# Opinions and conflict in social networks: models, computational problems and algorithms

## Lecture 4: Mining signed networks: theory and applications

Bertinoro International Spring School 2022

*Aristides Gionis*

`argioni@kth.se`

KTH Royal Institute of Technology

# course overview

- lecture 1: introduction
  - polarization in social media; methods for detecting polarization
- lecture 2: mathematical background
  - submodular maximization; spectral graph theory
- lecture 3: methods for mitigating polarization
  - maximizing diversity, balancing information exposure
- lecture 4: signed networks; theory and applications
- lecture 5: opinion dynamics in social networks

# overview of this lecture

- introduction / motivation
- theory of signed networks
- application
  - finding polarized communities in signed networks

# motivation of signed networks

# signed networks (or signed graphs)

graphs with edge signs,

i.e., edge labels can be either positive or negative

# signed graphs: motivation

**human interactions**

- friendly or antagonistic

image source: pxfuel.com

# signed graphs: motivation

**online social media**

- a graph of users (twitter, facebook, etc.)
- users may express like or dislike towards others
- can be used to study online polarization

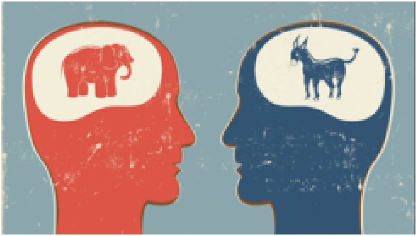
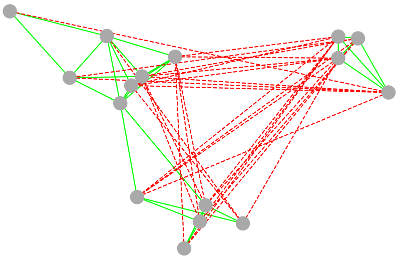


image source: iStockphoto.com

# signed graphs: motivation

**groups of humans**

- examples: political parties, countries, etc.
  - political polarization
  - country relations during war



New Guinea highland tribes graph [Read, 1954]

# signed graphs: motivation

**human language**

- a graph of words that captures synonym or antonym relations

“Happy”

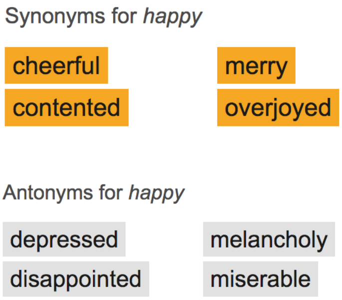



image source: thesaurus.com

# signed graphs: motivation

**molecular biology**

- a graph of proteins

  a protein activates or inhibits the functioning of another

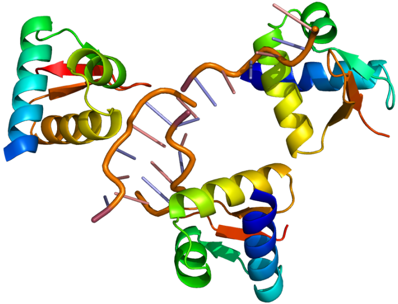


image source: commons.wikimedia.org

# signed graphs: motivation

**finance**

- a graph of securities (tradable assets)
- a security *correlates* positively/negatively with another
- here, "correlation" indicates that the prices of two securities move jointly

image source: vecteezy.com

# theory of signed networks

# outline

- we will discuss:
  - balance
  - spectrum
  - frustration

# balance

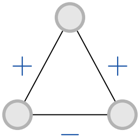
# signed networks

signed networks (or graphs): each edge labeled $+$ or $-$

definitions:

- $G = (V, E^+, E^-)$,
- $G = (V, E, \sigma)$, where $\sigma : E \to \{-, +\}$

adjacency matrix: $A = A_{E^+} - A_{E^-}$

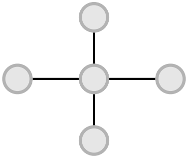



$$\mapsto \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}$$

# expressiveness of signed graphs

signed graphs can be quite expressive

example: star graph with $|E|$ edges

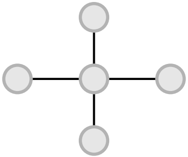


- number of possible unsigned graphs: 1
- number of possible signed graphs: ?
- number of non-isomorphic signed graphs: ?

