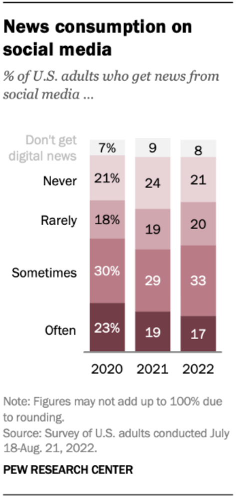
Adversaries with Limited Information in the Friedkin-Johnsen Model

Sijing Tu, Stefan Neumann, Aristides Gionis

KTH Royal Institute of Technology

August 9, 2023

Online Social Networks (OSNs)



- Online social networks (OSNs) have become ubiquitous parts of modern societies
- People use OSNs to stay in touch with their friends, to read news and discuss societal issues
- Some concerns arise, e.g.,
  - State actors use bots to adversarially attack societies and to increase social discord

Malicious Actors are Attacking OSNs

(The New York Times, 2016)

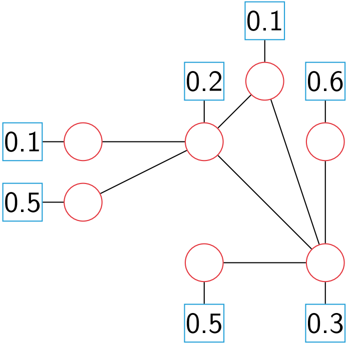
- 2016 Democratic National Committee email leak.
  - *Russian military and intelligence services have been using the Internet to sow discord and discredit legitimate political institutions.* (TIME, 2016)

Our Goal

Achieving a theoretical understanding of the consequence of malicious actors

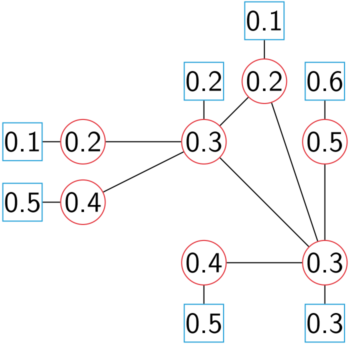
- Model social networks as graphs
- We formulate the *intervention* of the adversary as an *optimization problem*
  - *Intervention*: the effects after conducting some activities in the network:
    - Spreading fake news $\Rightarrow$ changing people' opinions
  - Objective function encodes the desired goal
  - Constraints encode the power of the *intervention*
- We consider the *opinion formation model* as an abstraction of users' opinions
  - DeGroot Model, Bounded Confidence Model, Friedkin-Johnsen Model...
  - The societal discord is measured by the opinions

A Glimpse on Friedkin-Johnsen Model



- The graph $G = (V, E, w)$, $\mathbf{L}$ is the graph Laplacian
- Nodes with their *fixed, private* innate opinions
- $s_u \in [0, 1]$

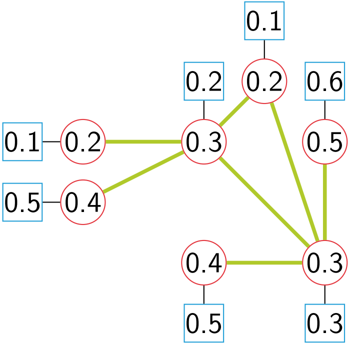
A Glimpse on Friedkin-Johnsen Model



- Due to peer pressure, the nodes *share different public* expressed opinions
- $z_u^t \in [0,1]$ changes on time t
- $z_u^{(t)} = \frac{s_u + \sum_{v \in N(u)} w_{uv} z_v^{(t-1)}}{1 + \sum_{v \in N(u)} w_{uv}}$
- At the equilibrium state $\mathbf{z}^* = (\mathbf{I} + \mathbf{L})^{-1}\mathbf{s}$

