POLE: Polarized Embedding for Signed Networks

Zexi Huang
University of California
Santa Barbara, CA, USA
zexi_huang@cs.ucsb.edu

Arlei Silva
Rice University
Houston, TX, USA
arlei@rice.edu

Ambuj Singh
University of California
Santa Barbara, CA, USA
ambuj@cs.ucsb.edu

ABSTRACT

From the 2016 U.S. presidential election to the 2021 Capitol riots to the spread of misinformation related to COVID-19, many have blamed social media for today's deeply divided society. Recent advances in machine learning for signed networks hold the promise to guide small interventions with the goal of reducing polarization in social media. However, existing models are especially ineffective in predicting conflicts (or negative links) among users. This is due to a strong correlation between link signs and the network structure, where negative links between polarized communities are too sparse to be predicted even by state-of-the-art approaches. To address this problem, we first design a partition-agnostic polarization measure for signed graphs based on the signed random-walk and show that many real-world graphs are highly polarized. Then, we propose POLE (POLarized Embedding for signed networks), a signed embedding method for polarized graphs that captures both topological and signed similarities jointly via signed autocovariance. Through extensive experiments, we show that POLE significantly outperforms state-of-the-art methods in signed link prediction, particularly for negative links with gains of up to one order of magnitude.

CCS CONCEPTS

• **Computing methodologies** → **Learning latent representations**; • **Information systems** → **Social networks**.

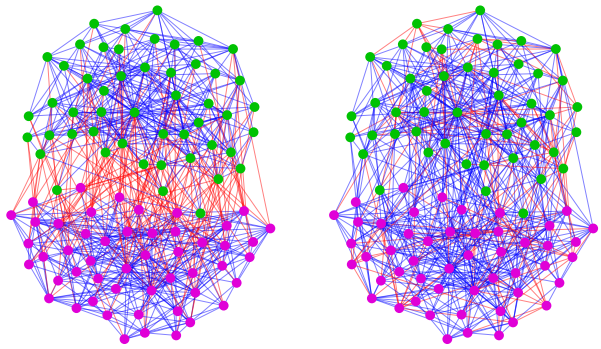
KEYWORDS

Representation learning; Signed embedding; Social polarization

ACM Reference Format:
Zexi Huang, Arlei Silva, and Ambuj Singh. 2022. POLE: Polarized Embedding for Signed Networks. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22), February 21–25, 2022, Tempe, AZ, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3488560.3498454

1 INTRODUCTION

Social media has made our world more polarized [11, 12, 51]. The events surrounding the 2016 U.S. election [16] and, more recently, the tragic U.S. Capitol riot [44] and spread of COVID-19 misinformation [36, 40], have illustrated the dangers of a deeply ideologically divided society. But if technology has led to the rise of polarization, can it also help us to solve it? More specifically, can recent advances in representation learning [22, 42] help us to address online polarization? One could argue these methods should be as effective for predicting conflicts as they are for recommending connections and content in social media platforms [13]. However, we will show that polarization leads to new challenges for representation learning.



(a) LFR-polarized (b) LFR-unpolarized

Figure 1: Two synthetic graphs with the same underlying topology but different link signs. Node colors depict communities and link colors depict signs. (a) is more polarized than (b) as its link signs are related to the community structure—negative links connecting two polarized communities.

Signed graphs are a powerful tool for analyzing social polarization, capturing both positive (friendly) and negative (hostile) connections between entities. They have been used to model relationships between politicians in the U.S. Congress [50] and interactions between Twitter users on political matters [30], both of which are known to harbor polarization [6, 49]. In signed graphs, polarization is often related to the emergence of conflicting communities [3], where nodes within each community form dense positive connections and nodes across different communities are sparsely connected via negative links. Figure 1 shows two signed graphs with the same underlying topology (based on the LFR benchmark [32]) but different link signs. The left graph is more polarized as all of its negative links connect the two communities, while link signs in the right graph are unrelated to the community structure.

To fight polarization, we first need to measure it. Existing measures [3, 52] are defined as objective functions for community detection, requiring community memberships as input. This dependence makes those measures less useful for real-world graphs, where ground-truth memberships are often unknown. To solve this problem, we propose a novel polarization measure based on the correlation between unsigned and signed random-walk dynamics. By varying the random-walk length, our measure naturally captures polarized community structure at different scales and does not

WSDM '22, February 21–25, 2022, Tempe, AZ, USA

ACM ISBN 978-1-4503-9132-0/22/02.
https://doi.org/10.1145/3488560.3498454

rely on specific partitions. We will demonstrate its effectiveness in characterizing both node and graph-level polarization.

