A Viral Marketing-Based Model For Opinion Dynamics in Online Social Networks

SIJING TU · STEFAN NEUMANN
{sijing, neum}@kth.se, KTH Royal Institute of Technology, Sweden

MOTIVATION

Phenomenon: Information spreading has been taken advantage to influnece people's opinions, for example,

- fake news about vaccination, and,
- state bots influencing the election.

Opinion-Based Network Indices which measures, for example, *average Opinion*, *polarization*, *disagreement*, and *sum of disagreement and controversy* in the social network.

Research Questions:

Information Spreading → How Much Influence? → Opinion Formation

- We quantify influence in terms of estimating and optimizing network indices.

PROBLEM SETTING

Problem Setting: The problem sets on a weighted undirected graph $G = (V, E, w)$. $\mathbf{L}$ is the graph laplacian, $\mathbf{D}_{u,u}$ is the sum of incident edge weights.

- Each user u has an *expressed* opinion $z_u \in [0, 1]$, which depends on the network, and a fixed *innate* opinion $s_u \in [0, 1]$.
- Each user turns into one of the three status upon exposing viral content: spread, acknowledge, and ignore.
- Two campaign contents: marketing, and polarizing.

Information Spreading: Independent Cascade Model [5].

- Each edge (u, v) is assigned with an influence probability p_{uv}, and a parameter δ indicating the tendency to share, seed nodes are inital spread nodes.
- At each time step, each spread node u gets one shot at influencing its non-spread neighbor v.
  - If v is ignore or acknowledge, with probability δp_{uv}, v switches to spread;
  - If v is ignore, with probability $(1 - \delta)p_{uv}$, v switches to acknowledge.

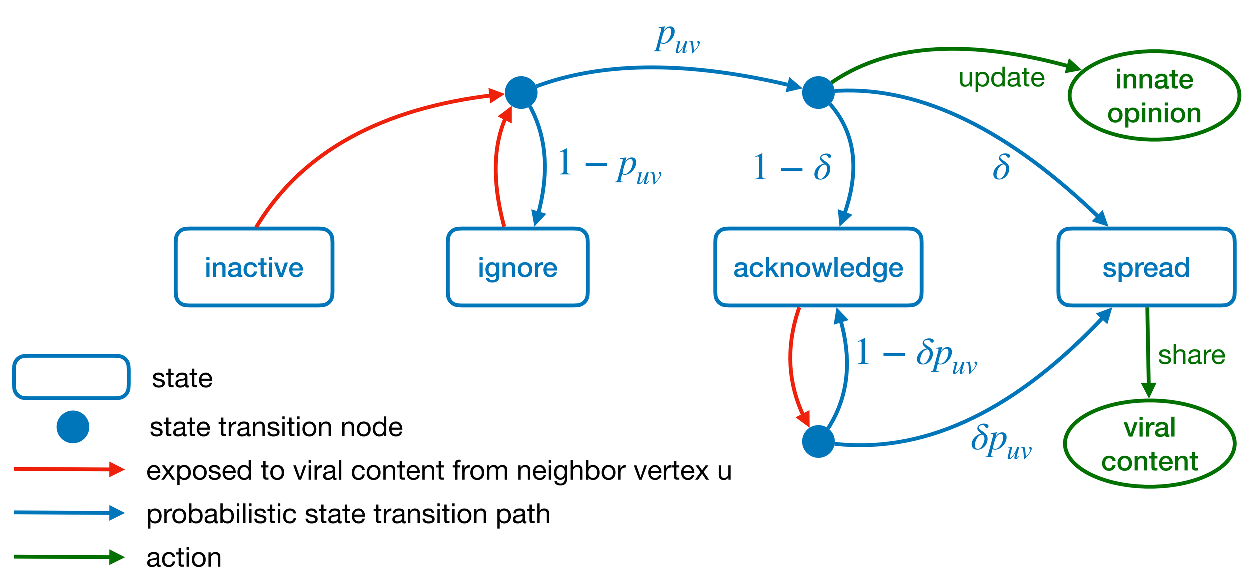
Updating Innate Opinions: Given parameter ϵ, once a node first switches to spread or acknowledge status.

