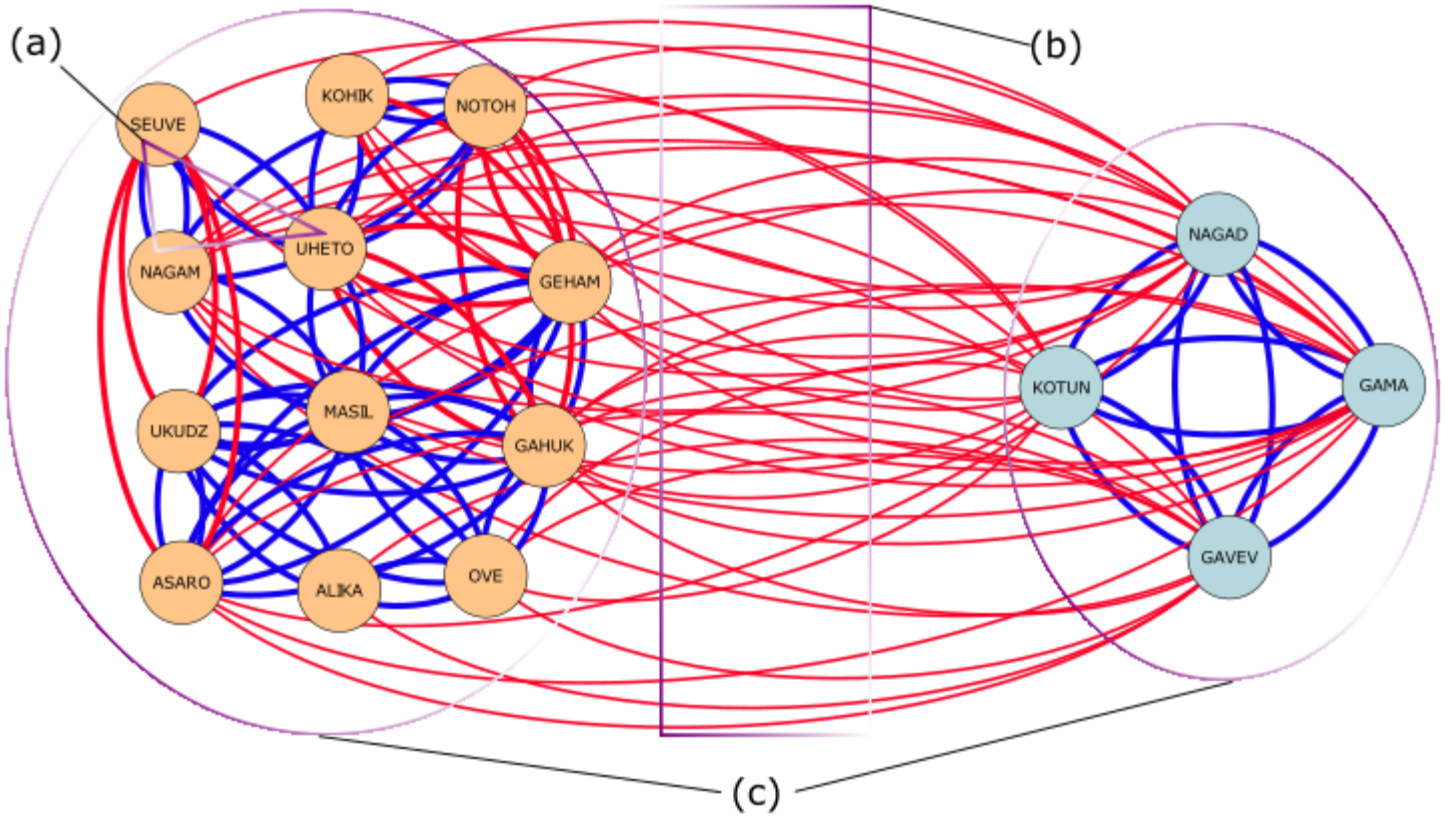
UNIVERSITY OF TORONTO

b. Meso-level balance (divisiveness)



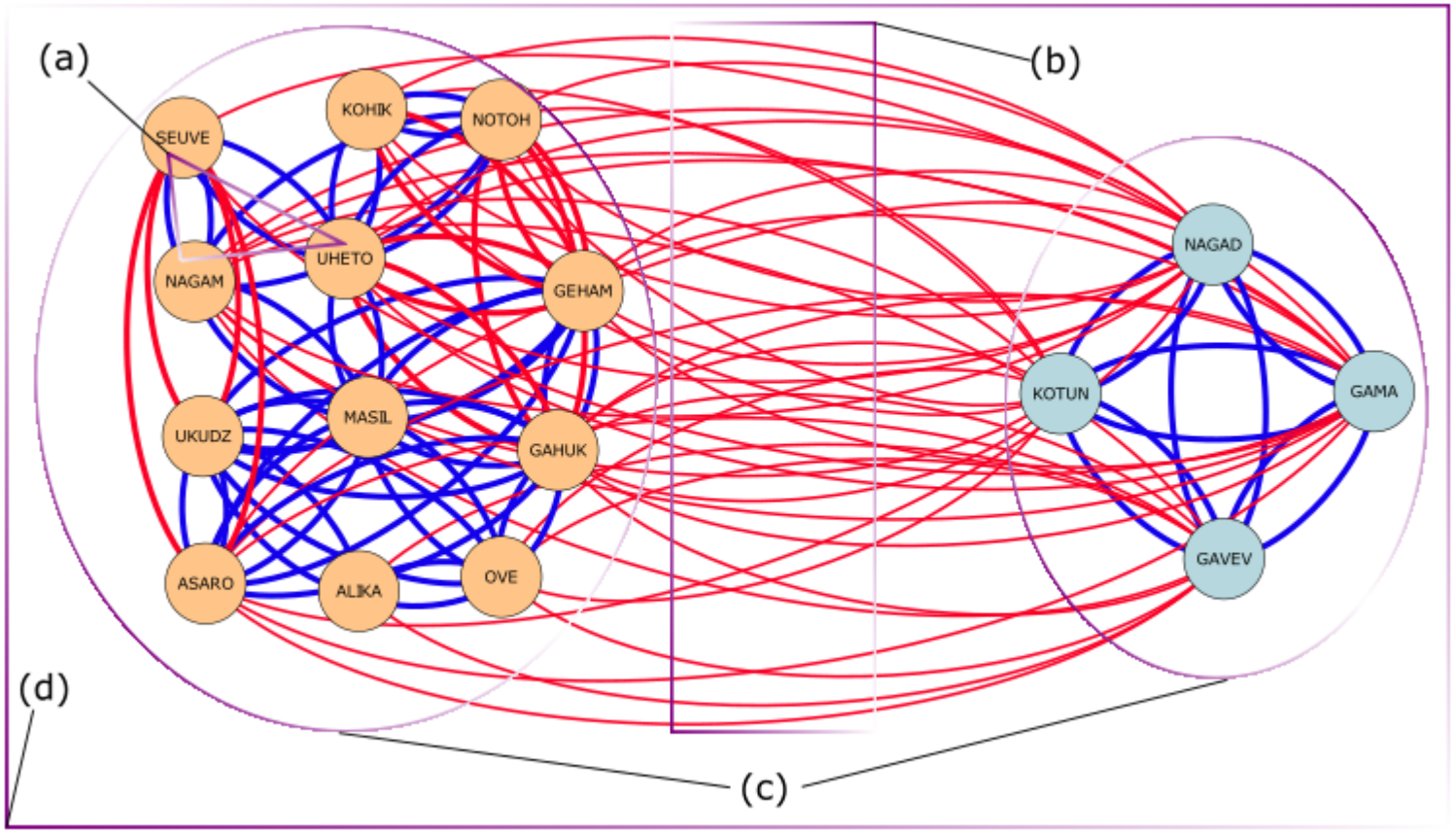
I suggest using the revised version of cohesiveness proposed by Emma Fraxanet et al. which has the advantage of controlling for the differences in fraction of negative edges.

c. Meso-level balance (cohesiveness)



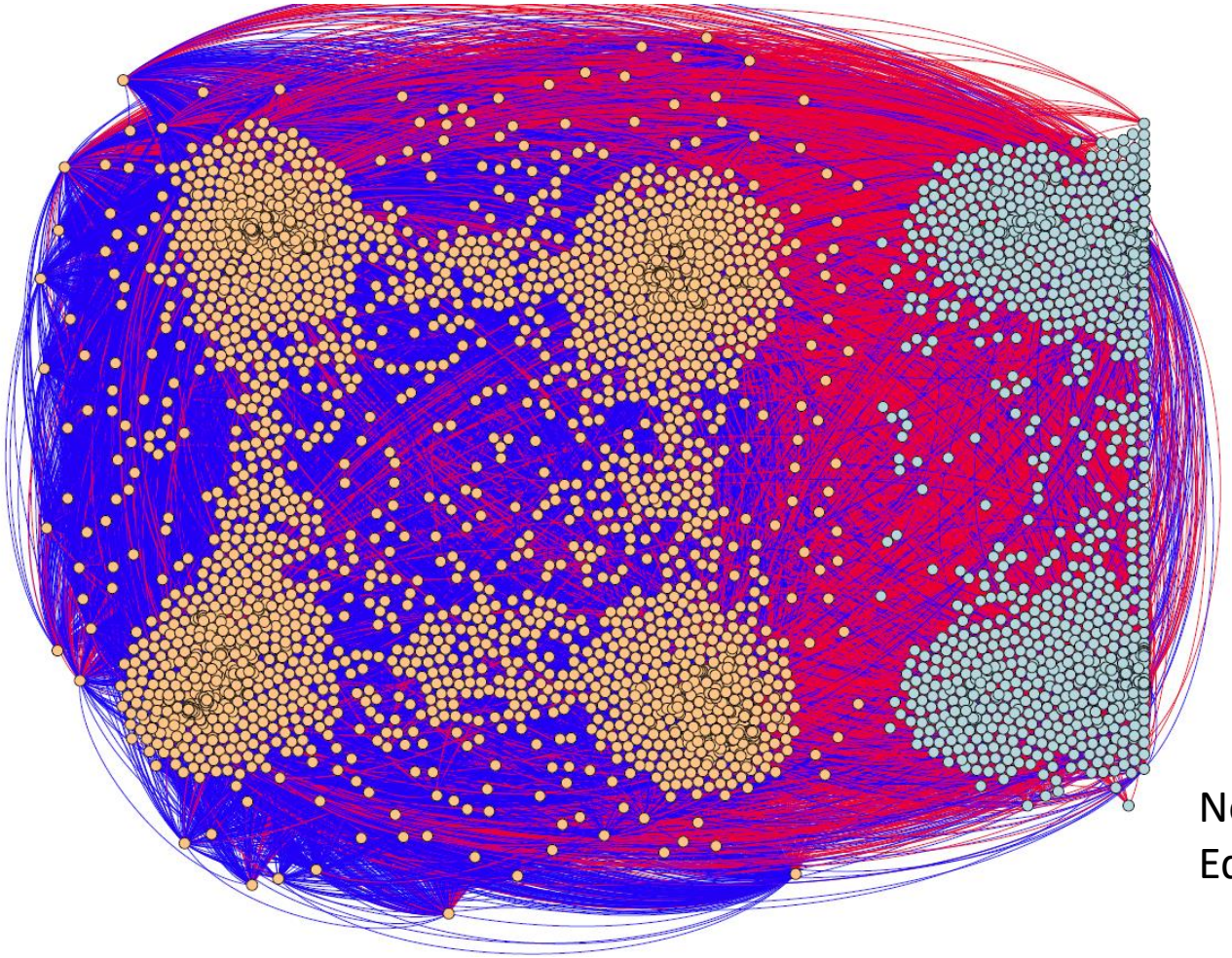
I suggest using the revised version of cohesiveness proposed by Emma Fraxanet et al. which has the advantage of controlling for the differences in fraction of negative edges.

