DISCOVERING OPINION INTERVALS FROM CONFLICTS IN SIGNED GRAPHS

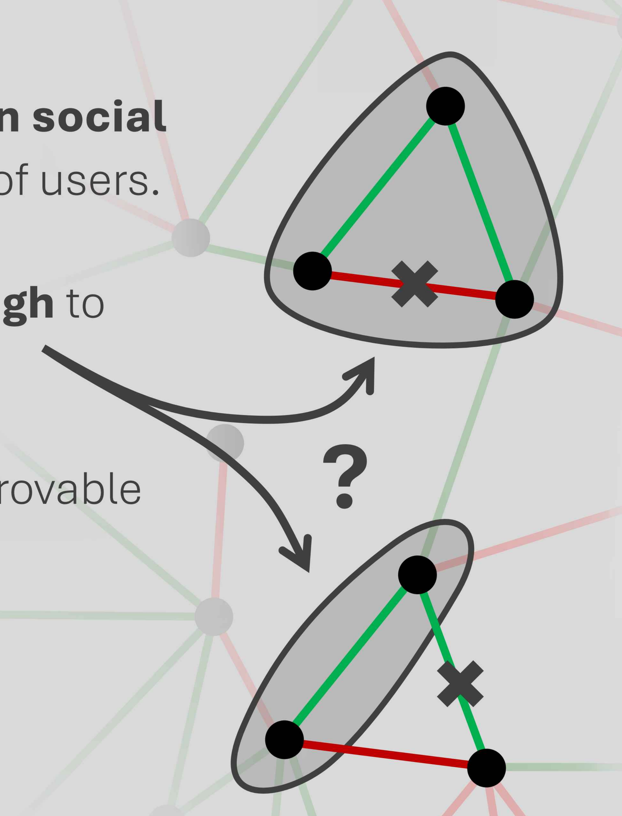
Peter Blohm*, Florian Chen*, Aristides Gionis, Stefan Neumann

WHAT are we doing?

We observe positive and negative **interactions in social networks** and use them to **learn the opinions** of users.

Clustering methods are not expressive enough to model non-transitive relationships.

We **model opinions as intervals** and provide provable and efficient algorithms to find them.



WHY should you care?

Online social networks have become **essential parts of modern societies** with billions of users.

There are a **substantial number of conflicts**, particularly between people with differing viewpoints.

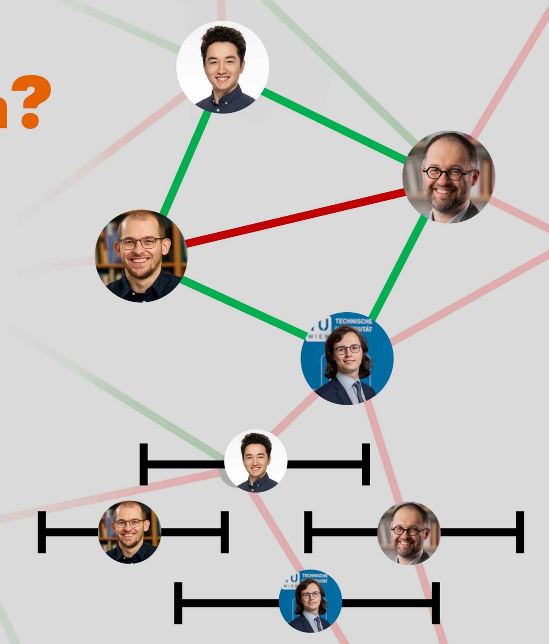
More expressive representations can drastically improve the **information learned from social networks**.

HOW do we address the problem?

We formulate **Best Interval Approximation** and show NP-hardness results.

In the setting of a fixed number of intervals, we provide a **PTAS and efficient heuristics**.

We demonstrate the **expressivity, scalability, and interpretability** of our approach.