Measure Societal Discord by Users' Opinions

For instance, disagreement, polarization



- The disagreement measures the differences between expressed opinions

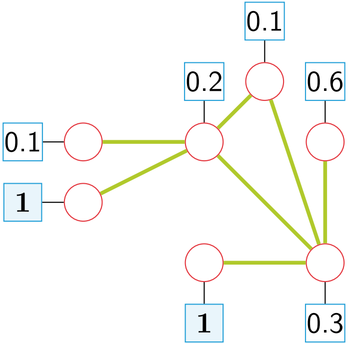
$$\mathcal{D}_{G,\mathbf{s}} = \sum_{(u,v)\in E} w_{u,v}(z_u^* - z_v^*)^2 = \mathbf{s}^\intercal \mathcal{D}(\mathbf{L})\,\mathbf{s}$$

- $\mathcal{D}(\mathbf{L}) = (\mathbf{L}+\mathbf{I})^{-1}\mathbf{L}(\mathbf{L}+\mathbf{I})^{-1}$
- Disagreement is determined by *network structure* and *innate opinions*.

How Much Disagreement Can Malicious Actors Sow on Online Social Networks

Maximize Disagreement with Full Information

Chen, Racz (TNSE'21) and Gaitonde, Kleinberg, Tardos (EC'20)



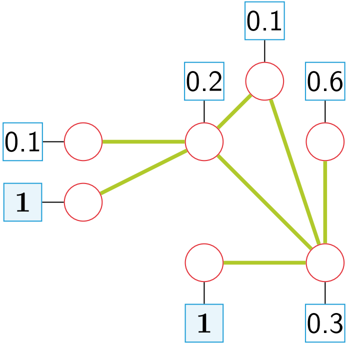
Problem (Full-information)

Maximize disagreement *by radicalizing k users' innate opinions, given the network structure and innate opinions.*

$$\max_{\mathbf{s}} \quad \mathbf{s}^{\intercal}\mathcal{D}(\mathbf{L})\,\mathbf{s},$$
$$s.t. \quad \|\mathbf{s}-\mathbf{s}_0\|_0 = k, \text{ and}$$
$$\mathbf{s}(u) \in \{\mathbf{s}_0(u), 1\} \text{ for all } u \in V.$$

Maximize Disagreement with Full Information

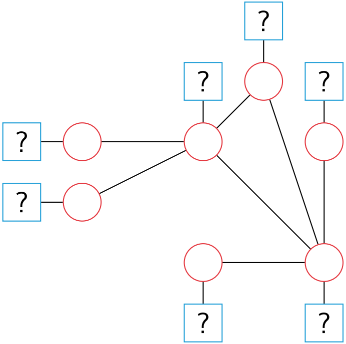
Chen, Racz (TNSE'21) and Gaitonde, Kleinberg, Tardos (EC'20)



- They show that disagreement increases by at most $8d_{\max}k$
- They conducted some heuristics, like greedy algorithm and changing the opinions of centrists
- Adversaries in the full-information setting are quite powerful

Our Paper:
Maximize Disagreement with Limited Information

Maximize Disagreement with Limited Information



True innate opinions $\mathbf{s}_0$ are very hard to obtain

- It still knows the network structure
- It still has the power to radicalize opinions

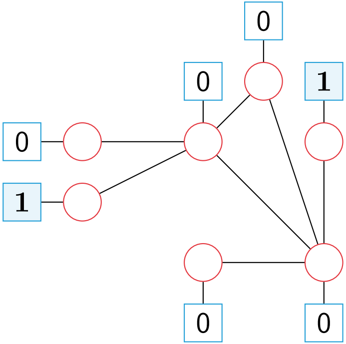
$$\max_{\mathbf{s}} \quad \mathbf{s}^{\intercal} \mathcal{D}(\mathbf{L})\, \mathbf{s},$$

$$s.t. \quad \|\mathbf{s} - ?\|_0 = k, \text{ and}$$

$$\mathbf{s}(u) \in \{?, 1\} \; \forall \; u \in V.$$

What strategy can the malicious actor apply?