d. Macro-level balance (network balance)



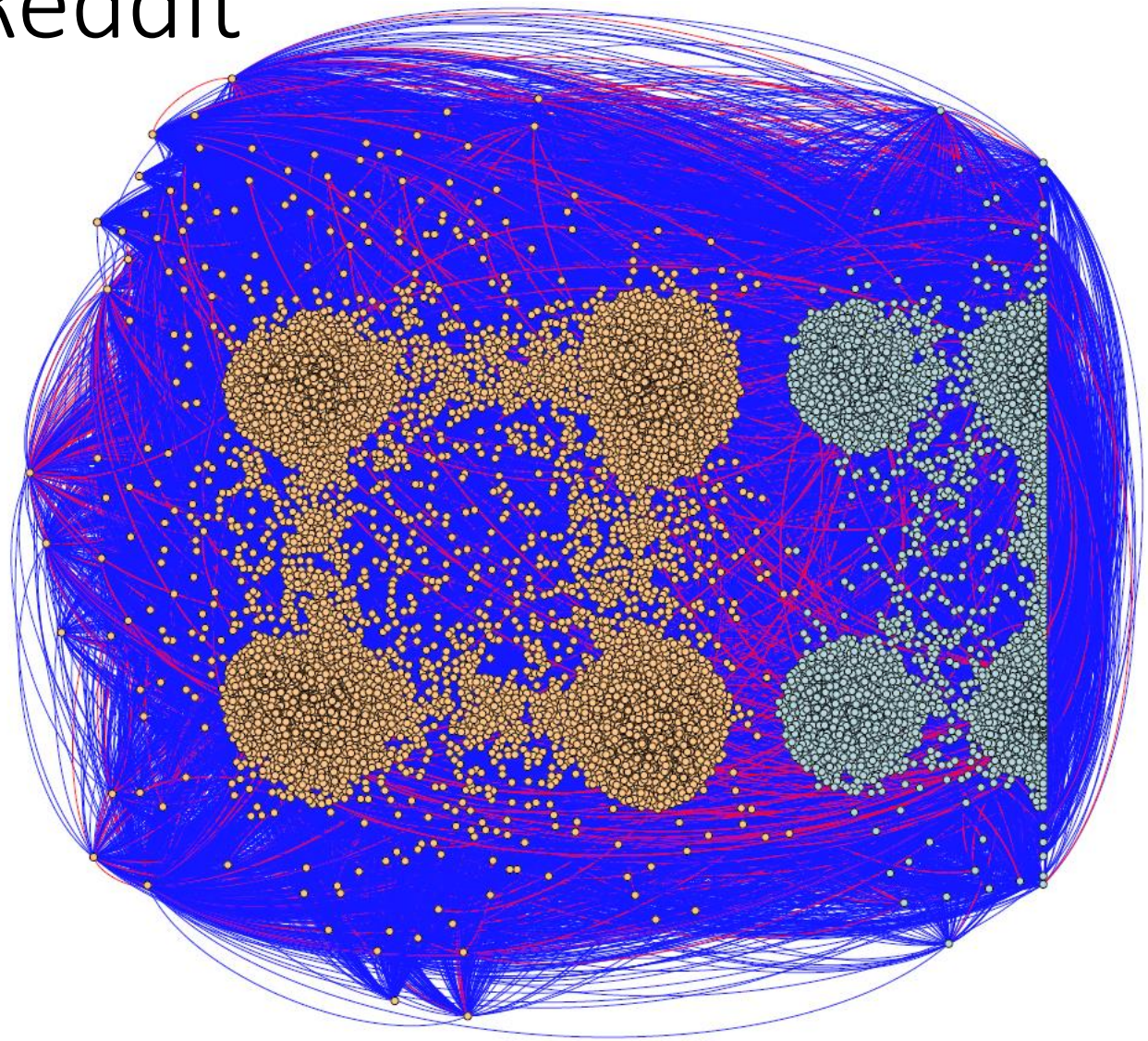
UNIVERSITY OF TORONTO

Wikipedia Elections



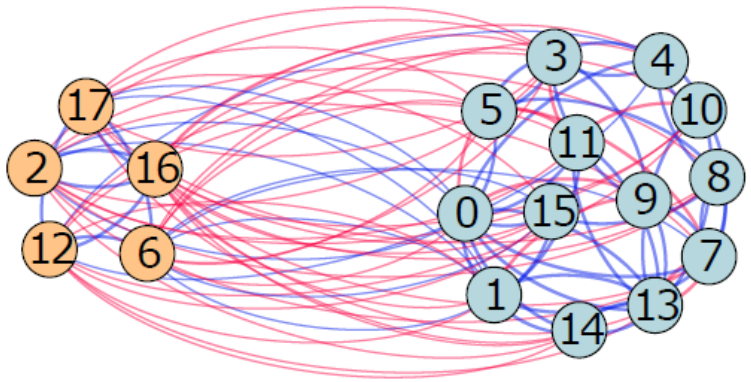
Nodes: Wikipedia editors
Edges: Approval and disapproval votes (for promotion)

Reddit

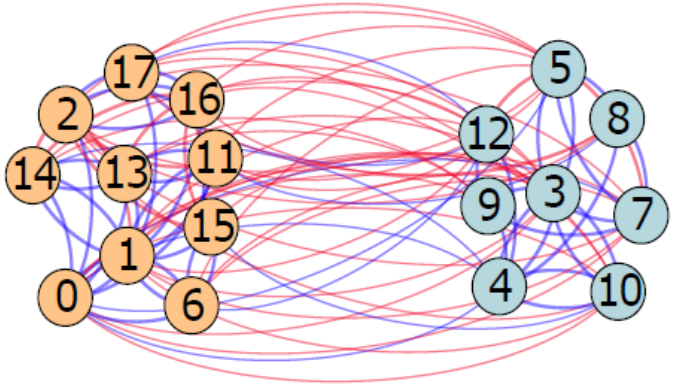


Nodes: Reddit users (from two subreddits)
Edges: Sentiments of the content exchanged

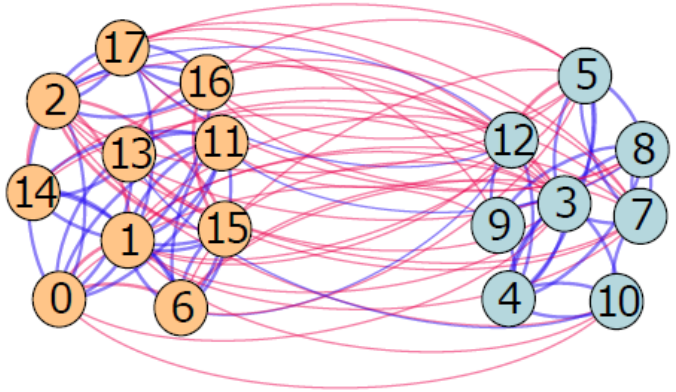
Temporal network (Sampson's monastery)



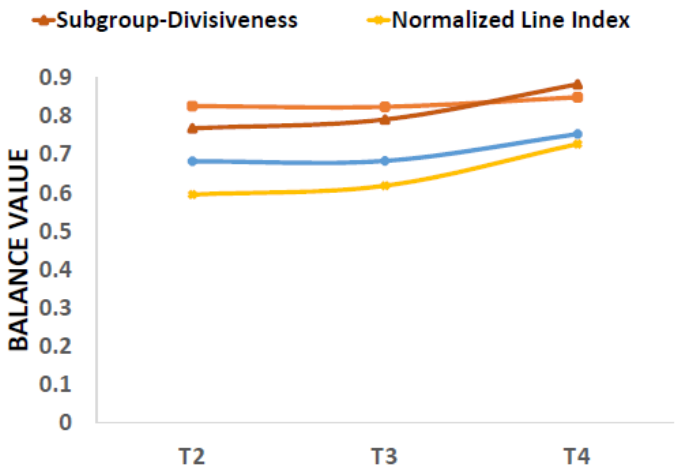
Time=T2

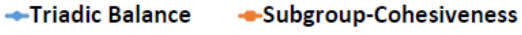


Time=T3



Time=T4



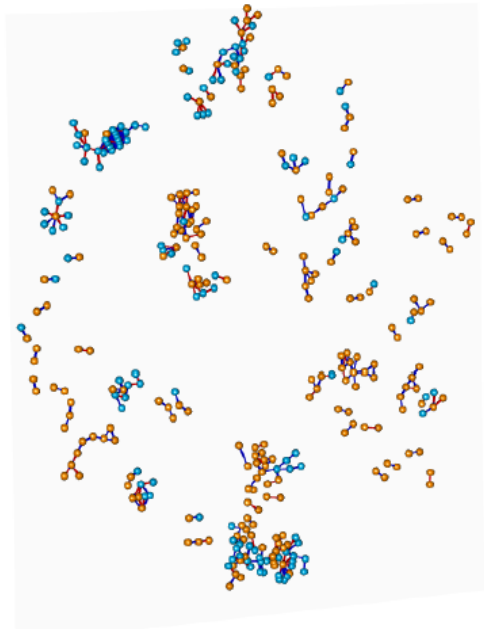


Nodes: people (monks)
Edges: top choices for like and dislike

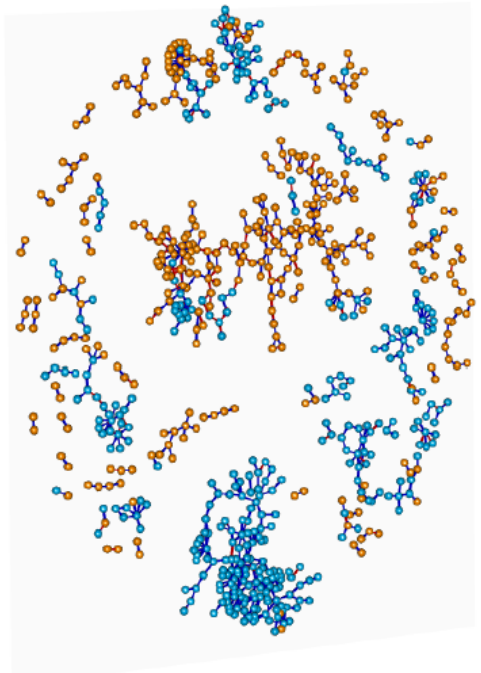
See the alternative/complementary interpretations about the disintegration of this network by Sofia Teixeira et al. in this paper
Talaga, S., Stella, M., Swanson, T.J. et al. Polarization and multiscale structural balance in signed networks. Commun Phys 6, 349 (2023).

