Rayleigh quotient maximization and its applications to social network analysis (50% seminar)

Presenter: Ruo-Chun Tzeng (KTH)

Main Supervisor: Aristides Gionis (KTH)
Co-Supervisor: Alexandre Proutière (KTH)

Opponent: Aditya Bhaskara (University of Utah)

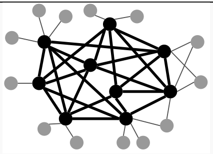
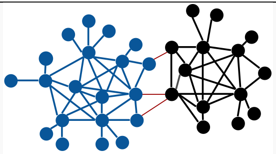
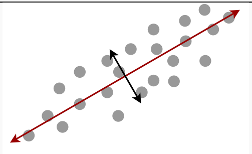
March 23rd, 2022

Introduction: Rayleigh quotient maximization over $\mathcal{T}$ (1/2)

Given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and a feasible set $\mathcal{T} \subseteq \mathbb{R}^n$, find any

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \mathcal{T} \backslash \{\mathbf{0}\} \right\}.$$

Examples

densest subgraph [1]	2-community detection [9]	PCA [7]
$\mathcal{T} = \{0,1\}^n$	$\mathcal{T} = \{-1,1\}^n$	$\mathcal{T} = \mathbb{R}^n$
		

Introduction: Rayleigh quotient maximization over $\mathcal{T}$ (2/2)

Given a symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathcal{T} \subseteq \mathbb{R}^n$, find any

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \mathcal{T} \backslash \{\mathbf{0}\} \right\}.$$

Discrete $\mathcal{T}$: graph applications

Questions of interests:

- What structures does $\mathbf{u}^\star$ capture?
- Is $\mathbf{u}^\star$ polynomial-time tractable?
- How well can we approximate $\mathbf{u}^\star$?

$\mathcal{T} = \mathbb{R}^n$: numerical linear algebra

Questions of interests:

- How to evaluate the quality of $\mathbf{u}^\star$?
- How well can we approximate $\mathbf{u}^\star$ under computational limitations?

Contributions

Given a symmetric $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathcal{T} \subseteq \mathbb{R}^n$, find any

$$\mathbf{u}^\star \in \operatorname{argmax}\left\{\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} : \mathbf{x} \in \mathcal{T}\backslash\{\mathbf{0}\}\right\}.$$

Part I - discrete $\mathcal{T}$

Studied a signed graph application.

- $\mathbf{u}^\star$ captures antagonistic patterns
- known to be APX-hard
- a provable approximation algorithm

Part II - $\mathcal{T} = \mathbb{R}^n$

Improved analysis of a numerical solver

- for the multiplicative gap $R(\hat{\mathbf{u}})$
- under memory-limited and pass-limited setting

(Part I) Tzeng et al. "Discovering conflicting groups in signed networks." In Proc. of NeurIPS 2020.
(Part II) Tzeng et al. "Improved analysis of randomized SVD for top-eigenvector approximation." In Proc. of AISTATS 2022.

Part I - an application in signed graphs

Given a symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathcal{T} \subseteq \mathbb{R}^n$, find any

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \mathcal{T} \backslash \{\mathbf{0}\} \right\}.$$

Part I - discrete $\mathcal{T}$

Studied a signed graph application.

- $\mathbf{u}^\star$ captures antagonistic patterns
- known to be APX-hard
- a provable approximation algorithm

Part II - $\mathcal{T} = \mathbb{R}^n$

Improved analysis of a numerical solver

- for the multiplicative gap $R(\hat{\mathbf{u}})$
- under memory-limited and pass-limited setting

(Part I) Tzeng et al. "Discovering conflicting groups in signed networks." In Proc. of NeurIPS 2020.
(Part II) Tzeng et al. "Improved analysis of randomized SVD for top-eigenvector approximation." In Proc. of AISTATS 2022.