But how can we fight polarization in a network? For unsigned graphs, one approach is to bridge communities to reduce the effect of echo chambers [9]. Signed graphs open new possibilities in this endeavor. In particular, if one can identify potential hostile links that are likely to be formed in the future, they can take preventive measures to reduce further polarization. The success of this approach relies on the accuracy of signed link prediction [5, 19, 53], where both existence and signs of future links are inferred. However, while unsigned and signed graph embedding methods have shown success at the related link-level tasks of (unsigned) link prediction [14, 20] and sign prediction [26, 54], they cannot be easily combined for signed link prediction in polarized graphs. That is because negative links mostly connect antagonistic communities and are much sparser than positive links, making it nearly impossible to predict their existence with unsigned embedding methods. And without link existence information, even a perfect sign prediction oracle is unable to predict the negative links.

To address the challenges mentioned above, we propose POLE, a novel signed embedding method for polarized graphs. The key feature that distinguishes our method from existing ones is that it captures signed and topological similarities jointly. Specifically, it guarantees that positively related pairs are more similar than unrelated topologically distant pairs, which are in turn more similar than negatively related pairs. This is accomplished by leveraging the signed random-walk that incorporates social balance theory [18] and extending autocovariance similarity [7, 45] to signed graphs. In this way, negative links can be predicted as the most dissimilar node pairs in the graph, at the other end of the similarity spectrum.

To summarize, our main contributions are:

- We design a novel partition-agn for signed graphs based on the si
- We analyze how existing signed signed link prediction for polariz
- We propose POLE, a novel signe polarized graphs that captures b similarities via signed autocovar
- We conduct an extensive exper method on six real-world signed POLE significantly outperforms s signed link prediction, especially

2 RANDOM-WALK ON SIGI

We introduce our formulation for rando the basis for our polarization measure a

2.1 Notations

A signed undirected weighted graph is $\mathcal{V} = \{1, \dots n\}$ denotes the set of n node the set of m links, with positive and ne represented by a signed weighted adjace $A_{uv} > 0$ (or < 0) if a positive (or nega connects nodes u and v, and $A_{uv} = 0$ o absolute adjacency matrix $|A|$ to constru D/d with $D_{uu} = d_u = \sum_v |A|_{uv}$ as the d

2.2 Signed Random-walk

Random-walks are widely used for unsigned graph embedding due to their ability to capture topological similarity at multiple structural scales. While there have been several attempts to extend random-walks to signed graphs [24, 59], we will define our own version to capture topological and signed similarity jointly and explicitly guarantee *polarized similarity consistency*—we will discuss this important property in more detail in the next subsection.

We start by recalling random-walks for unsigned graphs. A walk l is a sequence of nodes $\langle w_0, w_1, \dots w_t \rangle$ where $(w_\tau, w_{\tau+1}) \in \mathcal{E}$. The transition probability of the walk is the product of stepwise ones:

$$\begin{aligned} \text{Prob}(l) &= \prod_{(w_\tau, w_{\tau+1}) \in l} \text{Prob}(w_{\tau+1} | w_\tau) \\ &= \prod_{(w_\tau, w_{\tau+1}) \in l} |A|_{w_\tau w_{\tau+1}} / d_{w_\tau} \end{aligned} \tag{1}$$

Then, the t-step random-walk transition probability from node u to v can be expressed as the sum of the transition probabilities of all length-t walks between u and v, denoted as $\text{Walk}(u, v, t)$:

$$|M|_{uv}(t) = \sum_{l \in \text{Walk}(u,v;t)} \text{Prob}(l) \tag{2}$$

which serves as a measure of topological similarity between u and v. The Markov time t controls the scale of the walk.