UNIVERSITY OF TORONTO

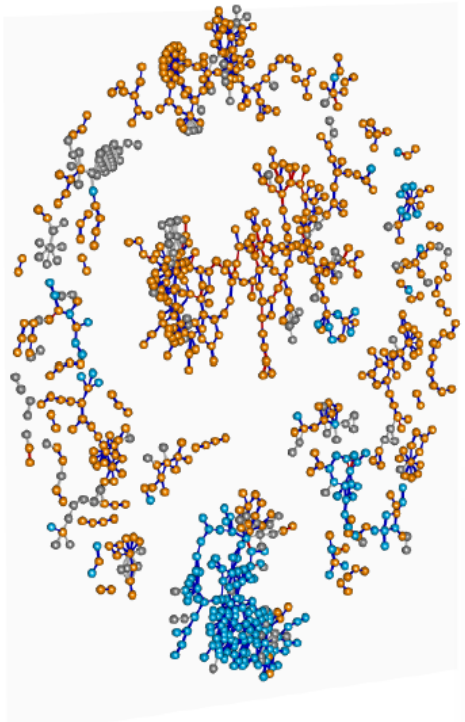
Multilayer network (Collin's philosophers)



Acquaintance



Master_Pupil



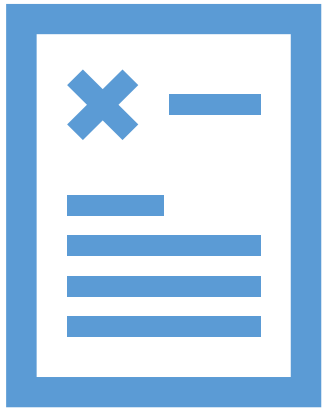
Flattened

Nodes: philosophers

Edges: positive and negative ties (based on close readings of historical texts)

UNIVERSITY OF TORONTO

Paper 6: Networks of US Congress legislators
20 signed networks: one for each session (2-year period) of the House of Representative



Bill co-sponsorship data

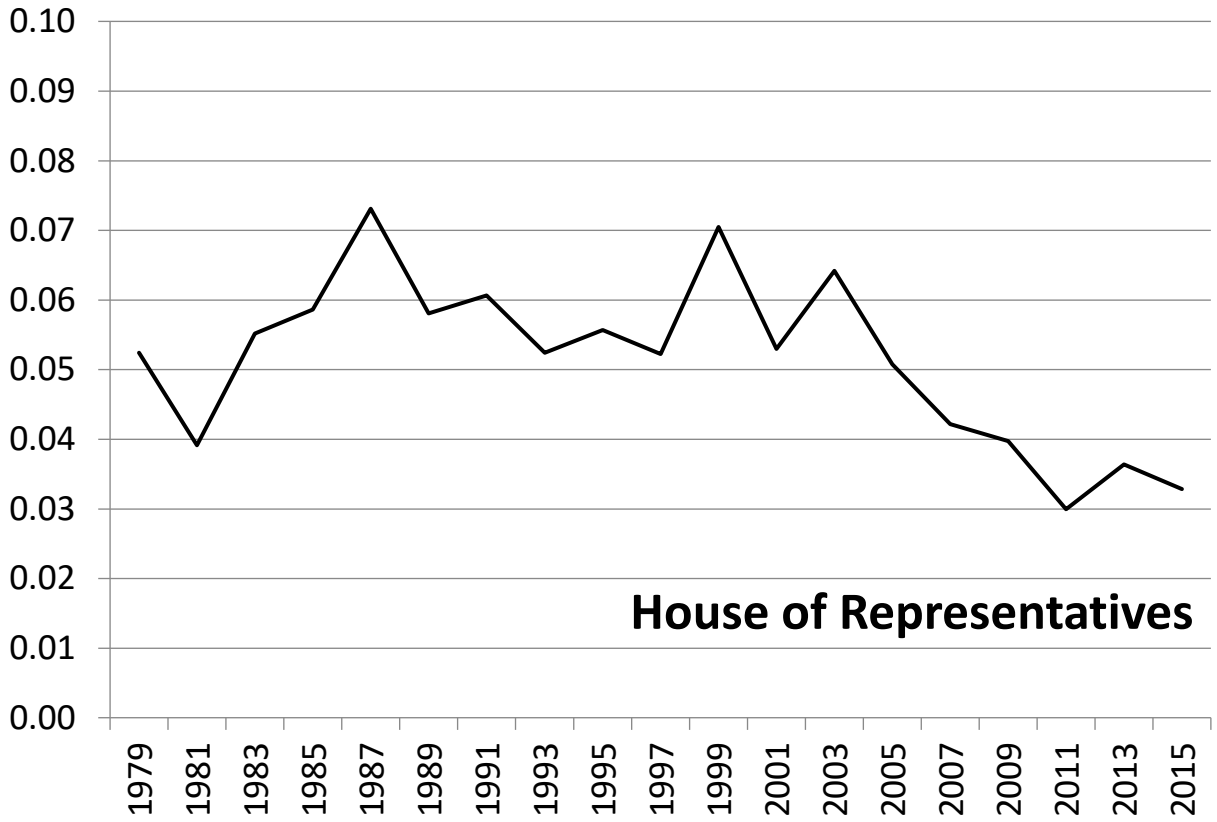
who has co-sponsored which bill

20 sessions: 1979-2017

~435 US Representatives ×
~6000 bills (co)sponsored

UNIVERSITY OF TORONTO

Bill passage rate in the US House of Representatives



Research Question:

Why is Congress so ineffective at passing legislation?

Two possible explanations

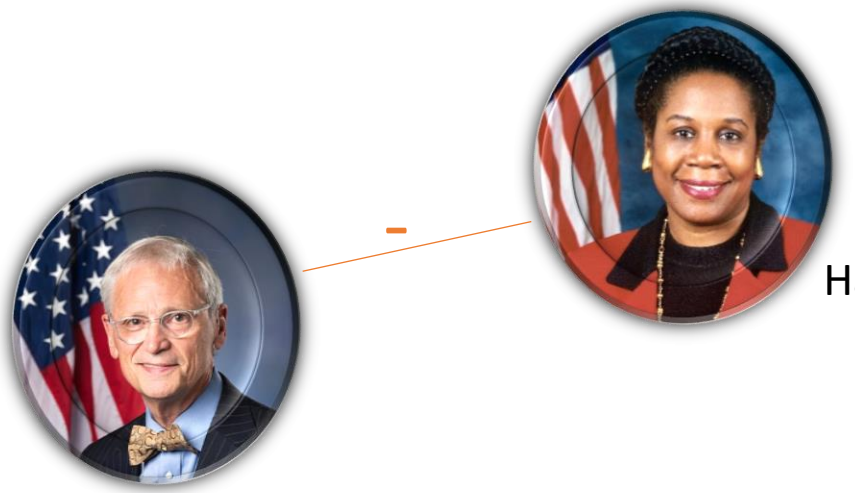
Party Control Hypothesis: It's easier to pass bills when the majority party holds a larger majority.

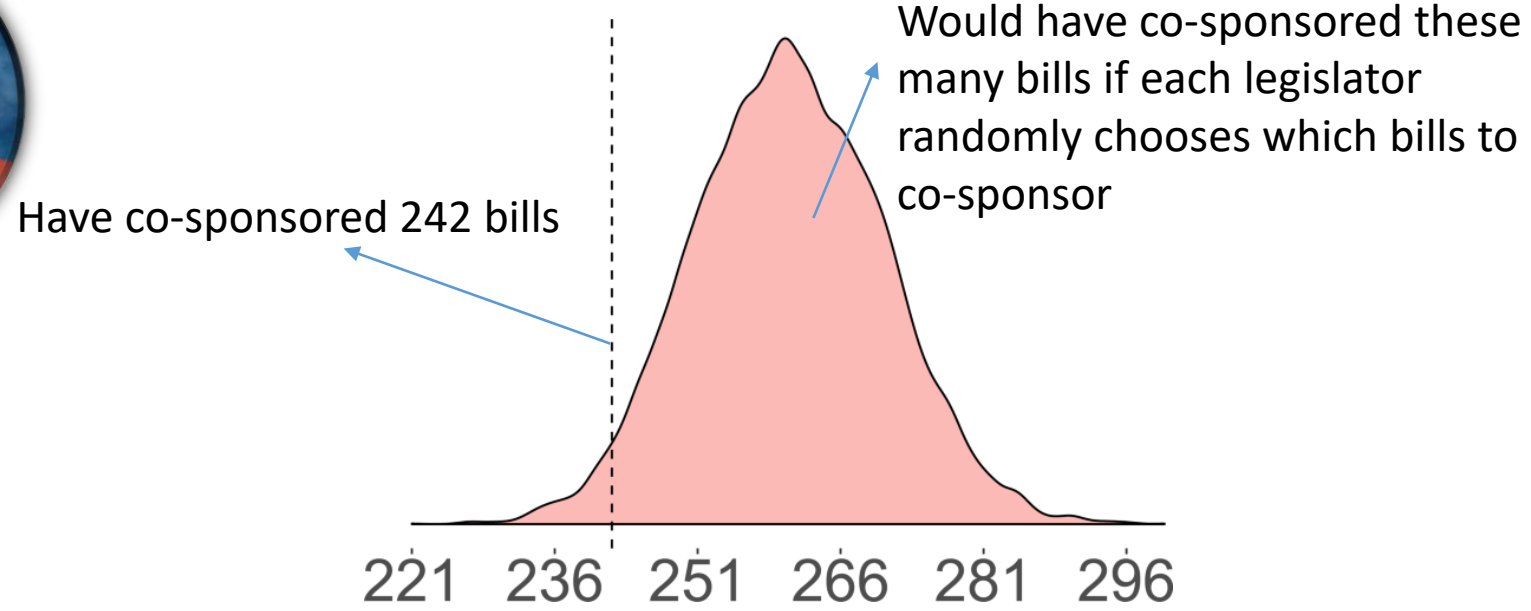
Coalition Partisanship Hypothesis: Coalitions matter. When Congress is polarized[1] into opposing coalitions, it's easier to pass bills when the largest coalition is more ideologically unified (i.e., partisan).

1. (political) polarization: the formation of non-overlapping ideologically homogeneous groups (Layman, Carsey, and Horowitz , 2006).

UNIVERSITY OF TORONTO

Inferring signed networks of US legislators





The fi-method proposed by Georges Andres at al. offers a similar way of inferring weighted signed graphs from interaction data. See Andres, G.; Casiraghi, G.; Vaccario, G.; Schweitzer, F. (2023). Reconstructing signed relations from interaction data. Scientific Reports 13, 20689.

US Senate network 1989-1991 (101st session)

Node colors=party affiliation:

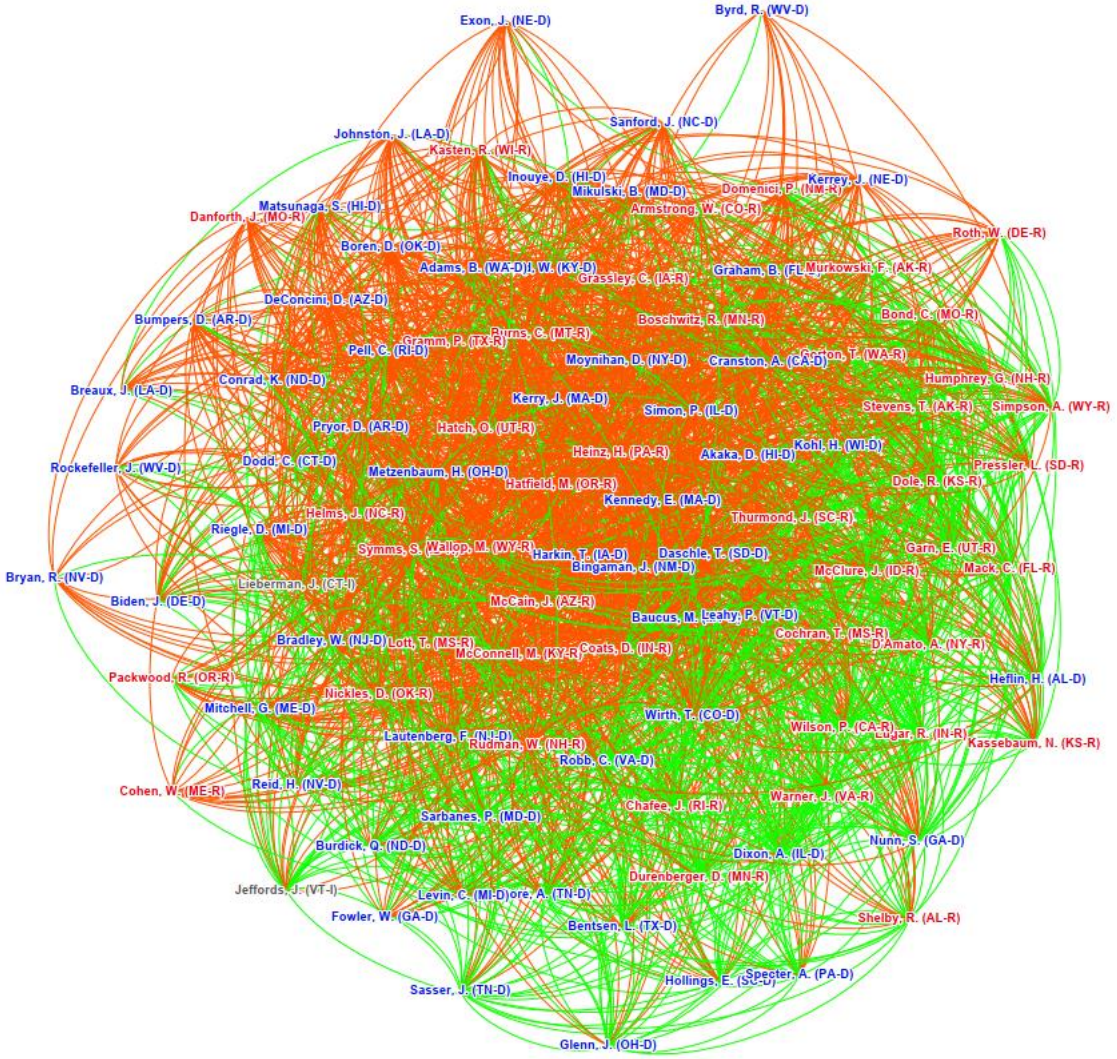
Republican

Democrat

Edge colors:

Collaboration

Lack of collaboration



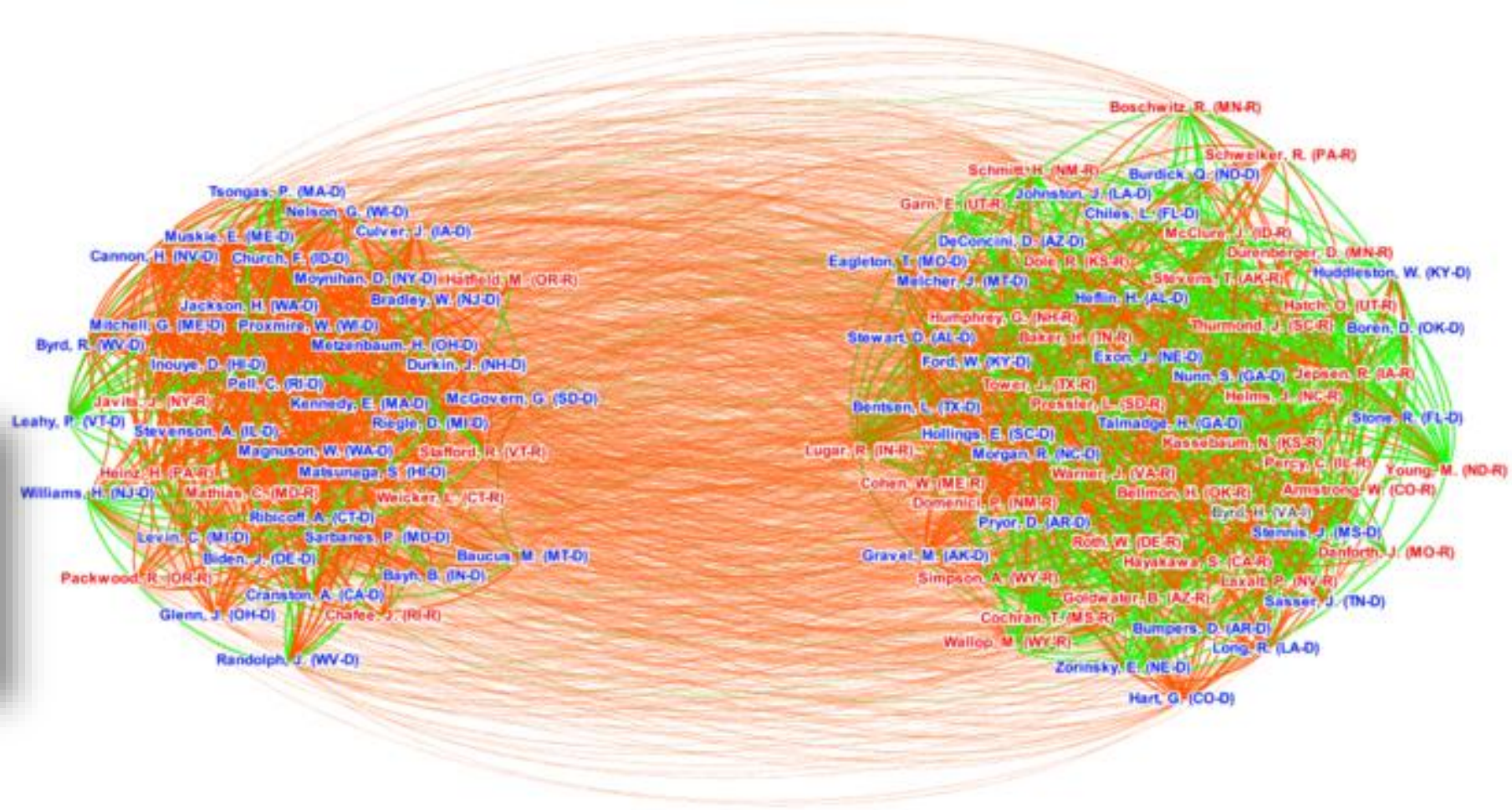
1979

Node colors=party affiliation:

- Republican
- Democrat

Edge colors:

- Collaboration
- Lack of collaboration



scientific reports

Explore our content ▾ Journal information ▾

nature › scientific reports › articles › article

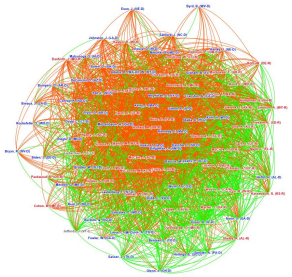
Article | Open Access | Published: 30 January 2020

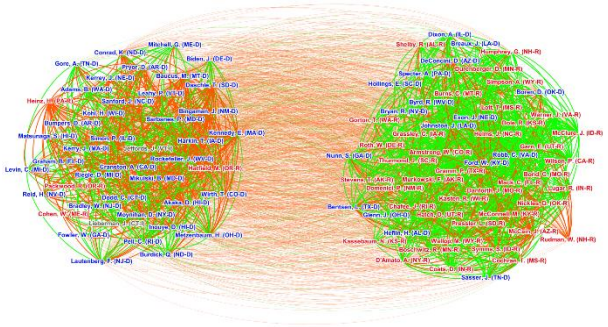
Detecting coalitions by optimally partitioning signed networks of political collaboration

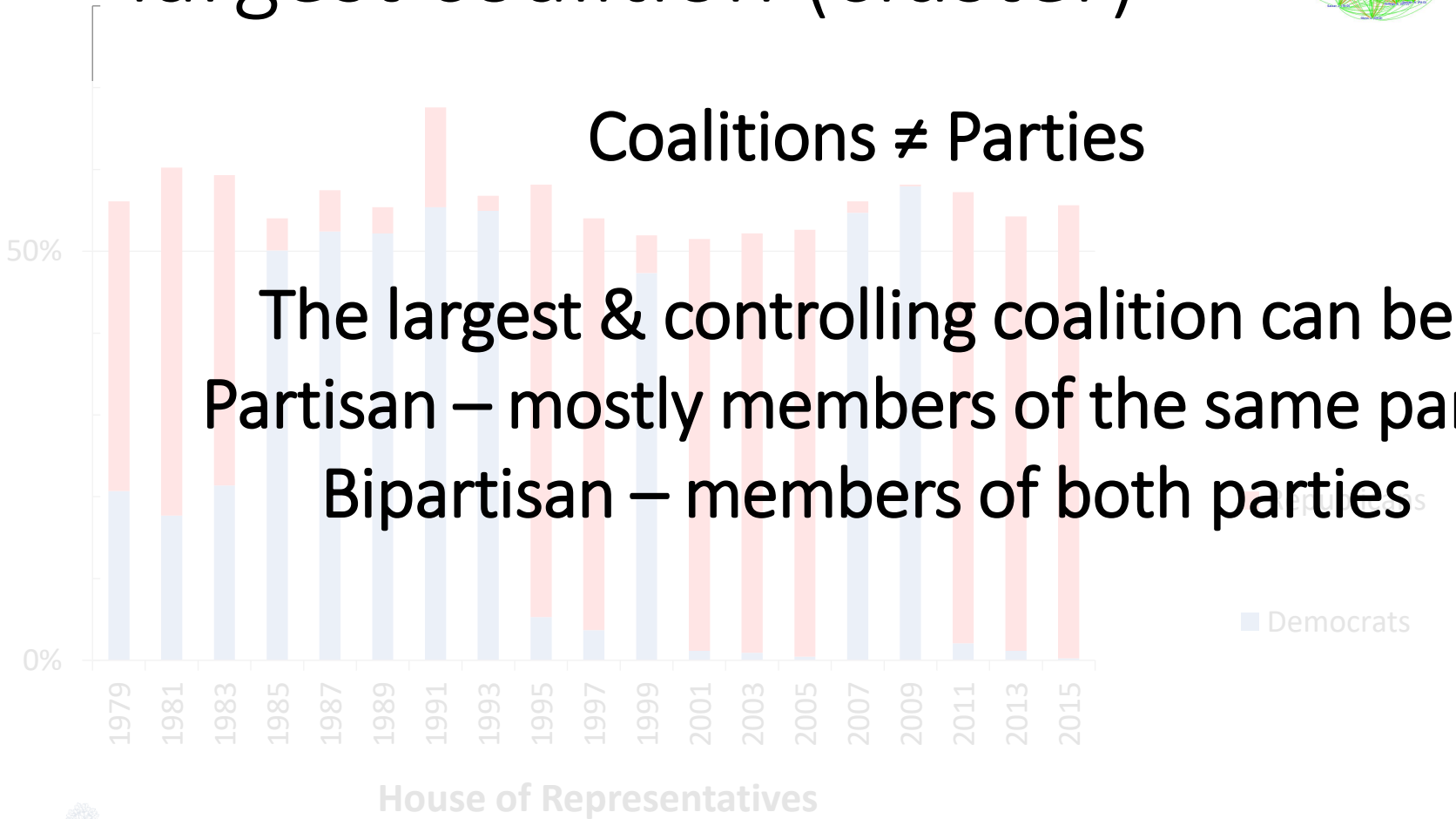
Samin Aref ✉ & Zachary Neal

UNIVERSITY OF TORONTO

Composition of the largest coalition (cluster)