Maximize Disagreement with Limited Information



Problem (Limited-information)

Maximize disagreement by radicalizing k users' innate opinions, given **only** *the network structure.*

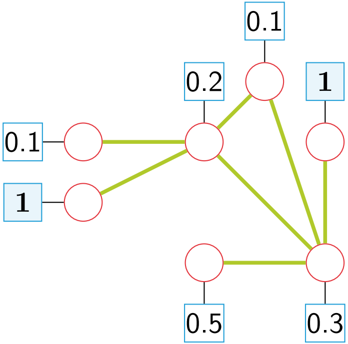
$$\max_{\mathbf{s}} \quad \mathbf{s}^{\intercal}\mathcal{D}(\mathbf{L})\,\mathbf{s},$$

$$s.t. \quad \|\mathbf{s}-\mathbf{0}\|_0 = k, \ \textit{and}$$

$$\mathbf{s} \in \{0,1\}^n.$$

It might select different nodes to radicalize.

Maximize Disagreement with Limited Information



Theorem (informal)

Assume that the initial innate opinions have small variance (achieved by technical assumptions):

- *$\mathcal{O}(1)$-approximation algorithm to Limited-information problem $\Rightarrow$ $\mathcal{O}(1)$-approximation solution Full-information problem.*
- *Adversaries with limited information are almost as powerful as with full information*

Limited-information Problem: Cardinality-Constrained Max-Cut Variant

On *graph* with positive and **negative** edge weights

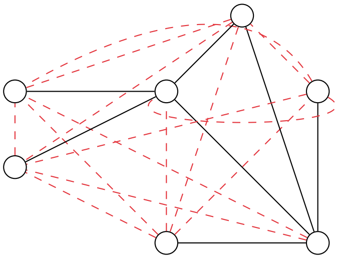


Figure: The graph Laplacian of the graph is $\mathcal{D}(\mathbf{L})$. Negative weights edges are in red.

Regard the disagreement matrix $\mathcal{D}(\mathbf{L})$ as a graph Laplacian

- This problem is $\mathbf{NP}$-hard and no longer submodular
- We apply Semidefinite Program (SDP) Relaxation and and hyperplane rounding to get a initial *cut*
- We bound the loss on *cut* on each greedy move

Theoretical results on Limited-information Problem

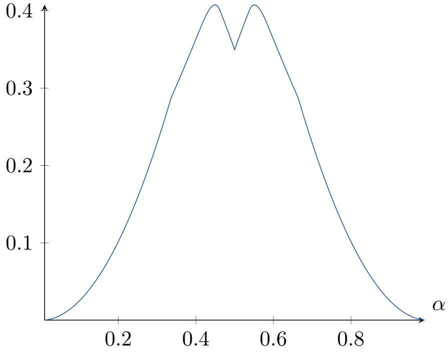


Figure: Approximation Ratio. $k = \alpha n$.

Theorem (informal)

If $k \in \Omega(n)$, there exists a randomized $\Omega(1)$-approximation algorithm for the limited-information problem that succeeds with high probability.

- We set $k \in \Omega(n)$ due to technical difficulties

Experiments

The Factor that Influences the Performance

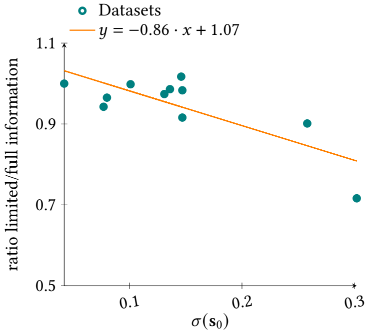
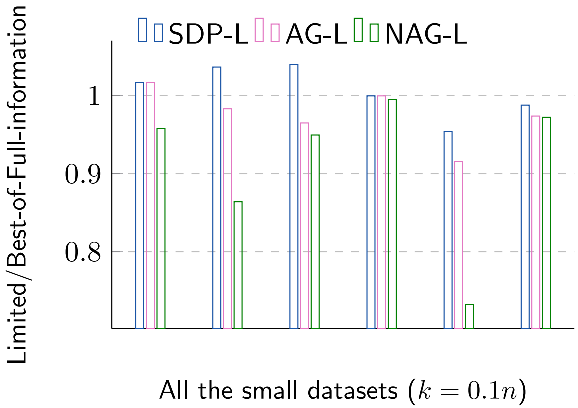