- marketing content: $\hat{\mathbf{s}}_u = \min\{\mathbf{s}_u + \epsilon, 1\}$;
- polarizing content: If $\mathbf{s}_u \geq \tau$, $\hat{\mathbf{s}}_u = \max\{\mathbf{s}_u - \epsilon, 0\}$, then embrace,$\hat{\mathbf{s}}_u = \min\{\mathbf{s}_u + \epsilon, 1\}$; If $\mathbf{s}_u < \tau$, then repel, adjusts $\hat{\mathbf{s}}_u = \max\{0, \mathbf{s}_u - \epsilon\}$.

Updating Expressed Opinions: Friedkin-Johnsen Model [3]. $\mathbf{z}^{(t+1)} = (\mathbf{D} + \mathbf{I})^{-1}(\mathbf{W}\mathbf{z}^{(t)} + \mathbf{s})$.

MODEL

Spread-acknowledge Model with respect to state transitioning and actions performed for a single node v. In the initial round, k seed nodes are in state spread, while the rest of nodes are in state inactive.



two-stage model:

- First stage: Performing information spread until no new users have changed their state to spread.
- Second stage: Updating users' expressed opinions.

> **Lemma.** With same seed nodes, **Spread-acknowledge Model** ≡ **two-stage model** on the distribution of innate and expressed opinions.

Network Indices:

Index	Notation	Matrix
Polarization	$\mathcal{P}(\mathbf{L})$	$(\mathbf{I}+\mathbf{L})^{-1}(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\intercal}{n})(\mathbf{I}+\mathbf{L})^{-1}$
Disagreement	$\mathcal{D}(\mathbf{L})$	$(\mathbf{L}+\mathbf{I})^{-1}\mathbf{L}(\mathbf{L}+\mathbf{I})^{-1}$
Disagreement–controversy	$\mathcal{I}^{dc}(\mathbf{L})$	$(\mathbf{L}+\mathbf{I})^{-1}$

Let $\mathcal{M}(\mathbf{L})$ denotes one of the matrices, let $\hat{\mathbf{s}}$ be the final innate opinions. Then $\mathbb{E}[\hat{\mathbf{s}}^\intercal \mathcal{M}(\mathbf{L})\, \hat{\mathbf{s}}]$ is the measure of corresponding index.

EXPERIMENTS

Experimental Results. We report the relative increase of each index in percent.

- Results for Marketing campaigns with $k = \lceil 0.5\% \cdot n \rceil$ seeds.

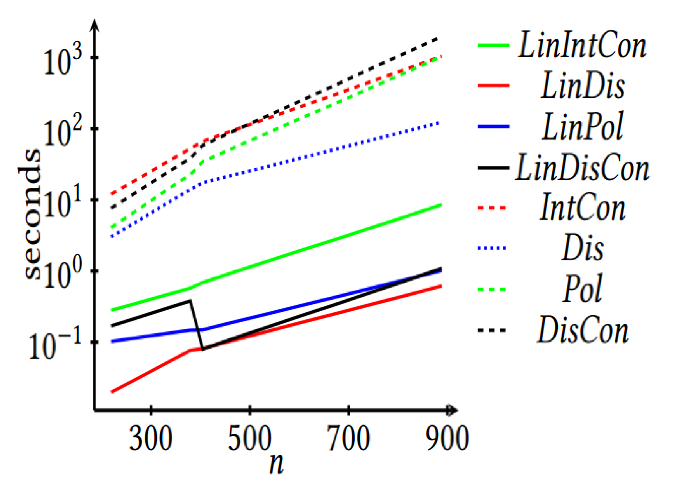
Dataset	Sum Index						Polarization Index						
	Sum	*LinDisCon*	*LinPol*	*MaxInflu*	*Random*	*FJ*	*Sum*	*LinDisCon*	*LinPol*	*MaxInflu*	*Random*	*FJ*	*FJUpp*
Netscience	**2.79**	2.75	0.74	2.78	0.27	0.11	3.15	3.18	**7.54**	3.17	-0.06	2.36	10.54
WikiVote	**4.14**	4.12	0.53	4.11	0.3	0.11	-0.64	-0.61	**3.83**	-0.58	-0.06	2.92	12.29
Brightkite	6.16	6.15	0.72	**6.17**	0.27	-	-0.17	-0.06	**4.27**	-0.24	-0.07	-	-
WikiTalk	9.27	9.27	1.73	**9.28**	0.29	-	-0.82	-0.71	**3.37**	-0.79	-0.09	-	-