Best Interval Approximation

Problem Formulation

Given a signed graph $G = (V, E^+ \cup E^-)$, assign each vertex an interval $\{I_v : v \in V\}$ with $I_v \subset \mathbb{R}$ to maximize the agree$(G, \{I_v : v \in V\})$, which is the number of

(1) positive edges $\{u, v\} \in E^+$ for which I_u and I_v overlap and

(2) negative edges $\{u, v\} \in E^-$ for which I_u and I_v are disjoint:

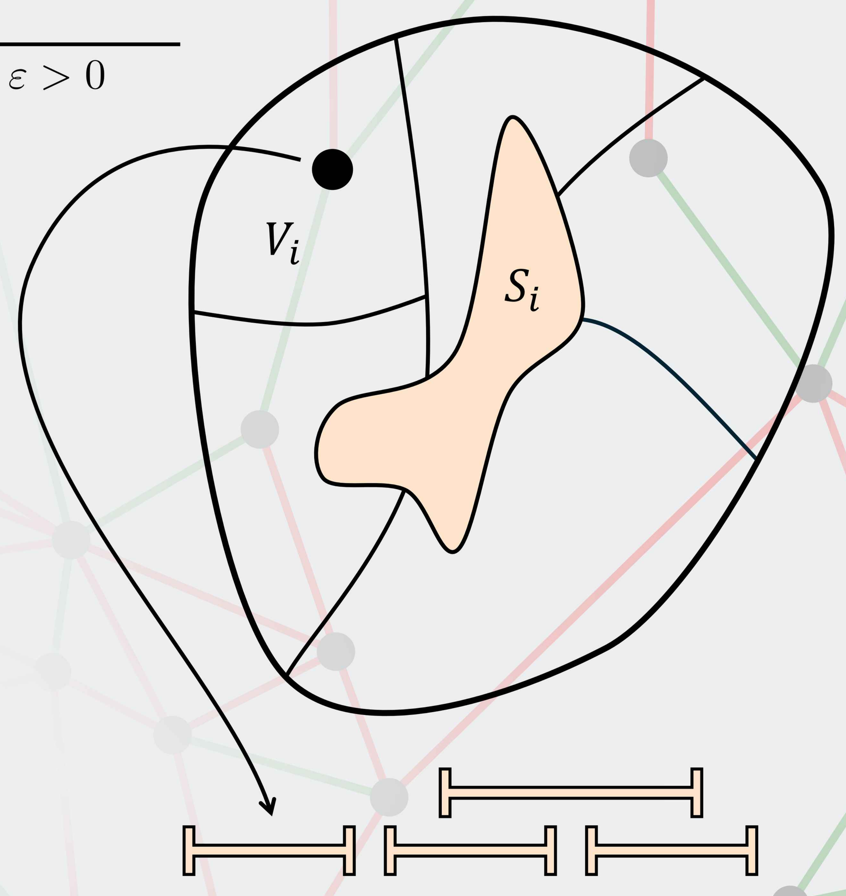
$$\text{agree}(G, \mathcal{I}) = \sum_{\{u,v\} \in E^+} \mathbb{1}(I_u \cap I_v \neq \emptyset) + \sum_{\{u,v\} \in E^-} \mathbb{1}(I_u \cap I_v = \emptyset)$$

Polynomial-Time Approximation Scheme

1+ϵ approximation for a fixed number of intervals

Algorithm 1: PTAS to maximize agreement for fixed k

Input: Complete signed graph $G = (V, E^+ \cup E^-)$, intervals $I_1, \ldots, I_k$, $\varepsilon > 0$
Result: An assignment ALG maximizing the agreement
Partition V into $m = O(\frac{1}{\varepsilon})$ sets $V_1, \ldots, V_m$;
Sample $S_i \subseteq V \setminus V_i$ of size $\tilde{O}(\frac{1}{\epsilon^2})$ for all $i = 1, \ldots, m$;
Initialize some arbitrary assignment ALG;
for *all partial assignments* PART_{S_i} *of all* S_i **do**
  for $i = 1, \ldots, m$ **do**
    Let ALG'_i be an empty assignment of V_i;
    for $u \in V_i$ **do**
      $\ell^* \leftarrow \text{argmax}_{\ell=1,\ldots,k}\ \text{agree}(u, \ell, (\text{PART}_{S_i}))$;
      $\text{ALG}'_i(u) \leftarrow \ell^*$
  Set $\text{ALG}' \leftarrow \bigcup_{i=1}^{m} \text{ALG}'_i$;
  if *agree*$(G, \text{ALG}') >$ *agree*(G, ALG) **then**
    Set ALG $\leftarrow$ ALG$'$;
return ALG;