# expressiveness of signed graphs

signed graphs can be quite expressive

example: star graph with $|E|$ edges



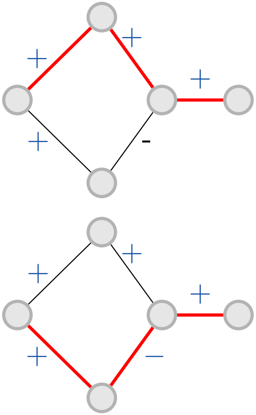
- number of possible unsigned graphs: $1$
- number of possible signed graphs: $2^{|E|}$
- number of non-isomorphic signed graphs: $|E| + 1$

# differences in signed graphs — shortest paths

signed graphs can be quite different...

consider shortest paths; how do we even define path length in signed graphs?

proposal: distinguish positive and negative paths (by product of edge signs).




positive path

negative path

finding all shortest simple signed paths is **NP**-complete!

if repetitions are allowed, $\mathcal{O}(|E| \log\log \frac{D}{d})$ algorithm

[Hansen, 1984]

# differences in signed graphs — densest subgraph

densest subgraph problem in unsigned graphs:

$$\max_{x \in \{0,1\}^n} \frac{x^T A x}{x^T x}$$

polynomial-time solvable [Goldberg, 1984]

# differences in signed graphs — densest subgraph

densest subgraph problem in unsigned graphs:

$$\max_{x \in \{0,1\}^n} \frac{x^T A x}{x^T x}$$

polynomial-time solvable [Goldberg, 1984]

densest subgraph problem in signed graphs:

$$\max_{x \in \{-1,0,1\}^n} \frac{x^T A x}{x^T x}$$

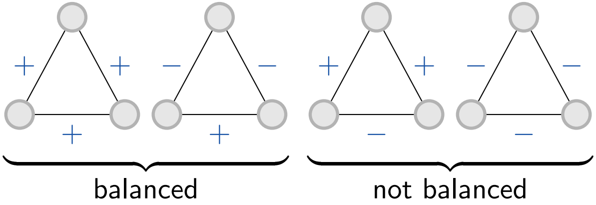
**NP**-hard [Bonchi et al., 2019, Tsourakakis et al., 2019]

## balance

# motivation

motivation: balance in social networks [Harary, 1953]

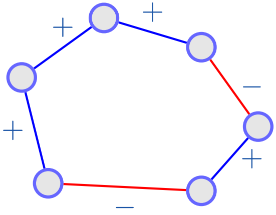
"*the friend of a friend is a friend*" (or "*the enemy of a friend is an enemy*")




the four possible non-isomorphic signed triangles

# motivation

balance applies to cycles of any length




$$+ \times - \times + \times - \times + \times + = +.$$

## definition of balanced cycle

a cycle is balanced if the product of its signs is positive

# motivation

## characterizations of balance

a signed graph $G$ is balanced if and only if

- there are no negative (unbalanced) cycles.
- there exists a sign-compliant partition $V = V_1 \cup V_2$, such that all $+$ edges within sets and all $-$ edges between sets
- all paths between any pair $u, v$ have same sign

some balanced graphs

# measures of partial balance

how can we measure partial balance?

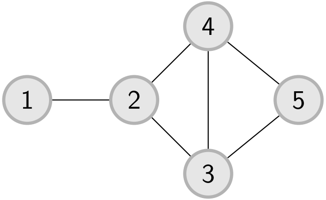
- fraction of balanced cycles [Cartwright and Harary, 1956, Giscard et al., 2017]
  - fraction of balanced triangles [Terzi and Winkler, 2011] (example in next slide)
- spectral methods (discussed later on)

check [Aref and Wilson, 2018] for an overview of partial measures of balance.