Figure: Standard deviation of opinions, $R^2 = 0.62$

Strong relationship between initial standard deviation of the innate opinions and performance of limited-information methods

SDP-L is the Best Among Limited-information Algorithms



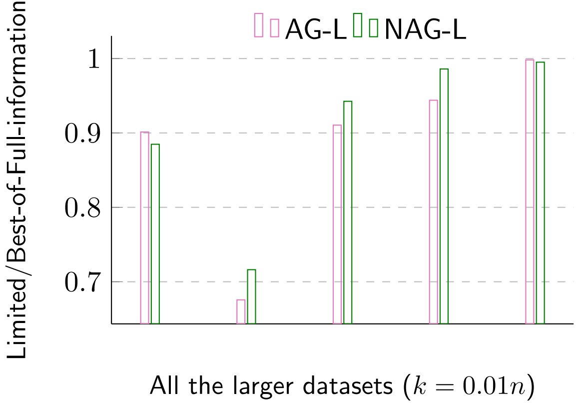
Limited-information algorithm:

SDP-L: SDP-based

AG-L: Adaptive-Greedy

NAG-L: NonAdaptive-Greedy

Limited-information is at Most a Factor of 1.4 Worse



Limited-information algorithm:

AG-L: Adaptive-Greedy

NAG-L: NonAdaptive-Greedy

Summary

- Our paper studies the adversary's potential to increase societal discord
- We formally prove that this adversary with limited information is almost the same powerful as the adversary with full information
- We propose a constant approximation ratio algorithm for the problem under the limited information when $k = \Omega(n)$
- We evaluate our algorithms in real-world datasets

Open Questions

- Can we design a faster algorithm?
- The graph with positive and negative edge weights is extremely dense, can we find a sparser substitute of it?
- Can we design an approximation algorithm when $k = o(n)$?
- Is it possible to verify our algorithm in real world?

Adversaries with Limited Information in the Friedkin–Johnsen Model

Sijing Tu*[1], Stefan Neumann†[1], and Aristides Gionis‡[1]

[1]KTH Royal Institute of Technology, Stockholm, Sweden

Appendix: How to Solve the Limited-information Problem?

Limited-information Problem is a MaxCut Variant

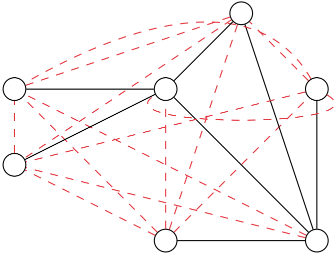
Observe (1) $\mathcal{D}(\mathbf{L})\,\mathbf{1} = \mathbf{0}$; and (2) $\mathcal{D}(\mathbf{L})$ is positive semidefinite.

$\mathcal{D}(\mathbf{L}) = D' - W'$, where $W'_{ii} = 0$ and
$D'_{ii} = \sum_j W'_{ij}$.
$\mathbf{s}^\intercal \mathcal{D}(\mathbf{L})\,\mathbf{s} = \frac{1}{2}\sum_{i,j} W'_{i,j}(s_i - s_j)^2$.

$$\max_{\mathbf{s}} \quad \mathbf{s}^\intercal \mathcal{D}(\mathbf{L})\,\mathbf{s},$$
$$s.t. \quad \|\mathbf{s}\|_0 = k, \text{ and}$$
$$\mathbf{s} \in [0,1]^n.$$

Limited-information Problem is a MaxCut Variant

$\mathcal{D}(\mathbf{L}) = D' - W'$, and we treat W' as a new weighted adjacency matrix on $G' = (V, W')$.