Conflicting group detection in signed graphs (1/8)

(Bonchi et al. 2019) 2-conflicting group detection

Given a signed adjacency matrix $\mathbf{A} \in \{-1, 0, 1\}^{n \times n}$, the 2-conflicting groups [3] are identified by the signs of

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \{-1, 0, 1\}^n \backslash \{\mathbf{0}\} \right\} \quad (1)$$

$\mathbf{A}_{i,j}\mathbf{x}_i\mathbf{x}_j = 1$:

- $+$ intra-group edges
- $-$ inter-group edges

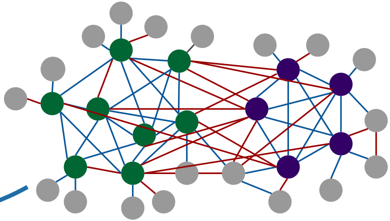
$\mathbf{A}_{i,j}\mathbf{x}_i\mathbf{x}_j = -1$:

- $-$ intra-group edges
- $+$ inter-group edges

antagonistic property:

intra-group: mostly $+$

inter-group: mostly $-$



Conflicting group detection in signed graphs (2/8)

(Bonchi et al. 2019) 2-conflicting group detection

Given a signed adjacency matrix $\mathbf{A} \in \{-1, 0, 1\}^{n \times n}$, the 2-conflicting groups [3] are identified by the signs of

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \{-1, 0, 1\}^n \backslash \{\mathbf{0}\} \right\} \tag{1}$$

Hardness result of (1)

NP-hard: independently proven by

(Bonchi et al. 2019) reduction from Correlation Clustering

APX-hard proven by (Bhaskara et al. 2012)

version 1

reduction from Max k-AND

version 2

reduction from ratio version of Unique Games Conjecture

Conflicting group detection in signed graphs (3/8)

(Bonchi et al. 2019) 2-conflicting group detection

Given a signed adjacency matrix $\mathbf{A} \in \{-1, 0, 1\}^{n \times n}$, the 2-conflicting groups [3] are identified by the signs of

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \{-1, 0, 1\}^n \setminus \{\mathbf{0}\} \right\} \tag{1}$$

Approximation algorithms

SDP-based (Bhaskara et al. 2012)

- general: $\tilde{\mathcal{O}}(n^{1/3})$
- bipartite: $\tilde{\mathcal{O}}(n^{1/4})$ (gap instance)

Eigenvector-based

- (Bonchi et al. 2019) $\mathcal{O}(n^{1/2})$
- (Tzeng et al. 2020) $\Omega(n^{1/2})$

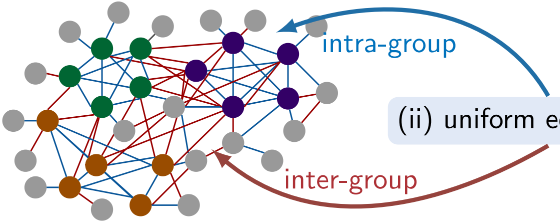
Conflicting group detection in signed graphs (4/8)

(Tzeng et al. 2020) formulation of k-conflicting group

k-conflicting groups are identified by the optimal solution to

$$\max_{\substack{\mathbf{X}\in\{0,1\}^{n\times k}\setminus\{\mathbf{0}\} \\ \mathbf{X}_{i,:}\in\{\mathbf{0},\mathbf{e}_1^T,\cdots,\mathbf{e}_k^T\},\forall i\in[n]}} \frac{\sum_{(i,j):\langle \mathbf{X}_{i,:},\mathbf{X}_{j,:}\rangle=1} \mathbf{A}_{i,j} + \frac{1}{k-1}\sum_{(i,j):\langle \mathbf{X}_{i,:},\mathbf{X}_{j,:}\rangle=0} \mathbf{A}_{i,j}}{\mathrm{Tr}(\mathbf{X}^T\mathbf{X})}.$$

(i) equally-sized groups



(ii) uniform edge density

Intuition: under (i)(ii), # inter-group edge $\approx (k-1)\times$ # intra-group edge

Conflicting group detection in signed graphs (5/8)

$$= \sum_{j=1}^{k-1} \frac{\|\mathbf{Y}_{:,j}\|_F^2}{\|\mathbf{Y}\|_F^2} \frac{\mathbf{Y}_{:,j}^T \mathbf{A}^{(j-1)} \mathbf{Y}_{:,j}}{\mathbf{Y}_{:,j}^T \mathbf{Y}_{:,j}}$$