# measures of partial balance — example: fraction of balanced triangles

reminder: counting triangles in unsigned graphs:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 3 & 1 & 1 & 2 \\ 1 & 1 & 3 & 2 & 1 \\ 1 & 1 & 2 & 3 & 1 \\ 0 & 2 & 1 & 1 & 2 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 0 & 3 & 1 & 1 & 2 \\ 3 & 2 & 6 & 6 & 2 \\ 1 & 6 & 4 & 5 & 5 \\ 1 & 6 & 5 & 4 & 5 \\ 2 & 2 & 5 & 5 & 2 \end{pmatrix}$$

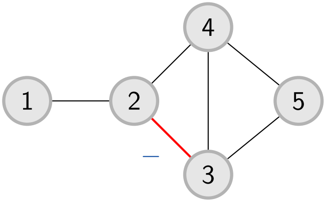



$A^3_{ii} = 2 \times \#\{\text{3-cycles adjacent to vertex } i\}$

# measures of partial balance — example: fraction of balanced triangles

counting triangles in signed graphs:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 1 & 0 & -1 & 1 & 0 \\ 0 & 3 & 1 & -1 & 0 \\ -1 & 1 & 3 & 0 & 1 \\ 1 & -1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & 2 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 0 & 3 & 1 & -1 & 0 \\ 3 & -2 & -4 & 4 & 0 \\ 1 & -4 & 0 & 5 & 3 \\ -1 & 4 & 5 & 0 & 3 \\ 0 & 0 & 3 & 3 & 2 \end{pmatrix}$$




$$A^3_{ii} = 2 \times (\#\{\text{balanced 3-cycles}\} - \#\{\text{unbalanced 3-cyles}\})$$

thus,

$$\frac{Tr(A^3) + Tr(|A|^3)}{2\,Tr(|A|^3)} = \text{fraction of balanced triangles}$$

[Terzi and Winkler, 2011]

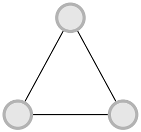
note: $|A|$ is the adj. matrix of the *underlying* (unsigned) graph

## spectrum

# spectral theory

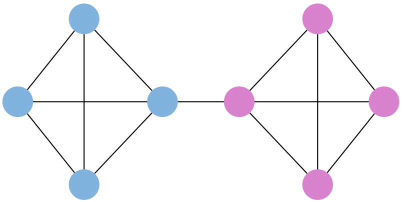
review of spectral theory for unsigned graphs:

Laplacian: $L = D - A$



$$Lv_1 = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0$$

- $\lambda_{min}(L) = 0$ (multiplicity of $0 = \#\{\text{of connected components}\}$)
- eigenvector $v_2$ gives a "good" partition (Cheeger inequality)

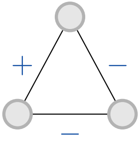


$$v_2 \approx \begin{pmatrix} -0.25 \\ -0.38 \\ -0.38 \\ -0.38 \\ 0.38 \\ 0.38 \\ 0.38 \\ 0.25 \end{pmatrix}, \quad \lambda_2(L) \approx 0.35$$

# spectral theory for signed graphs

Laplacian: $L = D - A$

| Unsigned | signed |
|---|---|
| $L$ is positive semidefinite | |
| $D_{ii} = \sum_j A_{ij}$ | $D_{ii} = \sum_j \lvert A_{ij} \rvert$ |
| $\lambda_{min}(L) = 0$ | $\lambda_{min}(L) \geq 0$ |

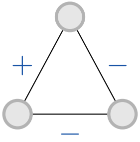



$$L = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

# spectral theory for signed graphs

Laplacian: $L = D - A$

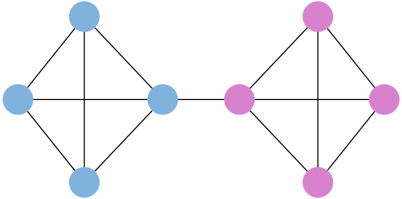
| Unsigned | signed |
|---|---|
| $L$ is positive semidefinite | |
| $D_{ii} = \sum_j A_{ij}$ | $D_{ii} = \sum_j \lvert A_{ij} \rvert$ |
| $\lambda_{min}(L) = 0$ | $\lambda_{min}(L) \geq 0$ |




$$L v_1 = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = 0$$

# spectral theory for signed graphs

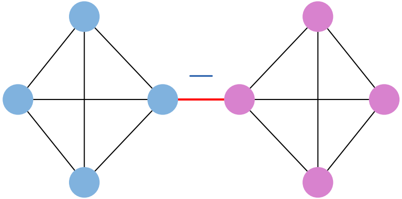
consider our previous graph;