Heuristics

Practical Improvements for PTAS

1. Do not sample S_i, iterate over **multiple epochs** instead of partial assignments:

for *all partial assignments* PART_{S_i} *of all* S_i **do** ⟶ **for** *each epoch* **do**

2. Reuse **incumbent solution** instead of partial assignments:

$\ell^* \leftarrow \text{argmax}_{\ell=1,\ldots,k}\ \text{agree}(u, \ell, (\text{PART}_{S_i}))$; ⟶ $\ell^* \leftarrow \text{argmax}_{\ell=1,\ldots,k}\ \text{agree}(u, \ell, (\text{ALG}))$;

3. And add a lot of **randomness**...

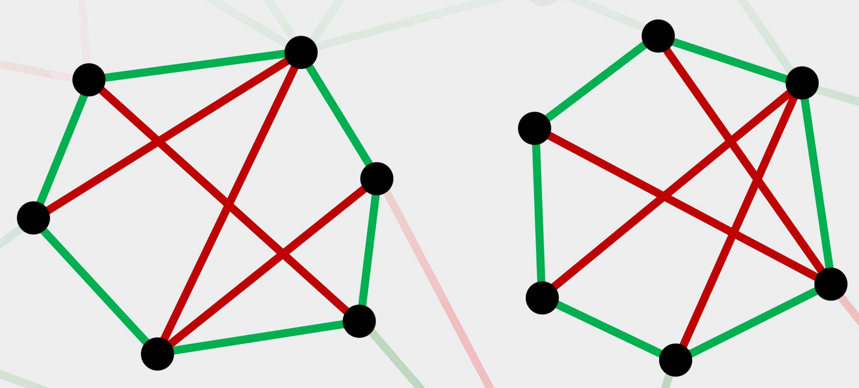
NP-Hardness

In general, hard even on "simple" instances

Theorem

Best Interval Approximation is NP-hard, even if $G^+ = (V, E^+)$ is a cycle. This follows via a reduction from Acyclic Digraph Partition.

Quiz:
Which graph can be perfectly represented by opinion intervals?