$$\max_{\mathbf{s}} \quad \mathbf{s}^\intercal \mathcal{D}(\mathbf{L})\,\mathbf{s},$$
$$s.t. \quad \|\mathbf{s}\|_0 = k, \text{ and}$$
$$\mathbf{s} \in [0,1]^n.$$

Limited-information Problem is a MaxCut Variant

How to solve the MaxCut with k nodes on one side on the G'? ($\mathbf{NP}$-hard, reduction from MaxCut)

- Applying Linear relaxation:
  $\frac{1}{2}$-approximation algorithm for all k.
- Applying Semidefinite relaxation:
  > 0.63-approximation algorithm,
  when $k = \frac{1}{2}n$.
- Applying SOS hierarchy:
  > 0.85-approximation algorithm,
  when $k = \Omega(n)$.

$$\max_{\mathbf{s}} \quad \mathbf{s}^\intercal \mathcal{D}(\mathbf{L})\, \mathbf{s},$$
$$s.t. \quad \|\mathbf{s}\|_0 = k, \text{ and}$$
$$\mathbf{s} \in [0,1]^n.$$

Solve Limited-information Problem with Semidefinite Programming

$$\max_{\mathbf{v}_1,\ldots,\mathbf{v}_n} \quad \frac{1}{4}\sum_{ij}\mathcal{D}(\mathbf{L})_{ij}\mathbf{v}_i^\intercal\mathbf{v}_j,$$

$$s.t. \quad \sum_{i<j}\mathbf{v}_i^\intercal\mathbf{v}_j = \frac{1}{2}n^2(1-2\alpha)^2 - \frac{n}{2},$$

$$\mathbf{v}_i \in \mathbb{R}^n, \qquad \|\mathbf{v}_i\|_2 = 1.$$

$$\max_{\mathbf{s}} \quad \mathbf{s}^\intercal\mathcal{D}(\mathbf{L})\,\mathbf{s},$$

$$s.t. \quad \|\mathbf{s}\|_0 = k,$$

$$\mathbf{s} \in [0,1]^n.$$

The scale does not influence approximation ratio.

Solve Limited-information Problem with Semidefinite Programming

$$\max_{\mathbf{v}_1,\dots,\mathbf{v}_n} \quad \frac{1}{4}\sum_{ij} \mathcal{D}(\mathbf{L})_{ij}\mathbf{v}_i^\intercal \mathbf{v}_j,$$

$$s.t. \quad \sum_{i<j} \mathbf{v}_i^\intercal \mathbf{v}_j = \frac{1}{2}n^2(1-2\alpha)^2 - \frac{n}{2},$$

$$\mathbf{v}_i \in \mathbb{R}^n, \qquad \|\mathbf{v}_i\|_2 = 1.$$

Algorithm 1: SDP-based algorithm

Solve the SDP, obtain $\mathbf{v}_1,\dots,\mathbf{v}_n$;

for $T = 1,\dots,\mathcal{O}(1/\epsilon\log(1/\epsilon))$ **do**

- Sample vector $\mathbf{r}$ with each entry $\sim \mathcal{N}(0,1)$;
- Set $S = \{i : \langle \mathbf{v}_i, \mathbf{r}\rangle \geq 0\}$ and $\bar{S} = V \setminus S$;
- **if** $|S| \neq \alpha n$ **then**
  - greedily move elements from S $(\bar{S})$ to $\bar{S}$ (S) until $|S| = \alpha n$

return *best over* T *trials*;

Solve Limited-information Problem with Semidefinite Programming

$$\max_{\mathbf{v}_1,\dots,\mathbf{v}_n} \quad \frac{1}{4}\sum_{ij}\mathcal{D}(\mathbf{L})_{ij}\mathbf{v}_i^\intercal\mathbf{v}_j,$$

$$s.t. \quad \sum_{i<j}\mathbf{v}_i^\intercal\mathbf{v}_j = \frac{1}{2}n^2(1-2\alpha)^2 - \frac{n}{2},$$

$$\mathbf{v}_i \in \mathbb{R}^n, \qquad \|\mathbf{v}_i\|_2 = 1.$$

Let M^* be the optimal solution of SDP. Let cut(S) be the cut if S consists of one partition, after hyperplane rounding.