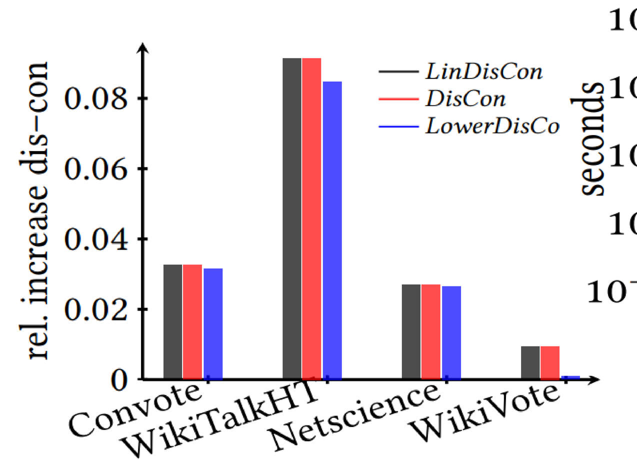
- Results for polarizing campaigns with $k = \lceil 0.5\% \cdot n \rceil$ seeds

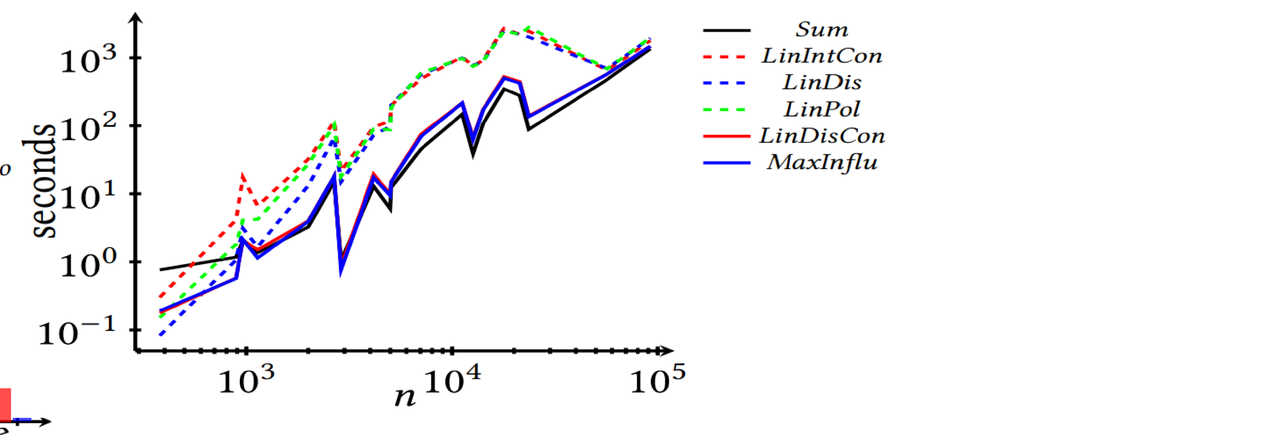
Dataset	Sum Index						Polarization Index						
	Sum	*LinDisCon*	*LinPol*	*MaxInflu*	*Random*	*FJ*	*Sum*	*LinDisCon*	*LinPol*	*MaxInflu*	*Random*	*FJ*	*FJUpp*
Netscience	**0.48**	0.46	0.04	0.34	-0.01	0.11	2.72	5.03	**7.62**	5.66	0.62	2.36	10.54
WikiVote	**0.33**	0.25	-0.28	-0.33	-0.02	0.11	3.14	5.83	**9.46**	9.14	0.64	2.92	12.29
Brightkite	**0.38**	0.24	-0.02	0.01	0.0	-	5.66	13.35	**15.86**	15.58	0.7	-	-
WikiTalk	**0.49**	0.29	0.02	0.02	0.0	-	13.46	25.84	**28.79**	28.57	0.73	-	-

Datasets. We obtain 16 datasets from public repositories: Konect, SNAP, and Network Repository. The size of the data set ranges from $0.2k$ to $92k$ nodes; $0.5k$ to $360k$ edges.