(Tzeng et al. 2020) an equivalent objective of k-conflicting group

$$\max_{\mathbf{Y} \in \mathbb{R}^{n \times (k-1)} \setminus \{\mathbf{0}\}} \frac{\mathrm{Tr}(\mathbf{Y}^T \mathbf{A} \mathbf{Y})}{\mathrm{Tr}(\mathbf{Y}^T \mathbf{Y})} \text{ subject to } \mathbf{Y}_{i,j} = \begin{cases} c_j(k-j) & \text{if } i \in S_j \\ 0 & \text{if } i \in \cup_{h=1}^{j-1} S_h \text{ or } i \notin \cup_{h \in [k]} S_h \\ -c_j & \text{if } i \in \cup_{h=j+1}^{k} S_h \end{cases}, \tag{2}$$

where $\{c_j\}_{j \in [k-1]}$ are fixed constants, and $S_1, \cdots, S_k$ are any k disjoint groups.

(Tzeng et al. 2020) a sequential algorithm called SCG

Suppose $S_1, \cdots, S_{j-1}$ are found, we find $S_j = \{i \in [n] : \mathbf{u}_i^\star = k - j\}$ by solving

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A}^{(j-1)} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \{-1, 0, k-j\}^n \setminus \{\mathbf{0}\} \right\}, \tag{3}$$

where $\mathbf{A}^{(j-1)}$ is the adjacency matrix after removing $\cup_{h \in [j-1]} S_h$ and $\mathbf{A}^{(0)} = \mathbf{A}$.

Conflicting group detection in signed graphs (6/8)

(Tzeng et al. 2020) The subproblem in SCG

Let $q \in [k-1]$ and $\mathbf{A} \in \{0,1\}^{n \times n}$ be the (modified) adjacency matrix.

$$\mathbf{u}^\star \in \operatorname{argmax}\left\{\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} : \mathbf{x} \in \{-1, 0, q\}^n \setminus \{\mathbf{0}\}\right\}. \tag{3}$$

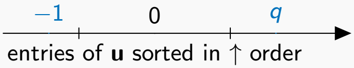
Solving (3): approximation algorithms

Eigenvector-based: let $\mathbf{u}$ be the leading eigenvector of $\mathbf{A}$.

- randomized: $\mathcal{O}(qn^{1/2})$-approx generalizes (Bonchi et al. 2019)

$$\tilde{\mathbf{u}}_i = \begin{cases} q \cdot \text{Bernoulli}(|\mathbf{u}_i|) & \text{if } \mathbf{u}_i > 0 \\ -1 \cdot \text{Bernoulli}(|\mathbf{u}_i|) & \text{if } \mathbf{u}_i < 0 \end{cases}$$

- deterministic:

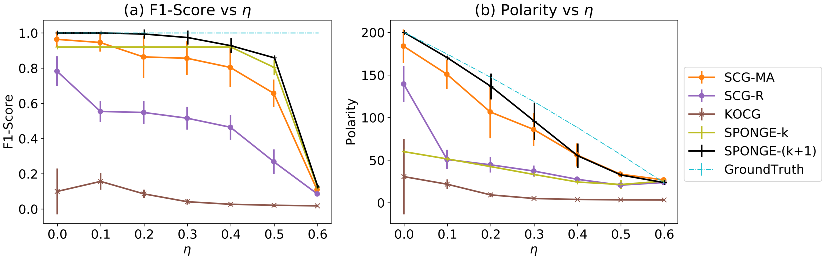


Conflicting group detection in signed graphs (7/8)

- Real-world networks:

	Bitcoin	WikiVote	Referendum	Slashdot	WikiConflict	Epinions	Wikipolitics
$\|V\|$	5 881	7 115	10 884	82 140	116 717	131 580	138 587
$\|E\|$	21 492	100 693	251 406	500 481	2 026 646	711 210	715 883
$\|E_-\|/\|E\|$	0.2	0.2	0.1	0.2	0.6	0.2	0.1
SCG-MA	**14.6**	**45.5**	**84.9**	**37.8**	**102.6**	**88.8**	**57.5**
SCG-R	5.0	9.7	39.8	7.3	16.2	39.4	5.5
KOCG [4]	4.4	5.5	8.8	2.6	4.5	8.7	4.8
SPONGE-k [5]	5.0	15.8	41.5	—	—	—	—
SPONGE-(k+1) [5]	0.8	1.0	1.0	—	—	—	—

- Synthetic:



Conflicting group detection in signed graphs (8/8)

Future work

- Can we improve the approximation guarantee to (3)?

$$\mathbf{u}^{\star} \in \operatorname{argmax}\left\{\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} : \mathbf{x} \in \{-1, 0, q\}^n \backslash \{\mathbf{0}\}\right\}. \tag{3}$$

- Can we design provable algorithm for k-conflicting group detection, (2)?
- What is the fundamental limit of the problem in synthetic model?
- Does our algorithm work well in sparse graphs?

Part II - an improved analysis for a numerical solver

Given a symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathcal{T} \subseteq \mathbb{R}^n$, find any

$$\mathbf{u}^\star \in \operatorname{argmax} \left\{ \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} : \mathbf{x} \in \mathcal{T} \backslash \{\mathbf{0}\} \right\}.$$

Part I - discrete $\mathcal{T}$

Studied a signed graph application.

- $\mathbf{u}^\star$ captures antagonistic patterns
- known to be APX-hard
- a provable approximation algorithm

Part II - $\mathcal{T} = \mathbb{R}^n$

Improved analysis of a numerical solver

- for the multiplicative gap $R(\hat{\mathbf{u}})$
- under memory-limited and pass-limited setting

(Part I) Tzeng et al. "Discovering conflicting groups in signed networks." In Proc. of NeurIPS 2020.
(Part II) Tzeng et al. "Improved analysis of randomized SVD for top-eigenvector approximation." In Proc. of AISTATS 2022.

Improved analysis of Randomized SVD (1/8)

usage → the 2-conflicitng group algo by [3] becomes $\mathcal{O}(R(\hat{\mathbf{u}})^{-1}n^{1/2})$-approx

Our metric of interest: multiplicative gap $R(\cdot)$

Given a symmetric $\mathbf{A} \in \mathbb{R}^{n\times n}$ with the largest eigenpair $(\lambda_1, \mathbf{u}_1)$, $\lambda_1 > 0$, define

$$R(\hat{\mathbf{u}}) = \lambda_1^{-1}\frac{\hat{\mathbf{u}}^T\mathbf{A}\hat{\mathbf{u}}}{\hat{\mathbf{u}}^T\hat{\mathbf{u}}}$$

Prior: $\tilde{\mathcal{O}}(n)$-space numerical solvers

For any $\mathbf{A} \succcurlyeq 0$, w.h.p.

no guarantee | $R(\hat{\mathbf{u}}) = \Omega(1)$ → q

$o(\ln n)$-pass | $\Omega(\ln n)$-pass

state-of-the-art: for any $\mathbf{A} \succcurlyeq 0$,

Randomized SVD [6]: $R(\hat{\mathbf{u}}) \geq 1 - \mathcal{O}(\ln n/q)$

Block Krylov [8]: $R(\hat{\mathbf{u}}) \geq 1 - \mathcal{O}((\ln n/q)^2)$

[10] $R(\hat{\mathbf{u}}) = \Theta(1)$ impossible unless $q = \Omega(\ln n)$

Our analysis of Randomized SVD

For any $\mathbf{A} \succcurlyeq 0$, w.h.p.

$R(\hat{\mathbf{u}}) = \Omega\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2q+1}}\right)$ | $R(\hat{\mathbf{u}}) = \Omega(1)$ → q

$o(\ln n)$-pass | $\Omega(\ln n)$-pass

For some indefinite $\mathbf{A}$, w.h.p.