- $\mathbb{E}[\frac{\text{cut}(S)}{M^*}] \geq \frac{\pi}{2}$.
- $\mathbb{E}[|S|\,|\bar{S}|] \geq 0.878n^2(1-\alpha)\alpha$.
- $\Rightarrow$ Markov inequality: $\exists S$ $\frac{\text{cut}(S)}{M^*} + \frac{|S|\,|\bar{S}|}{n^2\alpha(1-\alpha)} \geq (1-\epsilon)(\frac{\pi}{2}+0.878)$.

Solve Limited-information Problem with Semidefinite Programming

$$\max_{\mathbf{v}_1,\dots,\mathbf{v}_n} \quad \frac{1}{4}\sum_{ij} \mathcal{D}(\mathbf{L})_{ij}\mathbf{v}_i^\intercal \mathbf{v}_j,$$

$$s.t. \quad \sum_{i<j} \mathbf{v}_i^\intercal \mathbf{v}_j = \frac{1}{2}n^2(1-2\alpha)^2 - \frac{n}{2},$$

$$\mathbf{v}_i \in \mathbb{R}^n, \qquad \|\mathbf{v}_i\|_2 = 1.$$

Let M^* be the optimal solution of SDP. Let cut(S) be the cut if S consists of one partition, after hyperplane rounding. Let S' be the set of nodes in one partition.

- Markov inequality: $\exists S$ $\frac{\text{cut}(S)}{M^*} + \frac{|S|\left|\bar{S}\right|}{n^2\alpha(1-\alpha)} \geq (1-\epsilon)(\frac{\pi}{2}+0.878)$.
- Moving one node from S' to $\bar{S}'$ loses $\frac{2\text{cut}(S')}{|S'|}$.

Solve Limited-information Problem with Semidefinite Programming

$$\max_{\mathbf{v}_1,\ldots,\mathbf{v}_n} \quad \frac{1}{4}\sum_{ij} \mathcal{D}(\mathbf{L})_{ij} \mathbf{v}_i^\intercal \mathbf{v}_j,$$

$$s.t. \quad \sum_{i<j} \mathbf{v}_i^\intercal \mathbf{v}_j = \frac{1}{2} n^2 (1-2\alpha)^2 - \frac{n}{2},$$

$$\mathbf{v}_i \in \mathbb{R}^n, \qquad \|\mathbf{v}_i\|_2 = 1.$$

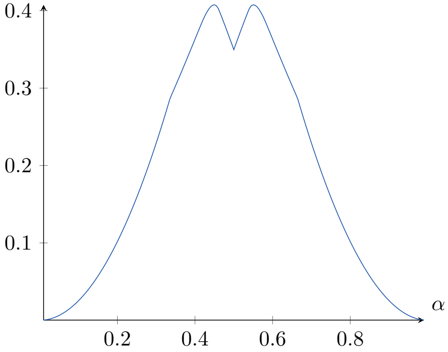


Figure: Approximation Ratio. $k = \alpha n$.

Interventions as Optimization Problems

Objective function encodes the desired goal; Constraints encode the power of the intervention.

- Musco, Musco, Tsourakakis (WebConf'18):
  - Suppose we can change the network structure such that the sum of degree keeps the same, then minimizing the sum of disagreement and polarization is a convex optimization problem.
- Chitra, Musco (WSDM'20):
  - If OSN providers repeatedly change the network structure to reduce disagreement, this will increase the polarization.
- Tu, Neumann (WebConf'22):
  - Model for simulating how viral content in OSNs impacts user opinions, and increase polarization.