Baselines. *MaxInflu* chooses the seed nodes that maximize the influence; *Random* selects seed nodes uniformly randomly; *FJ* [2] greedily maximizes the graph indices that allowes to change k innate user opinions arbitrarily much, and *FJUpp* [4] is an analytic upper bound.







ESTIMATING AND OPTIMIZING

Our objective is to compute $\mathbb{E}[\hat{\mathbf{s}}^\intercal \mathcal{M}(\mathbf{L})\, \hat{\mathbf{s}}]$. We also compute the sum of expressed opinions, i.e., $\mathbb{E}[\sum \hat{\mathbf{z}}_i]$.

Estimating

- We use Monte Carlo Simulation to estimate $\mathbb{E}[\hat{\mathbf{s}}]$.
- Then we apply fast algorithmr [7] based on Laplacian solver to calculate $\mathbb{E}[\hat{\mathbf{s}}^\intercal \mathcal{M}(\mathbf{L})\, \hat{\mathbf{s}}]$.

Optimizing (Maximizing network indices)

- We apply Sampling Technique based on Reverse Reachable set [1].
  - We derive an $(1 - \frac{1}{e} - \epsilon)$ approximation algrithm for maximizing *Sum*.
  - We derive a data-depdendent approximation algorithm, based on sandwich method.
    - *Dis*, *Pol*, *DisCon*;
  - We also design faster heuristics, and in practice, close to the approximation algorithm. We maximize $\mathbb{E}[2\mathbf{s}^\intercal \mathcal{M}(\mathbf{L})\, \Delta\hat{\mathbf{s}}]$, and evaluate on the objective function.
    - *LinDis*, *LinPol*, *LinDisCon*;

Regarding **sandwich method**. Let
$\mu_0(S) = \mathbb{E}[2\mathbf{s}^\intercal \mathcal{M}(\mathbf{L})\, \Delta\hat{\mathbf{s}} + \Delta\hat{\mathbf{s}}^\intercal \mathcal{M}(\mathbf{L})\, \Delta\hat{\mathbf{s}}]$,
$\mu_L(S) = \mathbb{E}[2\mathbf{s}^\intercal \mathcal{M}(\mathbf{L})\, \Delta\hat{\mathbf{s}}]$,
$\mu_U(S) = \mathbb{E}[2\mathbf{s}^\intercal \mathcal{M}(\mathbf{L})\, \Delta\hat{\mathbf{s}} + \Delta\hat{\mathbf{s}}^\intercal \mathcal{M}(\mathbf{L})^{U}\, \Delta\hat{\mathbf{s}}]$.

> **Theorem. [6]** Let $S^* = \arg\max_{|S| \leq k} \mu_0(S)$. Then $\mu_0(S) \geq \max\left\{\frac{\mu_0(S_U)}{\mu_U(S_U)}, \frac{\mu_L(S^*)}{\mu_0(S^*)}\right\}(1 - \frac{1}{e} - \epsilon)\, \mu_0(S^*)$.

REFERENCES

[1] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier. Maximizing social influence in nearly optimal time. In *SODA*, 2014.
[2] M. Chen and M. Z. Racz. Network disruption: maximizing disagreement and polarization in social networks. 2021.
[3] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 1990.
[4] J. Gaitonde, J. Kleinberg, and E. Tardos. Adversarial perturbations of opinion dynamics in networks. In *EC*, 2020.
[5] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. *Theory OF Computing*, 2015.
[6] W. Lu, W. Chen, and L. V. Lakshmanan. From competition to complementarity: comparative influence diffusion and maximization. *VLDB*, 2015.
[7] W. Xu, Q. Bao, and Z. Zhang. Fast evaluation for relevant quantities of opinion dynamics. In *WWW*, 2021.

ACKNOWLEDGEMENT

We would like to thank Aris Gionis who has been a mentor for the project, and for many valuable advices and suggestions he has provided.