$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \lambda_{min}(L) = 0.$$

# spectral theory for signed graphs

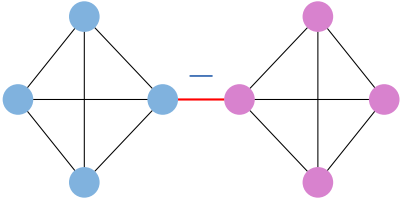
consider our previous graph; flip the sign of one edge:




$$v_1 = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \lambda_{min}(L) = 0.$$

# spectral theory for signed graphs

consider our previous graph; flip the sign of one edge:




$$v_1 = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \lambda_{min}(L) = 0.$$

this graph is balanced !

## spectral characterizations of balance

- connected and $\lambda_{min} = 0$ (or one zero eigenvalue per connected component)

# spectral theory

a taste of spectral analysis

lemma [Hou et al., 2003]
$\lambda_{max}(L(G)) \leq \lambda_{max}(L(G^-))$, where $G^-$ is the all-negative graph.

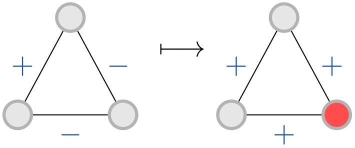
proof: $L(G^-)$ is the *signless Laplacian* of the underlying graph $(D_{|G|} + A_{|G|})$
so, $x^T L x = \sum_{(v_i,v_j)\in E}(x_i - \sigma(v_i, v_f)x_j)^2 \leq \sum_{(v_i,v_j)\in E}(|x_i| + |x_j|)^2)$

lemma [Hou et al., 2003]
$\lambda_{max}(L(G)) \leq 2(n-1)$, where n is the number of vertices

proof: $\lambda_{max}(G) = \lambda_{max}(D - A) \leq \lambda_{max}(D_G) + \lambda_{max}(-A_G) \leq n - 1 + n - 1$

# switching

switch $S \subseteq V$: flip edges between $S$ and $V \backslash S$

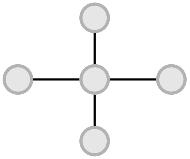



the spectrum is invariant with respect to switching.

$$A' = SAS^{-1}, \text{ where } S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

# spectral theory for signed graphs

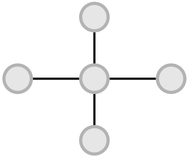
example: star graph



- number of possible graphs: $2^{|E|}$
- number of non-isomorphic graphs: $|E| + 1$
- number of distinct spectra: ?

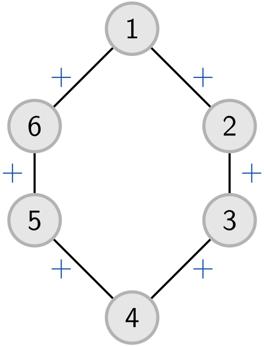
# spectral theory for signed graphs

example: star graph



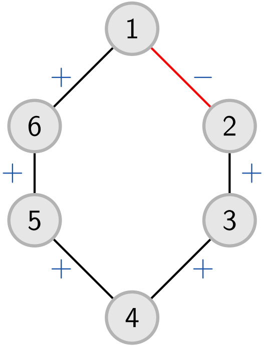
- number of possible graphs: $2^{|E|}$
- number of non-isomorphic graphs: $|E| + 1$
- number of distinct spectra: just the one!