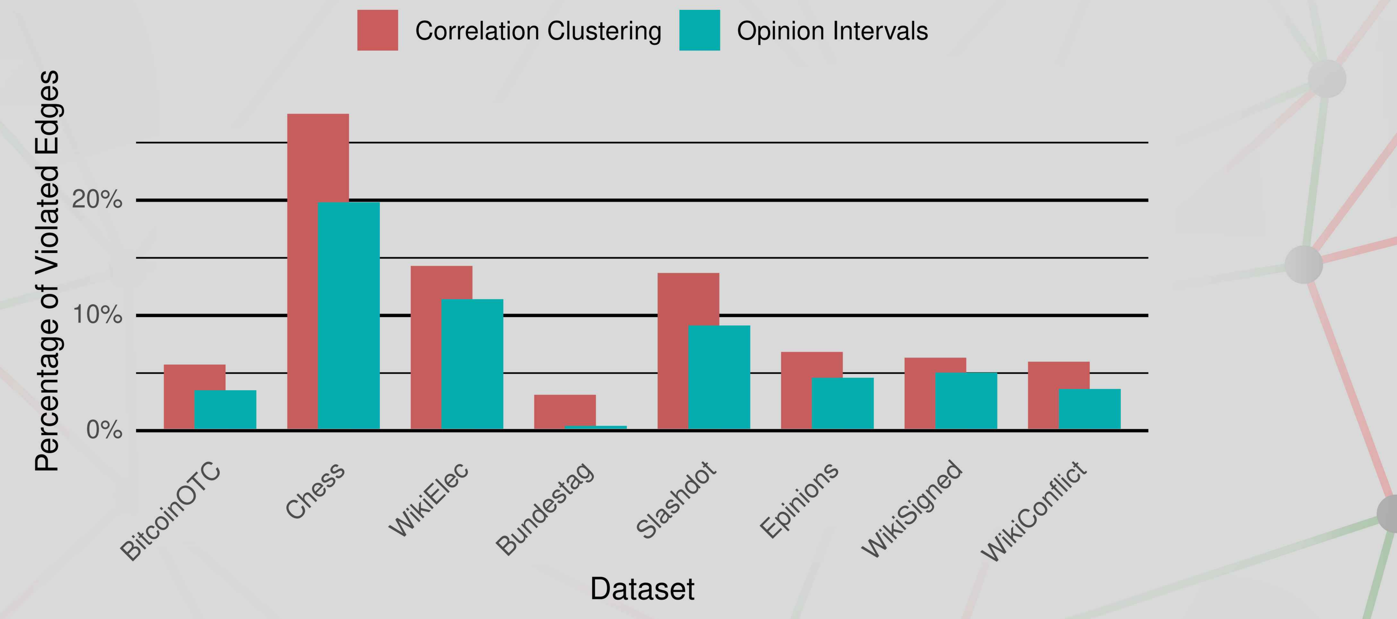


Expressivity Analysis

38% more expressive than Correlation Clustering

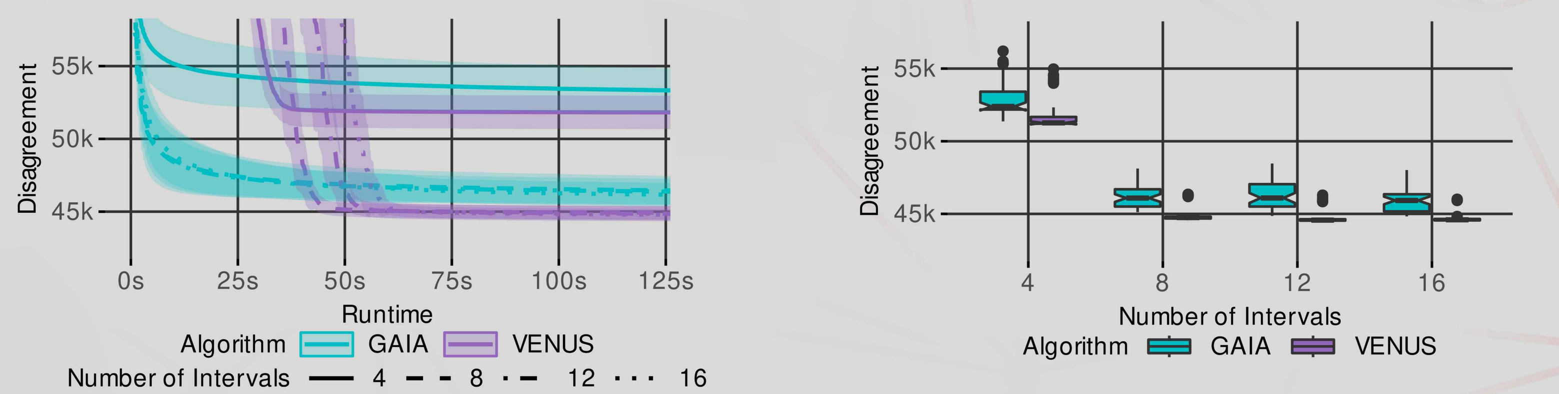
Our algorithms find representations where on average 38% fewer edges are misrepresented, i.e. violate conditions (1) and (2).

Dataset	$\|V\|$	$\|E\|$	Our algorithms: GAIA	Our algorithms: VENUS	Correlation Clustering baselines: GAEC	GAECKLj	SCMLEvo	RAMA	Improvement
BitcoinOTC	5 881	21 434	**3.32**	3.55	5.58	5.57	5.57	5.64	40.39%
Chess	7 301	32 650	19.82	**19.63**	28.64	28.10	27.33	39.98	28.17%
WikiElec	7 115	100 355	**11.24**	11.26	14.13	14.13	14.13	14.45	20.45%
Bundestag	1 480	397 497	**0.25**	**0.25**	3.06	2.95	2.95	3.72	91.53%
Slashdot	82 140	498 532	9.05	**8.94**	13.75	13.66	13.52	17.17	33.88%
Epinions	131 580	708 507	4.47	**4.42**	6.83	6.68	6.67	6.86	33.73%
WikiSigned	138 587	712 337	4.94	**4.85**	6.17	6.17	6.17	6.96	21.39%
WikiConflict	116 836	2 014 053	3.44	**3.43**	5.87	5.82	5.82	6.02	41.06%