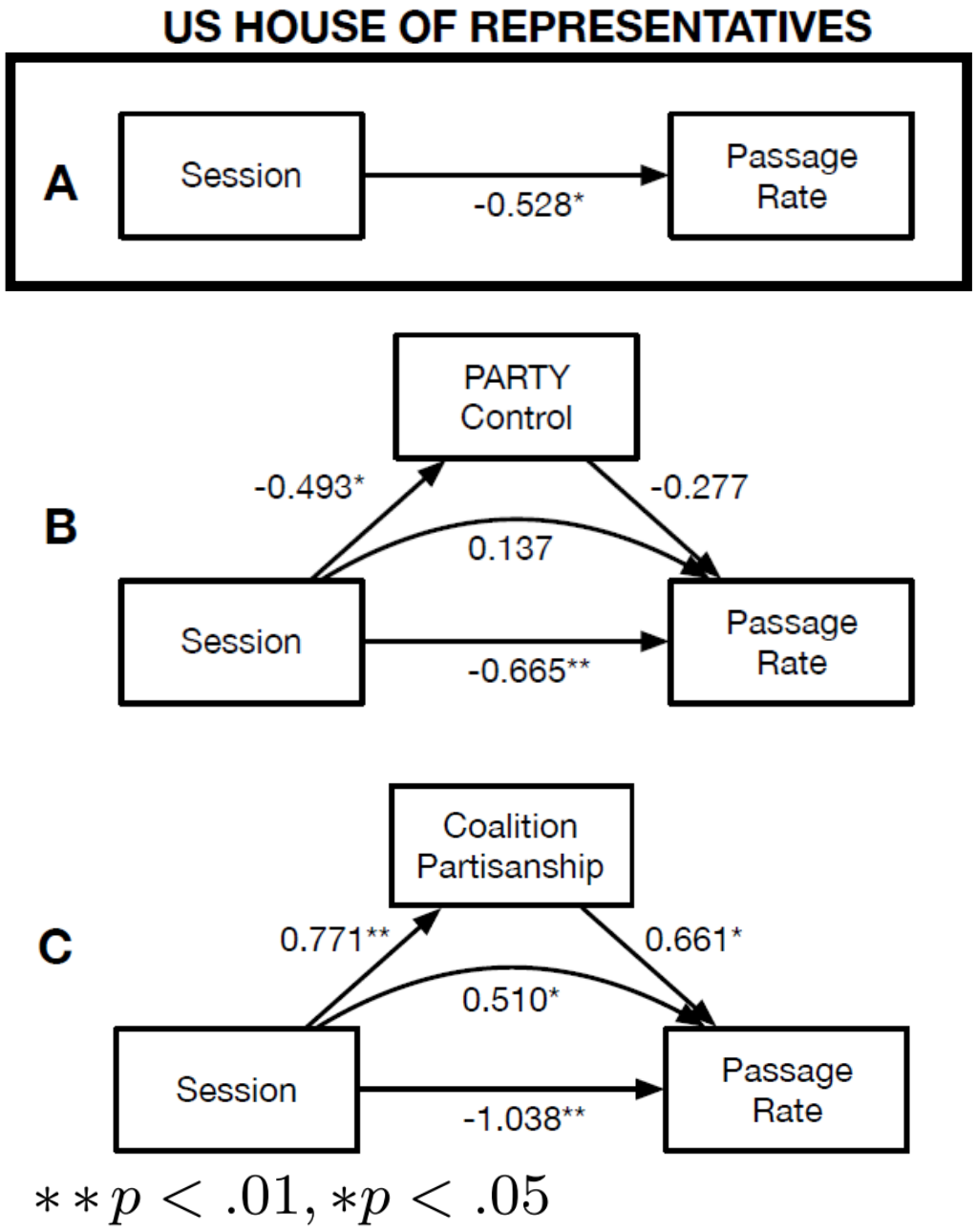


Coalitions ≠ Parties

The largest & controlling coalition can be:
Partisan – mostly members of the same party
Bipartisan – members of both parties

Findings

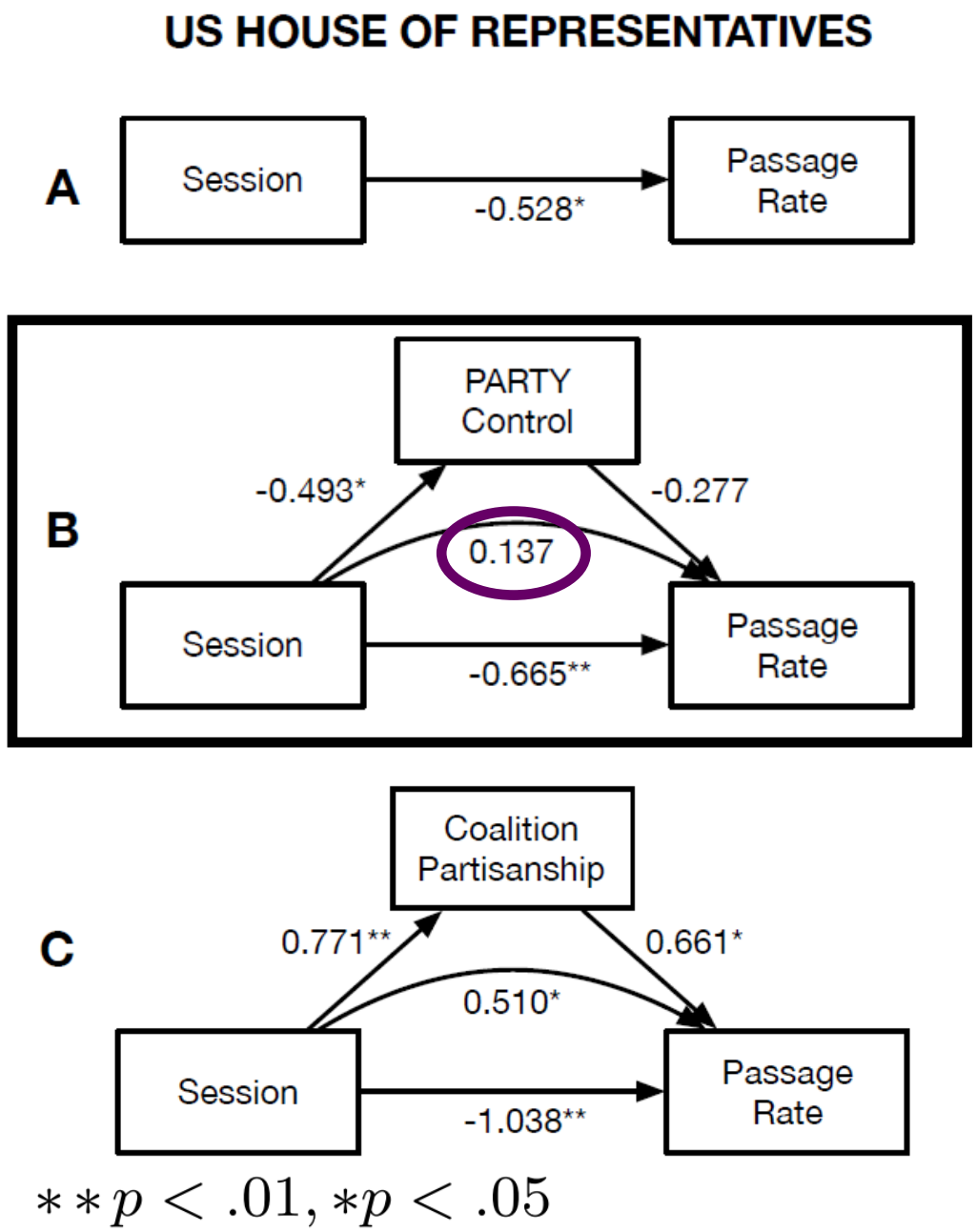
Over time, the proportion of introduced bills that pass into law (passage rate) has declined.



Findings

Over time, the proportion of introduced bills that pass into law (passage rate) has declined.

The size of the majority party's majority does not help explain why.

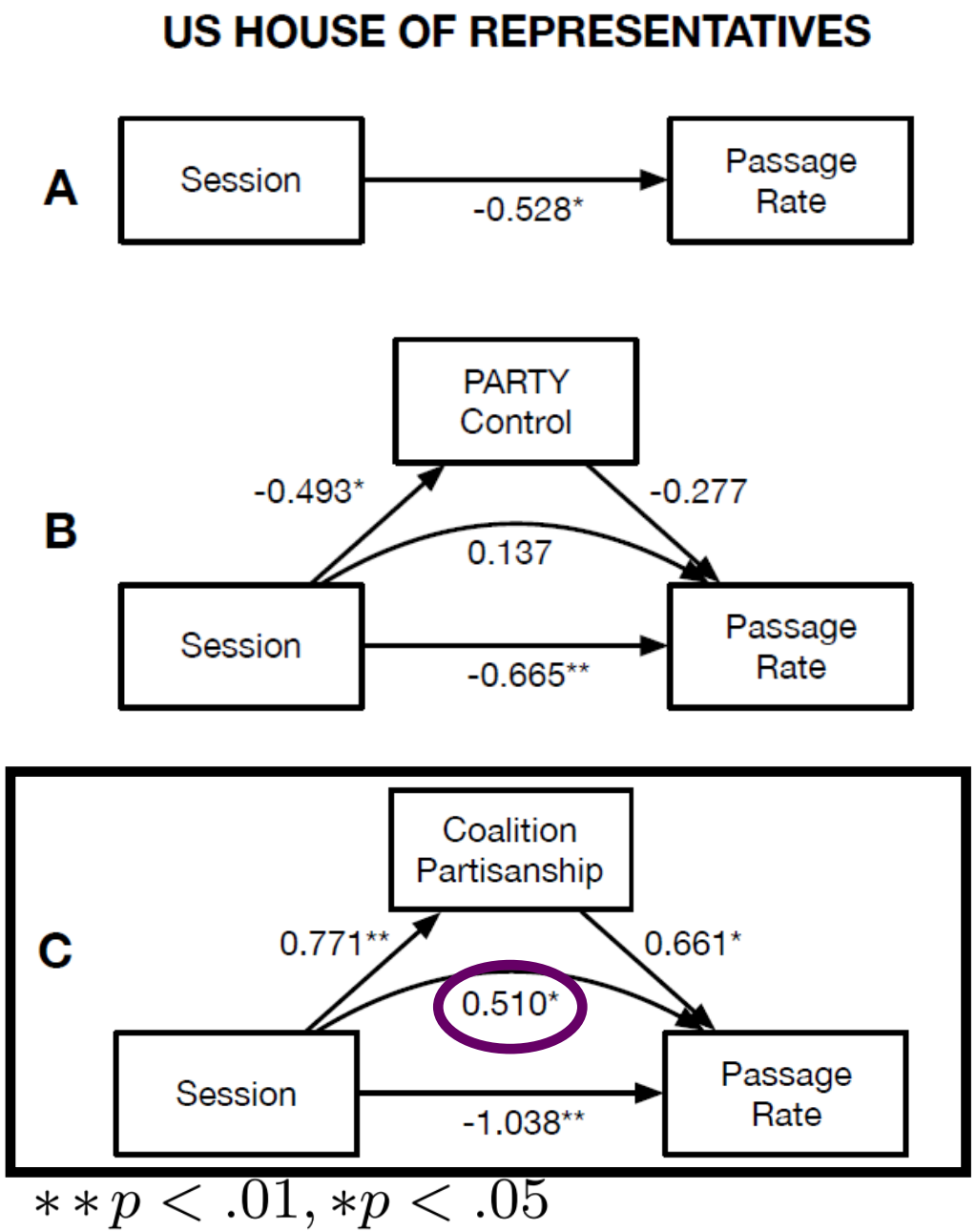


Findings

Over time, the proportion of introduced bills that pass into law (passage rate) has declined.