For signed graphs, we keep transition probabilities of walks for topological similarity and add an inferred sign for each walk to capture signed similarity, leading to

$$M_{uv}(t) = \sum_{l \in \text{Walk}(u,v;t)} \text{Sign}(l) \, \text{Prob}(l) \tag{3}$$

where $\text{Sign}(l) = \text{Sign}(\prod_{(w,w') \in l} A_{ww'})$ determines the sign of the

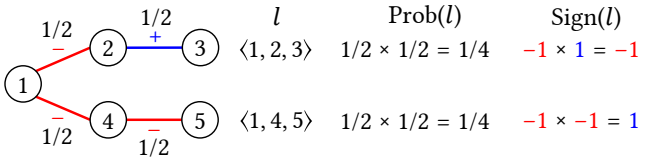


Figure 2: Examples of signed walks with corresponding transition probabilities (Prob**) and inferred signs (**Sign**).**

Note that after adding signs to walks, $M_{uv}(t) \in [-1, 1]$ is no longer a probability but rather captures a notion of signed similarity between u and v. In matrix form, we will still call $M(t) \in \mathbb{R}^{n \times n}$ the signed random-walk transition matrix even though it is not stochastic. A nice property of $M(t)$ is that it can be conveniently expressed in terms of the signed adjacency and degree matrices:

$$M(t) = \begin{cases} (D^{-1}A)^t & \text{for discrete random-walks} \\ \exp(-(I - D^{-1}A)t) & \text{for continuous random-walks} \end{cases} \tag{4}$$

where $I \in \mathbb{R}^{n \times n}$ is the identity matrix. We will use the continuous version of the transition matrix in the rest of the paper due to its finer granularity of the Markov time ($t \in [0, +\infty)$ instead of $t \in \mathbb{N}_0$).

2.3 Similarity Consistency

The key advantage of our signed random-walk is that it guarante *polarized similarity consistency*, a property critical to signed li prediction in polarized graphs. The *consistency* of similarity stat which type of node pairs should be more similar than other typ For example, unsigned embedding methods satisfy:

Property 1 (Topological similarity consistency). *Topolo cally close nodes are more similar than topologically distant ones.*

And signed embedding methods guarantee:

Property 2 (Signed similarity consistency). *Positively relat nodes are more similar than negatively related ones.*

These properties enable corresponding link-level downstream tas namely (unsigned) link prediction and sign prediction.

The transitions of our signed random-walk as a similarity met have a stronger consistency guarantee:

Property 3 (Polarized similarity consistency). *Positively lated node pairs are more similar than unrelated topologically distant pairs, which are in turn more similar than negatively related pairs.*

Property 3 is a direct implication of the transition probability defined in Equation 3. Large positive/negative values of $M_{uv}(t)$ indicate that u and v are connected via mostly positive/negative paths while small values mean that they are unrelated and topologically distant. Property 3 also holds for a similarity defined by dot products between columns of M. More specifically, $\langle M_{:u}, M_{:v} \rangle$ is large and positive (or negative) if and only if u and v have transitions of same (or different) sign(s) to a common set of nodes. And the similarity is small if their transitions reach a distinct set of nodes.

3 A MEASURE OF POLARIZATION

In this section, we propose and evaluate a measure of polarization based on the signed random-walks we have just introduced.

3.1 Random-walk Based Polarization

A signed network is polarized if it comprises antagonistic communities with dense positive connections in each community and sparse negative ones across communities. Unlike previous works [3, 52] that define polarization as a quality measure of graph partitions, our aim is to measure polarization based solely on graph structure. To achieve this, we apply a soft characterization of the polarized community structure in the graph based on signed random-walks.

The key observation that supports our polarization measure is that polarization increases the correlation between unsigned and signed random-walk dynamics. Consider the two graphs shown in Figure 3. While they share the same underlying topology, (a) is more polarized than (b) as its negative link connects the two communities. We then analyze how their signed random-walk transitions differ from unsigned ones. For node 1 in graph (a), only signs of its inter-community transitions are changed, whose magnitudes are smaller, if not negligible, compared to the intra-community transitions. On the other hand, the negative link in the less polarized graph (b) significantly affects the signed intra-community transitions. As a consequence, the signed and unsigned transitions in the polarized graph are more correlated than in the less polarized one.

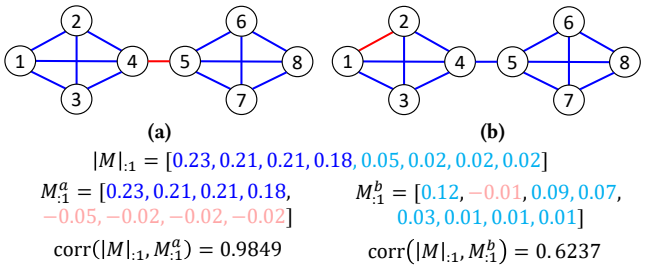


Figure 3: Illustration of our polarization measure based on signed random-walk dynamics ($t = 3$). The transitions for unsigned ($|M|$) and signed (M) random-walks are more correlated in the polarized network (a), where a negative link connects two antagonistic communities. On the other hand, the correlation is lower in the less polarized network (b).