Performance

Heuristic solutions found within seconds (results on Slashdot)



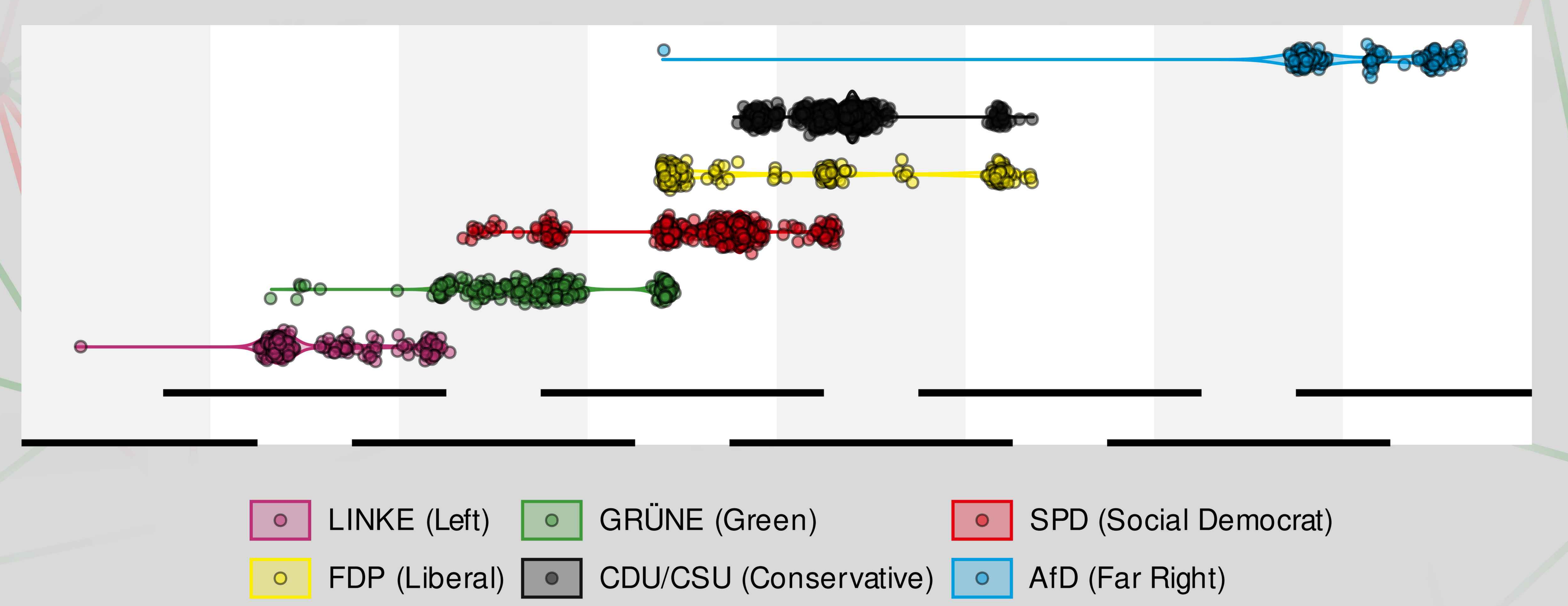
GAIA ... Greedy Agreement Interval Assignment: Greedy Heuristic

VENUS ... Variable ENergy Uphill Search: Simulated Annealing

German Parliament Case Study

Interpretable solutions in the real world

We are not only able to relate individual politicians to each other, but also parties, and rediscover the political spectrum in the German parliament.



Aalto University
University of Oxford
KTH Vetenskap och Konst
TU Wien
Neural Information Processing Systems
W|W|T|F
erc European Research Council
WASP Wallenberg AI, Autonomous Systems and Software Program