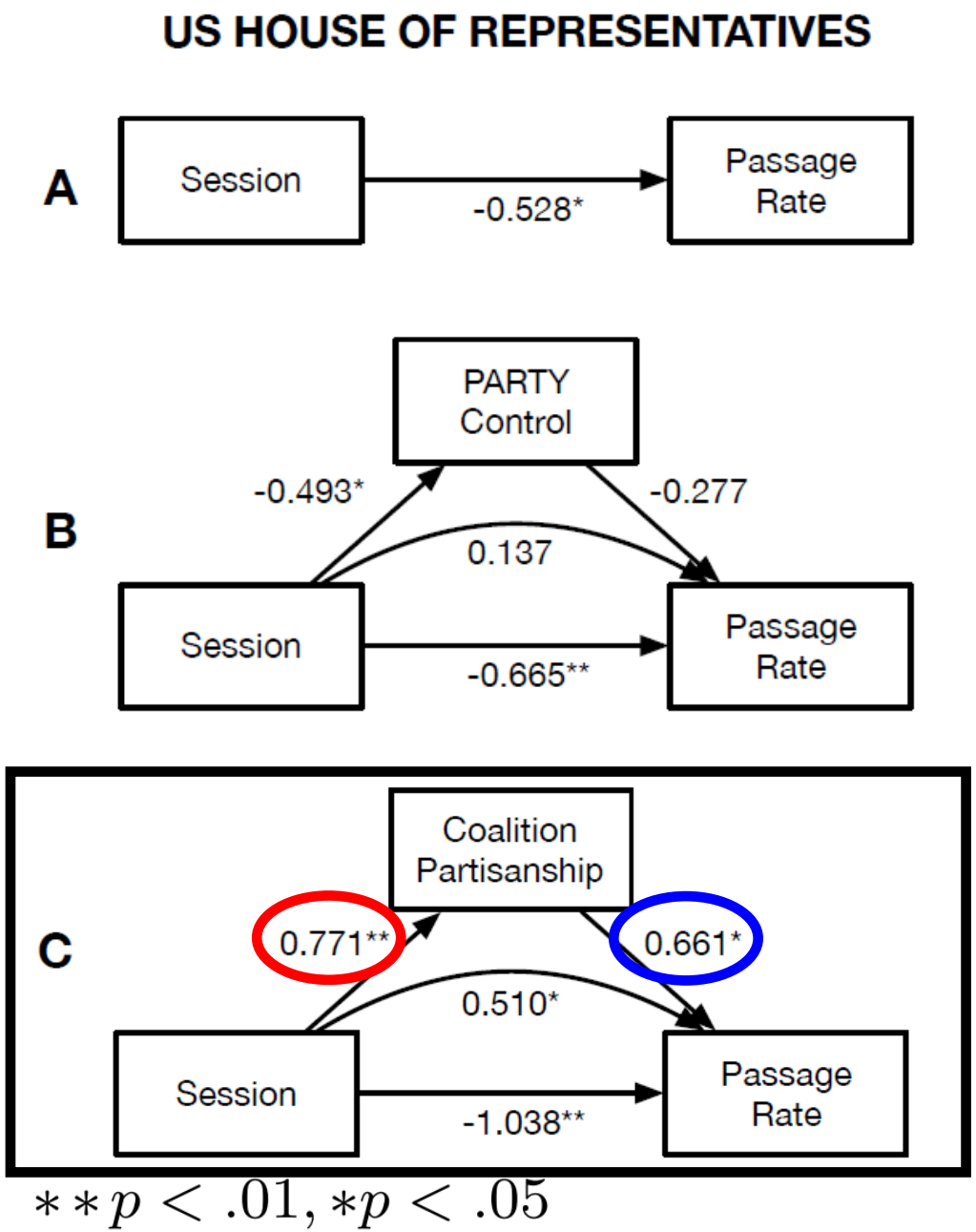
The size of the majority party's majority does not help explain why.

But, the partisanship of the largest coalition does.



Findings

Over time, the largest coalition has become more partisan, which makes it easier to pass bills.

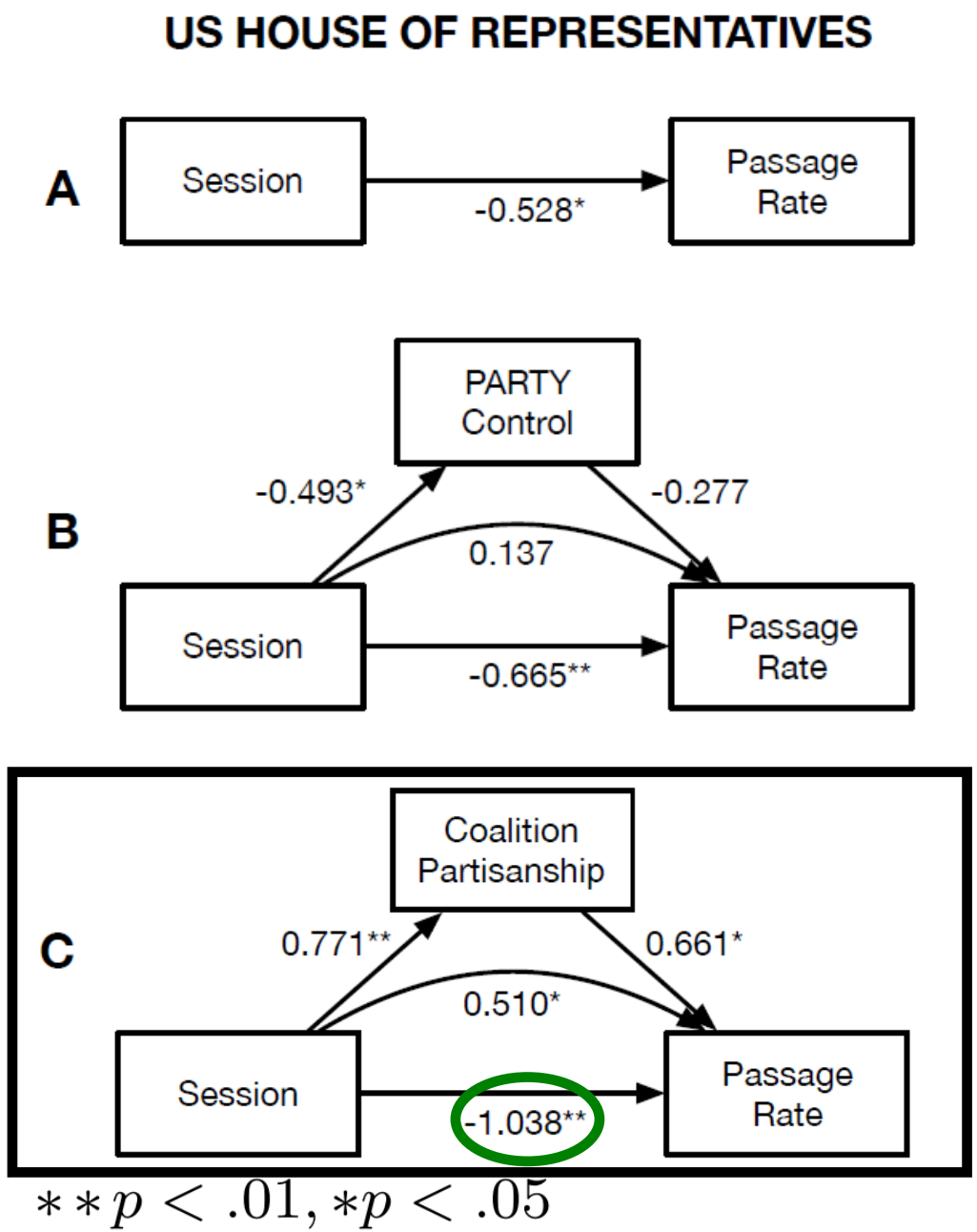


Findings

Over time, the largest coalition has become more partisan, which makes it easier to pass bills.

<u>The Silver Lining</u>

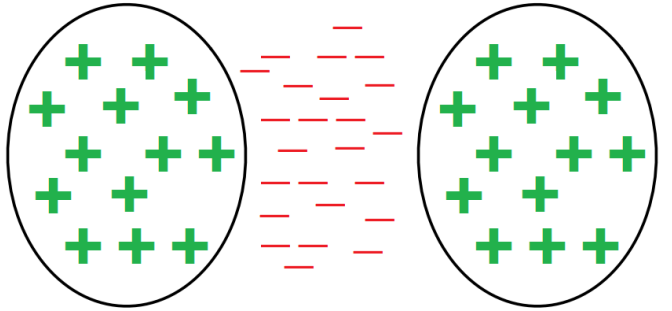
If it weren't for the increasing partisanship, Congress would be even more ineffective!



Paper 7: Classic balance vs. generalized balance

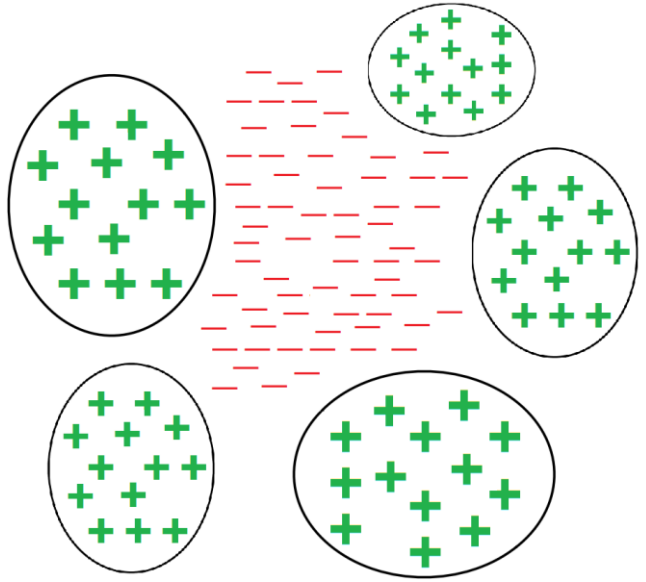
Structurally balanced iff:

The network can be partitioned into 2 clusters with positive edges being within clusters and negative edges being between clusters (Cartwright-Harary 1956)



Weakly balanced (*k*-balanced) iff:

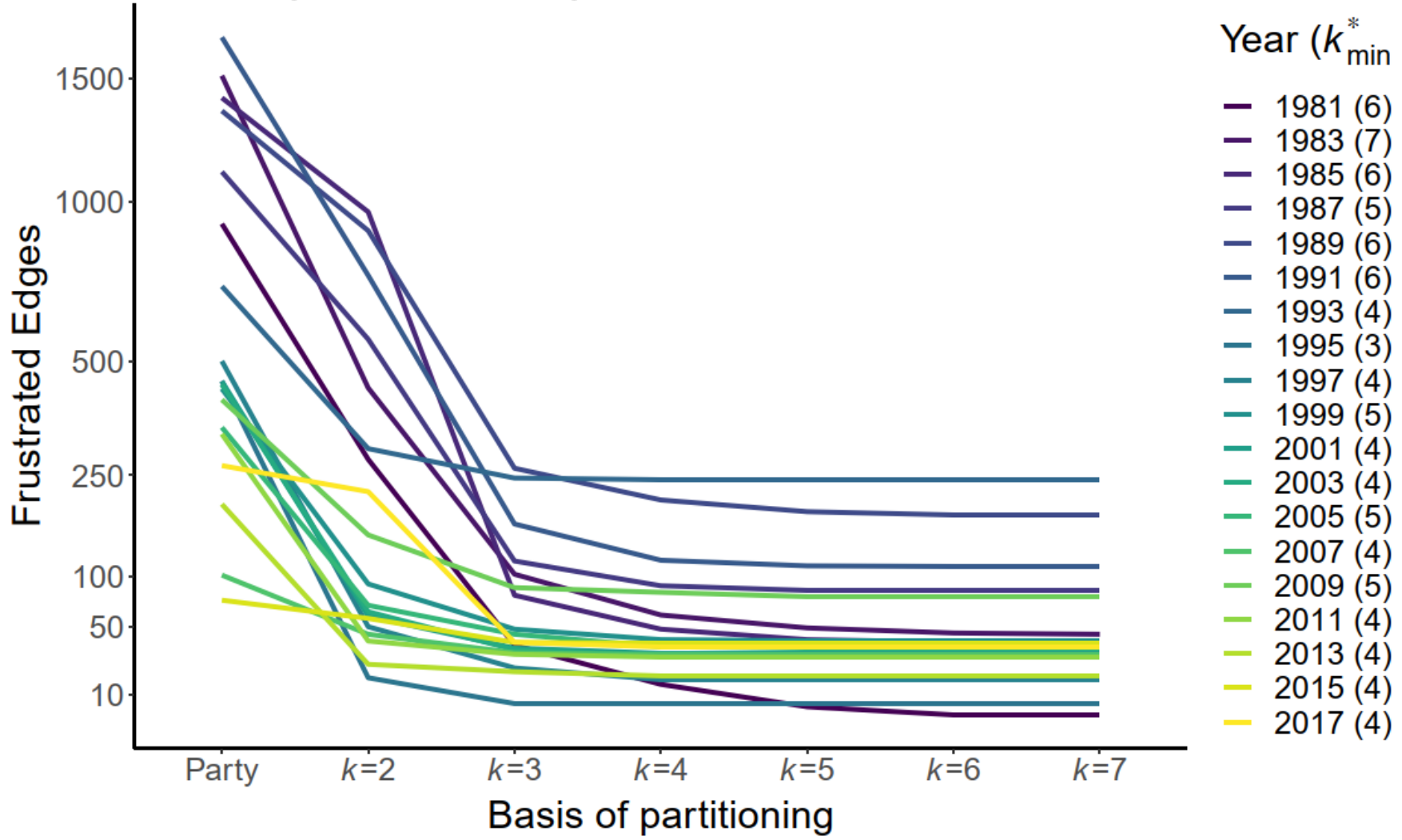
The network can be partitioned into k clusters with positive edges being within clusters and negative edges being between clusters **(Davis 1967)**