$R(\hat{\mathbf{u}}) = \Omega\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2q+1}}\right)$ | $R(\hat{\mathbf{u}}) = \Omega(1)$ → q

$o(\ln n)$-pass | $\Omega(\ln n)$-pass

Improved analysis of Randomized SVD (2/8)

$(q+1)$-pass $\mathcal{O}(nd)$-space

Intepreting Randomized SVD [6] for top-eigenvector approximation

Algorithm: RSVD $(\mathbf{A}, q, d)$

1. $\mathbf{Y} \leftarrow \mathbf{A}^q \mathbf{S}$ where $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$;
2. $\mathbf{Y} = \mathbf{QR}$;
3. $\mathbf{B} \leftarrow \mathbf{Q}^T \mathbf{AQ}$;
4. $\hat{\mathbf{u}} = \mathbf{Q}\, \mathbf{u}_1(\mathbf{B})$;
5. return $\hat{\mathbf{u}}$;

Fact: $\mathbb{E}[(\mathbf{u}_1^T \mathbf{S}_{:,j})^2] = \cdots = \mathbb{E}[(\mathbf{u}_n^T \mathbf{S}_{:,j})^2]$

(Step 1: line 1)

$\mathbf{Y}_{:,j} = \mathbf{A}^q \mathbf{S}_{:,j} = \sum_{i=1}^{n} \lambda_i^q (\mathbf{u}_i^T \mathbf{S}_{:,j}) \mathbf{u}_i, \ \forall j \in [d]$

(Step 2: line 2-4)

$\hat{\mathbf{u}} = \text{argmax}\{\mathbf{v}^T \mathbf{A} \mathbf{v} : \mathbf{v} \in \text{range}(\mathbf{Y}) \cap \mathbb{S}^{n-1}\}$

Effect of $q \uparrow$:

$\mathbf{Y}_{:,j}$ align more to eigenspace of λ_1

Effect of $d \uparrow$:

(i) $\uparrow$ the concentration around $\mathbb{E}[R(Y_{:,j})]$

(ii) $\uparrow \mathbb{E}[R(\hat{\mathbf{u}})]$ (we make this explicit)

Recall: $R(\hat{\mathbf{u}}) = \lambda_1^{-1} \frac{\hat{\mathbf{u}}^T \mathbf{A} \hat{\mathbf{u}}}{\hat{\mathbf{u}}^T \hat{\mathbf{u}}}$

Improved analysis of Randomized SVD (3/8)

$\geq e^{-\mathcal{O}(\frac{\ln n}{2q+1})} \geq 1 - \mathcal{O}(\frac{\ln n}{q})$

(Theorem 1) For $\mathbf{A} \succcurlyeq 0$, $R(\hat{\mathbf{u}}) = \left(\Omega\left(\frac{d}{n}\right)\right)^{\frac{1}{2q+1}}$ w.p. at least $1 - e^{-\Omega(d)}$.

Our technique: a reduction to random projection length

$$R(\hat{\mathbf{u}}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\sum_{i \in [n]} \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i \in [n]} \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2},$$

where $\alpha_i = \frac{\lambda_i}{\lambda_1}, \forall i \in [n]$.

$$\Rightarrow \cos^2 \theta(\mathbf{u}_1, \mathbf{S}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{a} \rangle^2}{\sum_{i \in [n]} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}$$

(Lemma [13]) For $\mathbf{v} \in \mathbb{S}^{n-1}$ and $d \ll n$, w.p. at least $1 - e^{-\Omega(d)}$,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right).$$

Improved analysis of Randomized SVD (4/8)

(Theorem 2) $\exists \mathbf{A} \succcurlyeq 0$ such that $R(\hat{\mathbf{u}}) = \mathcal{O}\left(\left(\frac{d}{n}\right)^{\frac{1}{2q+1}}\right)$ w.p. at least $1 - e^{-\Omega(d)}$.