# example: cycle



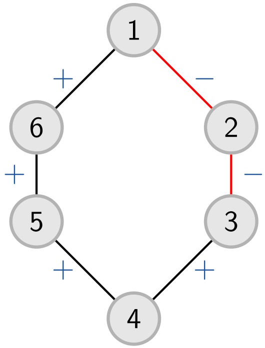

spectrum: $(0, 1, 1, 3, 3, 4)$

# example: cycle




spectrum: $(0.27, 0.27, 2, 2, 3.73, 3.73)$

# example: cycle

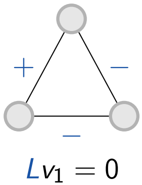


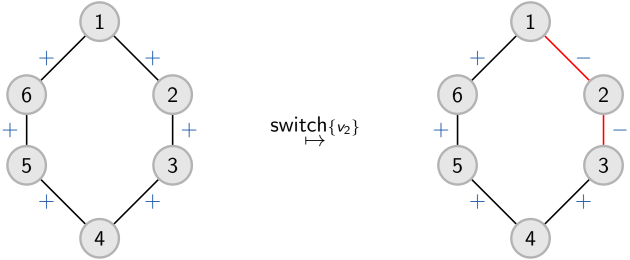

spectrum: $(0, 1, 1, 3, 3, 4)$

# switching

## spectral characterizations of balance

1. connected and $\lambda_{min} = 0$ (or one zero-eigenvalue per connected component)
2. spectrum of $G$ = spectrum of $|G|$ (underlying graph) [Acharya, 1980]
3. switching equivalent to all-positive




what's more: $\lambda_{min}(G) \leq \lambda_{max}(H)$, where $H$ is the smallest subgraph to remove to make $G$ balanced [Li and Li, 2016]

# frustration

# frustration

- we distinguish
  - vertex frustration ($f_v$): # vertices need to remove to achieve balance

    also known as frustration number
  - edge frustration ($f_e$): # edges need to remove to achieve balance

    also known as frustration index
- spectral frustration inequalities [Belardo, 2014]

$$\lambda_{min}(L) \leq f_v \leq f_e$$

proof left as exercise

# frustration

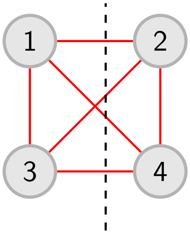
- how hard is vertex frustration?

  i.e., we want to find the minimum vertex frustration?

- minimization problem is **NP**-hard
- dual problem: finding the largest balanced subgraph
  - dual problem admits a 2-approximation on complete graphs [Bai and Wu, 2012]
- other than that, not that much is known

# frustration

- how hard is edge frustration?
- consider an all-negative signed graph

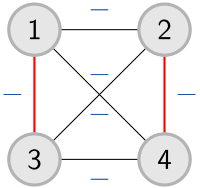



- finding the minimum $f_e$ is finding the $\textsc{MaxCut}$, and thus **NP**-complete
- corollary: all-negative is balanced $\Leftrightarrow$ it is bipartite

# frustration

- how hard is edge frustration?
- every signed graph $G = (V, E)$ contains a balanced subgraph with at least $\frac{|E|}{2} + \frac{|V|-1}{4}$ edges
- $f_e$ is (UG)-hard to approximate to any constant
- FPT: find subgraph of $|E| - k$ in $\mathcal{O}(2^k m^2)$

  [Hüffner et al., 2007]
- dual problem: can be approximated with factor $\mathcal{O}(\sqrt{\log n})$ (MinUnCut problem) [Agarwal et al., 2005]



## detecting polarization in signed networks

[Bonchi et al., 2019]

# discovering polarization in signed networks

what should polarization look like in signed networks ?

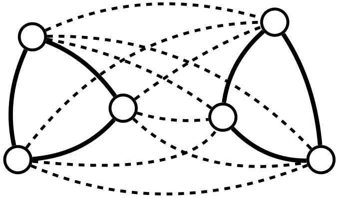
- assume two conflicting communities
- lots of $+$ edges within communities, lots of $-$ across
- few $-$ edges within communities, few $+$ across

# discovering polarization in signed networks

what should polarization look like in signed networks ?