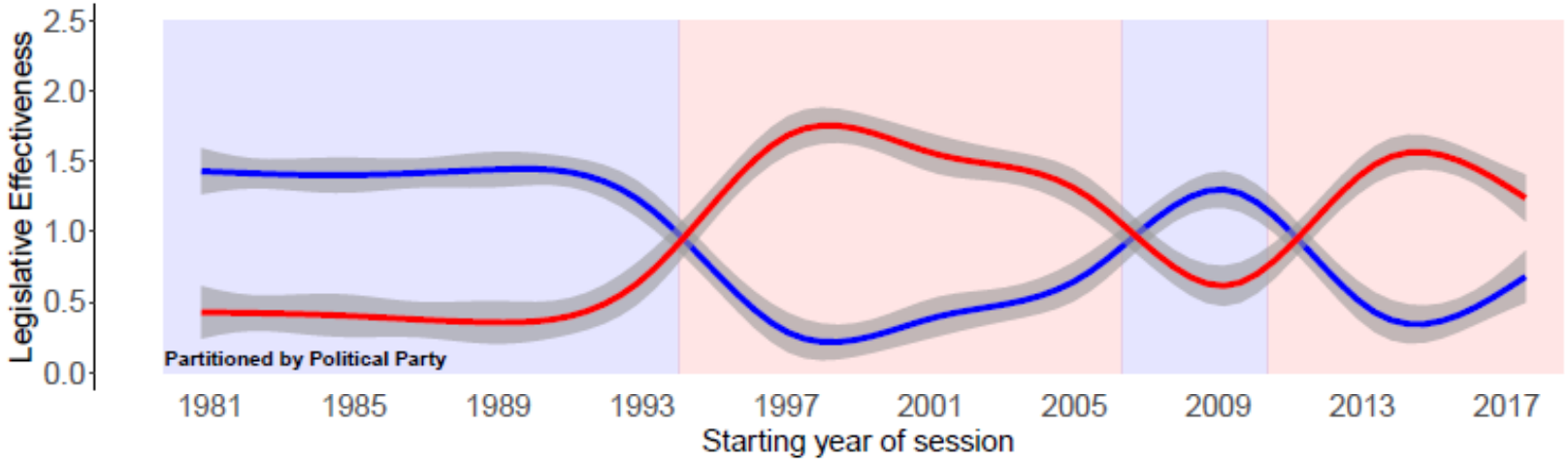
See github.com/saref/clusterability-index for exact calculation of the clusterability index (and k-clusterability index) in networks with up to 30k edges (weak balance)

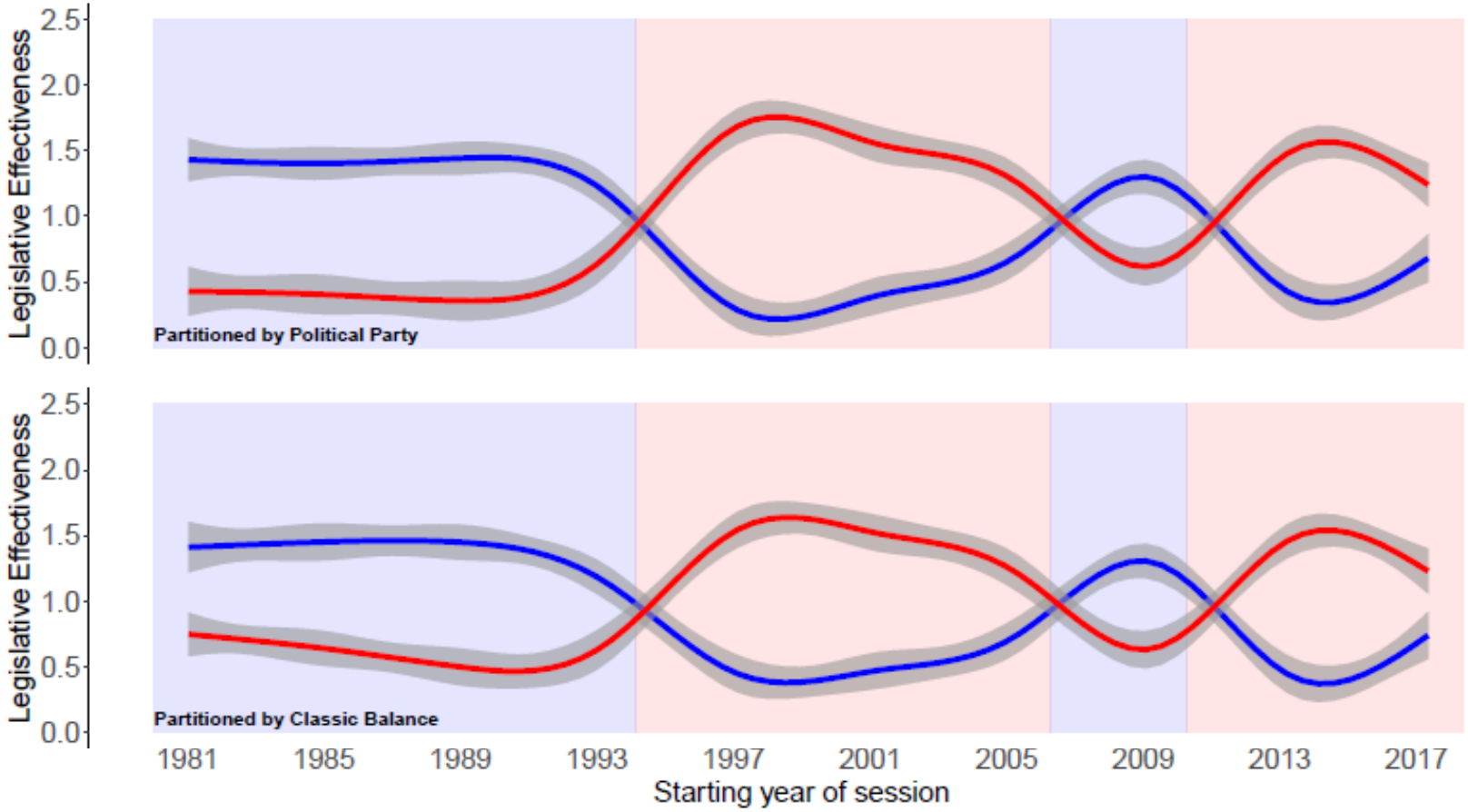
See Angela Fontan et. al 2024 for a closed-form expression that approximates clusterability index in specific types of larger networks (weak balance)