(Theorem 3) For $\mathbf{A} \succcurlyeq 0$ with (i_0, γ)-power-law decay, $i_0 \in [n]$ and $\gamma > 1/2q$,

$$R(\hat{\mathbf{u}}) = \Omega\left(\left(\frac{d}{d+i_0}\right)^{\frac{1}{2q+1}}\right) \text{ w.p. at least } 1 - e^{-\Omega(d)}.$$

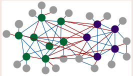
(Assumption 1) $\exists \kappa \in (0, 1]$ such that $\sum_{i=2}^{n} \lambda_i^{2q+1} \geq \kappa \sum_{i=2}^{n} |\lambda_i|^{2q+1}$.

(Theorem 4) For $\mathbf{A}$ with (i_0, γ)-power-law decay, $i_0 \in [n]$ and $\gamma > 1/2q$, and satisfying Assumption 1, there exists a constant $c_\kappa > 0$ such that

$$R(\hat{\mathbf{u}}) = \Omega\left(c_\kappa \left(\frac{d}{d+i_0}\right)^{\frac{1}{2q+1}}\right) \text{ w.p. at least } 1 - e^{-\Omega(\sqrt{d}\kappa^2)}.$$

Improved analysis of Randomized SVD (5/8)

e.g., $|S_1| + |S_2| = \Theta(n)$



Exploiting prior knowledge of large $\langle \mathbf{u}_1, \mathbf{1} \rangle^2$

Question: If $\langle \mathbf{u}_1, \mathbf{1} \rangle^2 = \Theta(n)$, is there a better choice of $\mathbf{S}$ other than $\mathcal{N}(0,1)^{n \times d}$?

Algorithm: RSVD($\mathbf{A}, q, d$)

$\mathbf{Y} \leftarrow \mathbf{A}^q \mathbf{S}$ where $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$;
$\mathbf{Y} = \mathbf{Q}\mathbf{R}$;
$\mathbf{B} \leftarrow \mathbf{Q}^T \mathbf{A} \mathbf{Q}$;
$\hat{\mathbf{u}} = \mathbf{Q}\, \mathbf{u}_1(\mathbf{B})$;
return $\hat{\mathbf{u}}$;

(Hint: $\mathbf{Y}_{:,j} = \mathbf{A}^q \mathbf{S}_{:,j} = \sum_{i=1}^{n} \lambda_i^q (\mathbf{u}_i^T \mathbf{S}_{:,j}) \mathbf{u}_i, \ \forall j \in [d]$)

Improved analysis of Randomized SVD (6/8)

Algorithm: RandSum($\mathbf{A}, q, d, p$)

1. $\mathbf{S}_1 \sim \mathcal{N}(0,1)^{n \times \lceil \frac{d}{2} \rceil}$, $\mathbf{S}_2 \sim \text{Bernoulli}(p)^{n \times \lfloor \frac{d}{2} \rfloor}$;
2. $\mathbf{S} \leftarrow [\mathbf{S}_1 \quad \mathbf{S}_2]$;
3. return RSVD($\mathbf{A}$,$\mathbf{S}$,q,d);

(Theorem 5) For $\mathbf{A} \succcurlyeq 0$, RandSum($\mathbf{A}$,$q$,$d$,$p$) returns $\hat{\mathbf{u}}$ satisfying

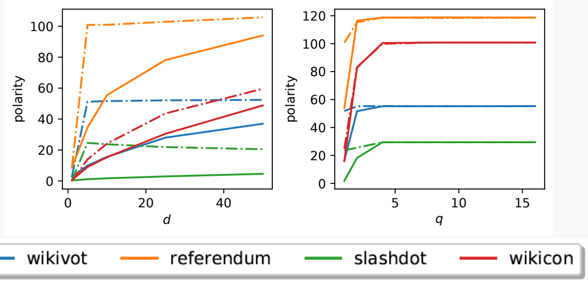
$$R(\hat{\mathbf{u}}) = \left(\Omega\left(\frac{\max\left\{d, \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2\right\}}{n}\right)\right)^{\frac{1}{2q+1}} \quad \text{with prob.} \quad \geq 1 - e^{-\Omega(d)}.$$

Theorem 5 generalizes to indefinite $\mathbf{A}$ under an assumption similar to **Assumption 1**.