- assume two conflicting communities
- lots of $+$ edges within communities, lots of $-$ across
- few $-$ edges within communities, few $+$ across

"ideal" polarized structure in a signed network:



# discovering polarization in signed networks

- problem formulation: given a signed network $G = (V, E_+, E_-)$ find disjoint subsets of vertices $S_1, S_2 \subseteq V$ to maximize the number of *"congruent"* edges minus the *"non-congruent"* ones
- alternatively: find $\mathbf{x} : V \to \{-1, 0, 1\}$ to maximize $\mathbf{x}^\top A\,\mathbf{x}$, where $A$ is the adjaceny matrix of $G$
- interested in density rather than absolute number
- 2PC problem: find $\mathbf{x} : V \to \{-1, 0, 1\}$ to maximize the ratio

$$\frac{\mathbf{x}^\top A\,\mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

# discovering polarization in signed networks

- 2PC problem: find $\mathbf{x} : V \to \{-1, 0, 1\}$ to maximize

$$\frac{\mathbf{x}^\top A \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

- problem can be seen as discrete eigenvalue
- also related to correlation clustering
- lemma: 2PC problem is **NP**-hard

# eigenvector algorithm

- EIGEN:

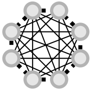
1. compute principal eigenvector $\mathbf{v}$ of $A$
2. set $x_i = sign(v_i)$
3. output $\mathbf{x}$

- EIGEN is $\mathcal{O}(n)$ approximation
- enhancement: set $x_i = 0$ if $|v_i| \leq \tau$

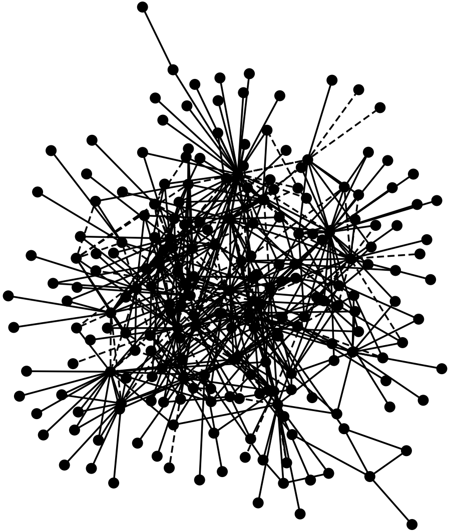
# randomized eigenvector algorithm

- RANDOM-EIGEN:

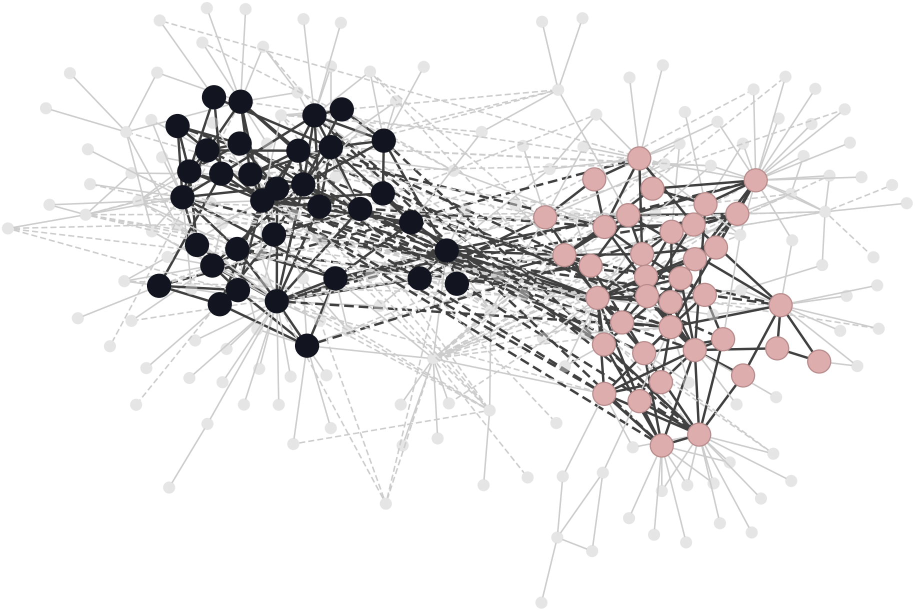
1. compute principal eigenvector $\mathbf{v}$ of $A$
2. set $x_i = sign(v_i)$ with probability $|v_i|$, o/w $x_i = 0$
3. output $\mathbf{x}$

- RANDOM-EIGEN provides a $\mathcal{O}(\sqrt{n})$ approximation to optimal solution
- analysis is tight; example: 
- enhancement: $\mathbf{v} \mapsto \|v\|_1 \mathbf{v}$
  - think of the case $v_i = v_j$ for all $i, j$