Partitioning the legislators into clusters

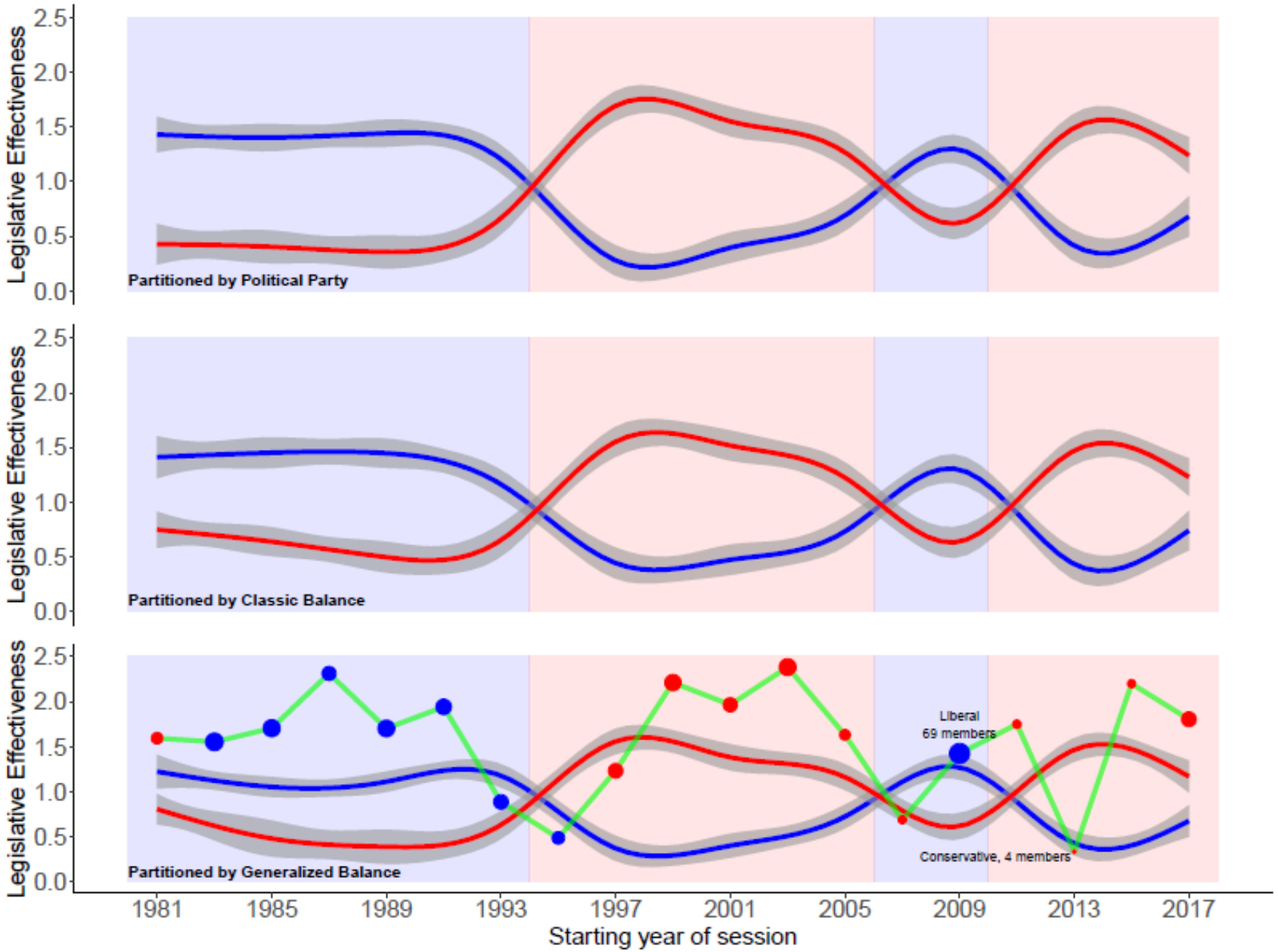


UNIVERSITY OF TORONTO

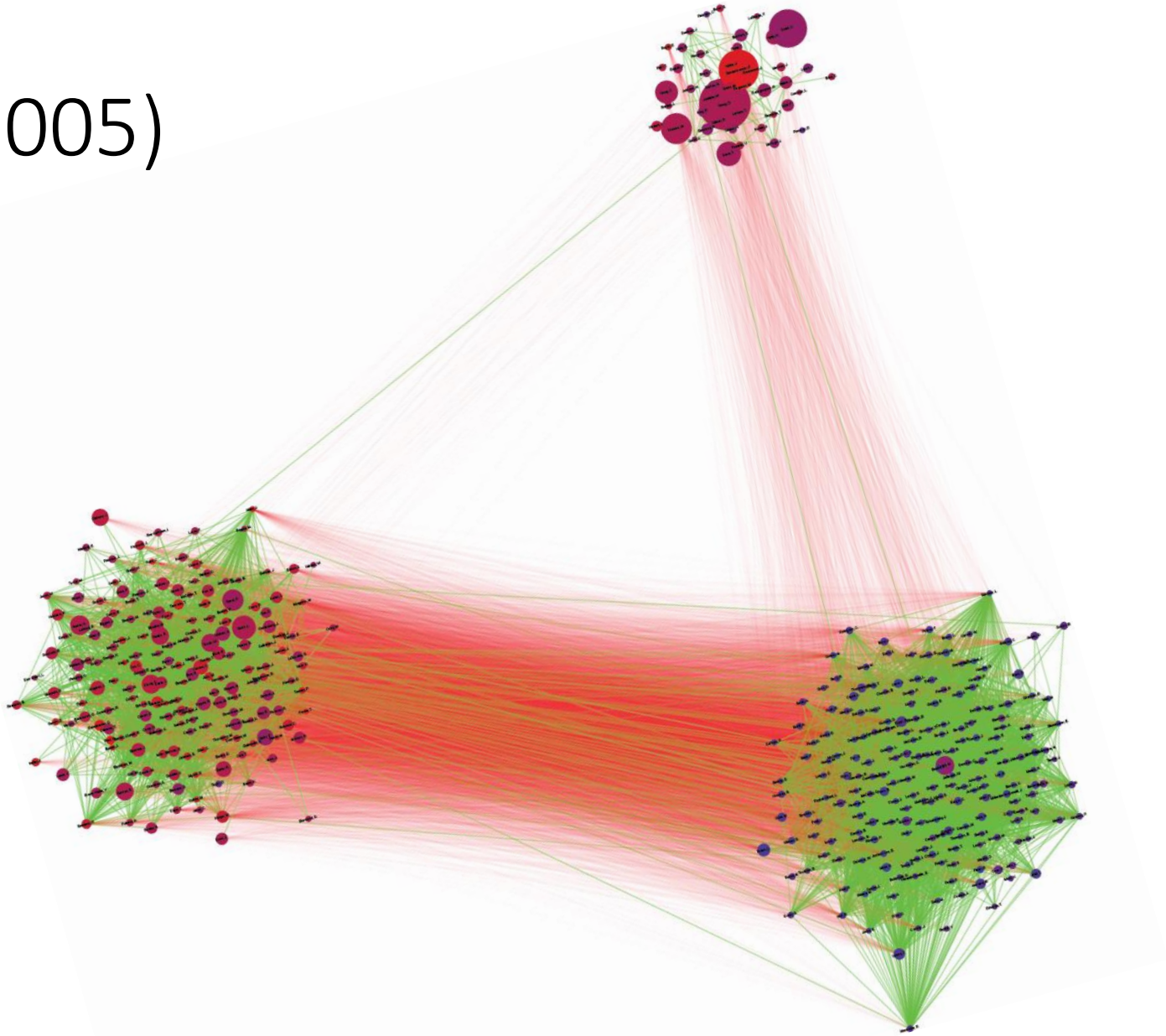

Legislative Effectiveness
2.5
2.0
1.5
1.0
0.5
0.0
Partitioned by Political Party
1981
1985
1989
1993
1997
2001
2005
2009
2013
2017
Starting year of session



UNIVERSITY OF TORONTO


Legislative Effectiveness
2.5
2.0
1.5
1.0
0.5
0.0
Partitioned by Political Party
Partitioned by Classic Balance
Partitioned by Generalized Balance
Liberal
69 members
Conservative, 4 members
1981
1985
1989
1993
1997
2001
2005
2009
2013
2017
Starting year of session

Session 108 (2003-2005)



scientific reports

Explore content ▾ About the journal ▾ Publish with us ▾

nature › scientific reports › articles › article

Article | Open Access | Published: 07 October 2021

Identifying hidden coalitions in the US House of Representatives by optimally partitioning signed networks based on generalized balance

Samin Aref ✉ & Zachary P. Neal

UNIVERSITY OF TORONTO

Peer-reviewed papers ⟶ ArXiv.org
Models and code ⟶ github.com/saref
Network data ⟶ FigShare and OSF, saref.github.io

Measuring balance arxiv.org/abs/1509.04037

Computing frustration index arxiv.org/abs/1710.09876

Frustration index of large networks arxiv.org/abs/1611.09030

Applications of frustration and partitioning arxiv.org/abs/1712.04628

Multilevel evaluation of balance arxiv.org/abs/2005.09925

Bi-partitioning dense signed networks arxiv.org/abs/1906.01696

k-partitioning signed networks arxiv.org/abs/2105.01913

UNIVERSITY OF TORONTO

Special thanks to my collaborators:

Mark C. Wilson
Information and Computer Sciences
University of Massachusetts

Ly Dinh
School of Information
University of South Florida

Andrew J. Mason
Engineering Science
University of Auckland

Rezvaneh (Shadi) Rezapour
Information Science
Drexel University

Zachary Neal
Department of Psychology
Michigan State University

Jana Diesner
Human-Centred Computing
Department of Governance
Technical University of Munich

UNIVERSITY OF TORONTO

Abelson RP, Rosenberg MJ (1958) Symbolic psycho-logic: A model of attitudinal cognition. Behavioral Science 3(1):1–13.

Aref S, Neal ZP (2020) Detecting coalitions by optimally partitioning signed networks of political collaboration. Scientific Reports, 10(1):1–10.

Aref S, Mason AJ, Wilson MC (2018) Computing the line index of balance using integer programming optimisation in Optimization Problems in Graph Theory, ed. Goldengorin B. (Springer), pp. 65–84.

Aref S, Mason AJ, Wilson MC (2020) A Modeling and Computational Study of the Frustration Index in Signed Networks. Networks, 75(1):95-110.

Aref S, Wilson MC (2016) Measuring partial balance in signed networks. Journal of Complex Networks 6(4):566–595.

Aref S, Wilson MC (2019) Balance and frustration in signed networks. Journal of Complex Networks 7(2):163–189.

Aref S, Dinh, L., Rezapour, R., & Diesner, J. (2020). Multilevel structural evaluation of signed directed social networks based on balance theory. Scientific reports, 10(1), 1-12.

Aref S, Neal ZP (2021). Identifying hidden coalitions in the US House of Representatives by optimally partitioning signed networks based on generalized balance. Scientific reports, 11(1), 1-9.

Cartwright D, Harary F (1956) Structural balance: a generalization of Heider's theory. Psychological Review 63(5):277–293

Cartwright D, Harary F (1976) Balance and clusterability: An overview. In P. W. Holland and S. Leinhardt, editors, Social networks: Surveys, advances, and commentaries, pages 25–50. Academic Press.

Davis, JA (1967) Clustering and structural balance in graphs. Human Relations 20 (2): 181–187.

Demaine, E. D., Emanuel, D., Fiat, A. & Immorlica, N. Correlation clustering in general weighted graphs. Theoret. Comput. Sci. 361, 172–187 (2006).

Harary F (1959) On the measurement of structural balance. Behavioral Science 4(4):316–323.

Heider F (1944) Social perception and phenomenal causality. Psychological Review 51(6):358–378.

Iacono, G, (2010) Determining the distance to monotonicity of a biological network: a graph-theoretical approach. IET Systems Biolog., 4, 3, 223–235.

Neal ZP (2014) The backbone of bipartite projections: Inferring relationships from coauthorship,co-sponsorship, co-attendance and other co-behaviors. Social Networks 39:84–97.

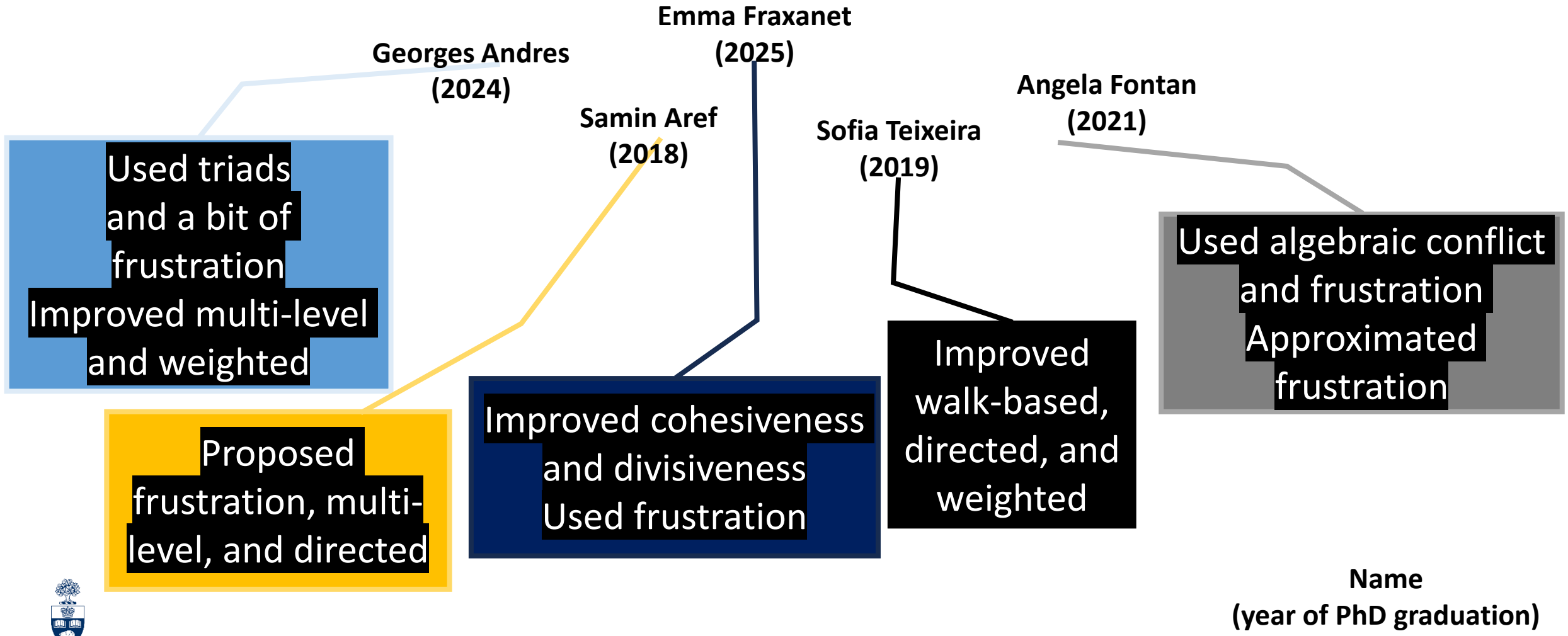
Neal ZP (2020). A sign of the times? Weak and strong polarization in the US Congress, 1973–2016. Social Networks, 60, 103-112.

Terzi E, Winkler M (2011) A spectral algorithm for computing social balance in Proceedings of International Workshop on Algorithms and Models for the Web-Graph, WAW 2011, eds. Frieze A, Horn P, Prałat P. (Springer), pp. 1–13.

Zaslavsky T (1987) Balanced decompositions of a signed graph. Journal of Combinatorial Theory, Series B 43(1):1–13.

UNIVERSITY OF TORONTO

With their co-authors, these early-career researchers have…



Questions?


aref@mie.utoronto.ca


@SaminAref


saref.github.io

UNIVERSITY OF TORONTO