Improved analysis of Randomized SVD (7/8)

Experiment: 2-conflicting group detection [3, 11]

	WikiVot	Referendum	Slashdot	WikiCon
$\lvert V \rvert$	7 115	10 884	82 140	116 717
$\lvert E \rvert$	100 693	251 406	500 481	2 026 646
(γ, i_0)	(4.6, 15)	(4.5, 16)	(5.3, 17)	(2.8, 22)
κ	0.397	0.620	0.204	0.034
$\cos\theta(\mathbf{u}_1, \mathbf{1}_n)$	0.378	0.399	0.194	0.193



- RSVD: solid line
- RandSum: dashed line

Improved analysis of Randomized SVD (8/8)

Future work

- Do the results generalize to (row/column)-stochastic matrices?
- Do the results of RandSum hold for any non-centered subgaussian distributions?
- Can we extend the analysis to top-k eigenvectors approximations?
- What is the fundamental limit of $R(\hat{\mathbf{u}})$ for any q-pass $\tilde{\mathcal{O}}(n)$-space algorithm?
- Can we reduce the space complexity while keeping the same guarantees?

Summary

Part I - k-conflicting group detection

- We formulate the problem of k-conflicting group detection.
- We propose an algorithm that sequentially solves a sub-problem which generalizes the problem considered by Bonchi et al. [3] and Bhaskara et al. [2].
- We demonstrate the effectness of our algorithm.

Part II - improved analysis of Randomized SVD

- We improve the analysis of RSVD, in the regime of $o(\ln n)$ passes, and give the first analysis of $R(\cdot)$ for indefinite matrices.
- We study the property of Bernoulli random projection and demonstrate its usefulness to the task of conflicting group detection [3, 11].

Reference I

Aris Anagnostopoulos, Luca Becchetti, Adriano Fazzone, Cristina Menghini, and Chris Schwiegelshohn.
Spectral relaxations and fair densest subgraphs.
In *Proc. of CIKM*, 2020.

Aditya Bhaskara, Moses Charikar, Rajsekar Manokaran, and Aravindan Vijayaraghavan.
On quadratic programming with a ratio objective.
In *Proc. of ICALP*, 2012.

Francesco Bonchi, Edoardo Galimberti, Aristides Gionis, Bruno Ordozgoiti, and Giancarlo Ruffo.
Discovering polarized communities in signed networks.
In *Proc. of CIKM*, 2019.

Reference II

Lingyang Chu, Zhefeng Wang, Jian Pei, Jiannan Wang, Zijin Zhao, and Enhong Chen.
Finding gangs in war from signed networks.
In *Proc. of KDD*, 2016.

Mihai Cucuringu, Peter Davies, Aldo Glielmo, and Hemant Tyagi.
Sponge: A generalized eigenproblem for clustering signed networks.
In *Proc. of AISTATS*, 2019.

Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp.
Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions.
SIAM review, 2011.

Ian T Jolliffe.
Principal components in regression analysis.
In *Principal component analysis*. Springer, 1986.

Reference III

Cameron Musco and Christopher Musco.
Randomized block krylov methods for stronger and faster approximate singular value decomposition.
In *Proc. of NeurIPS*, 2015.

Mark EJ Newman.
Modularity and community structure in networks.
Proc. of NAS, 2006.

Max Simchowitz, Ahmed El Alaoui, and Benjamin Recht.
Tight query complexity lower bounds for pca via finite sample deformed wigner law.
In *Proc. of STOC*, 2018.

Ruo-Chun Tzeng, Bruno Ordozgoiti, and Aristides Gionis.
Discovering conflicting groups in signed networks.
In *Proc. of NeurIPS*, 2020.

Reference IV

- Ruo-Chun Tzeng, Po-An Wang, Florian Adriaens, Aristides Gionis, and Chi-Jen Lu.
  Improved analysis of randomized svd for top-eigenvector approximation.
  arXiv preprint arXiv:2202.07992, 2022.

- Roman Vershynin.
  High-dimensional probability: An introduction with applications in data science.
  Cambridge university press, 2018.