# example: US congress



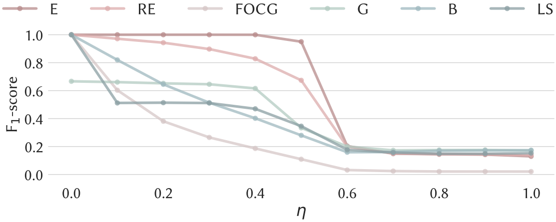
# example: US congress



# results on planted polarized communities

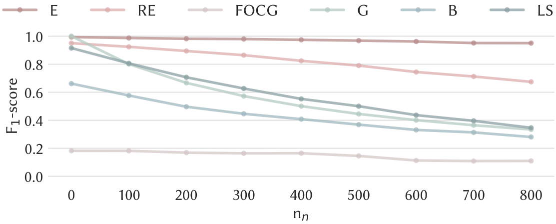
$F_1$-score as a function of noise parameter $\eta$

$(n_c = 100,\ n_n = 800)$




$F_1$-score as a function of the number of noisy vertices $n_n$

$(n_c = 100,\ \eta = 0.5)$



# references I


Acharya, B. D. (1980).
Spectral criterion for cycle balance in networks.
*Journal of Graph Theory*, 4(1):1–11.

Agarwal, A., Charikar, M., Makarychev, K., and Makarychev, Y. (2005).
O (sqrt log n) approximation algorithms for min uncut, min 2cnf deletion, and directed cut problems.
In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 573–581.

Aref, S. and Wilson, M. C. (2018).
Measuring partial balance in signed networks.
*Journal of Complex Networks*, 6(4):566–595.

Bai, C.-H. and Wu, B. Y. (2012).
Finding the maximum balanced vertex set on complete graphs.
In *THE 29TH WORKSHOP ON COMBINATORIAL MATHEMATICS AND COMPUTATION THEORY*. Citeseer.

Belardo, F. (2014).
Balancedness and the least eigenvalue of Laplacian of signed graphs.
*Linear Algebra and its Applications*, 446:133–147.

# references II


Bonchi, F., Galimberti, E., Gionis, A., Ordozgoiti, B., and Ruffo, G. (2019).
Discovering polarized communities in signed networks.
In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 961–970.

Cartwright, D. and Harary, F. (1956).
Structural balance: a generalization of heider's theory.
*Psychological review*, 63(5):277.

Giscard, P.-L., Rochet, P., and Wilson, R. C. (2017).
Evaluating balance on social networks from their simple cycles.
*Journal of Complex Networks*, 5(5):750–775.

Goldberg, A. V. (1984).
*Finding a maximum density subgraph*.
University of California Berkeley.

Hansen, P. (1984).
Shortest paths in signed graphs.
In *North-Holland mathematics studies*, volume 95, pages 201–214. Elsevier.

# references III


Harary, F. (1953).
On the notion of balance of a signed graph.
*The Michigan Mathematical Journal*, 2(2):143–146.

Hou, Y., Li, J., and Pan, Y. (2003).
On the Laplacian eigenvalues of signed graphs.
*Linear and Multilinear Algebra*, 51(1):21–30.

Hüffner, F., Betzler, N., and Niedermeier, R. (2007).
Optimal edge deletions for signed graph balancing.
In *International Workshop on Experimental and Efficient Algorithms*, pages 297–310.

Li, H. S. and Li, H. H. (2016).
A note on the least (normalized) laplacian eigenvalue of signed graphs.
*Tamkang Journal of Mathematics*, 47(3):271–278.

Read, K. E. (1954).
Cultures of the central highlands, new guinea.
*Southwestern Journal of Anthropology*, 10(1):1–43.

# references IV


Terzi, E. and Winkler, M. (2011).
A spectral algorithm for computing social balance.
In *International Workshop on Algorithms and Models for the Web-Graph*, pages 1–13.

Tsourakakis, C. E., Chen, T., Kakimura, N., and Pachocki, J. (2019).
Novel dense subgraph discovery primitives: Risk aversion and exclusion queries.
*arXiv preprint arXiv:1904.08178*.