We define the *node-level polarization* as the (Pearson) correlation between a node's signed and unsigned random-walk transitions:

$$\mathrm{Pol}(u; t) = \mathrm{corr}(|M|_{:u}(t), M_{:u}(t)) \tag{5}$$

We can then define the *graph-level polarization* as the mean node-level polarization for all nodes in the graph:

$$\mathrm{Pol}(G; t) = \underset{u \in G}{\mathrm{mean}}(\mathrm{Pol}(u; t)) \tag{6}$$

The random-walk based polarization proposed here has two main advantages. First, it is partition-agnostic and does not depend on the availability of ground-truth communities or the outcome of community detection algorithms. Second, it can measure polarization at different structural scales by tuning the Markov time t—large t captures polarization between macro-level communities (e.g., hostility between political parties) while small t measures it at a micro scale (e.g., disagreement between factions within a party). In the next subsection, we will demonstrate that our measure is effective in characterizing both node and graph-level polarization.

3.2 Polarization of Real-world Graphs

3.2.1 Node-level polarization. We apply our polarization measure to characterize nodes in a political network (Congress) [50]. Signed links in this network represent (un/)favorable interactions between U.S. congresspeople on the House floor in 2005. The statistics of the network are shown in Table 1.

We rank all nodes by their polarization scores at a fixed $t = 10$. The top 20 least and most polarized nodes are shown in Table 2 and Table 3 (both in Appendix A.1). The congressperson with the smallest score (-0.6542) is Henry Cuellar, a self-described moderate-centrist [39]. As a Democrat, he voted with President Trump (Republican) nearly 75% of the time, advocating his fellow Democrats to embrace a more conservative voting record [37]. The second least polarized person is Jane Harman (with -0.5376). She was described as a centrist, particularly on defense and intelligence issues [43]. She was also once called the best Republican in the Democratic Party [46]. In general, we found that our polarization measure captures the political views (centrists vs. extremists) of the congresspeople based on their signed interactions.

3.2.2 Graph-level polarization. We also apply our graph-level polarization measure to characterize real-world signed graphs. In addition to Congress, we consider five other networks, with their statistics summarized in Table 1:

- WoW-EP8 [27]: Interaction network of editors in the eighth legislature of the European Parliament. Link signs indicate whether they collaborate or compete with each other.
- Bitcoin-Alpha and Bitcoin-OTC [28]: Trust networks of Bitcoin traders on the platforms *Bitcoin Alpha* and *Bitcoin OTC*. Link weights and signs are based on users' rating of each other.
- Referendum [30]: Interaction network of Twitter users on the 2016 constitutional referendum in Italy. Link signs are inferred from the stance of the users.
- Wiki-RfA [56]: Voting network of Wikipedia users on adminship. Link signs indicate whether users vote for/against each other.

Table 1: An overview of the datasets.

	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{E}_-\rvert/\lvert\mathcal{E}\rvert$
Congress	219	523	20.46%
WoW-EP8	789	116,009	18.63%
Bitcoin-Alpha	3,772	14,077	9.31%
Bitcoin-OTC	5,872	21,431	14.71%
Referendum	10,864	251,396	5.09%
Wiki-RfA	11,275	169,925	22.04%

As reference for graph-level polarization, we construct two synthetic graphs, one polarized and the other unpolarized, with the same underlying topology but different link signs, as shown in Figure 1. The topology is based on the LFR benchmark [32] with the following parameters: two structural communities with 50 nodes, an average node degree of 12, and an inter-community link ratio of 0.15. Then, for the polarized version (LFR-polarized), we assign link signs fully based on the structural communities—negative for inter-community links and positive for intra-community links. For the unpolarized version (LFR-unpolarized), we first assign nodes to two random communities with the same cut size as structural communities. We then assign link signs based on the random communities following the same rule. In this way, both graphs have the same proportion of negative links and the same social balance—proportion of closed triangles in the graph that satisfy the social balance theory [34]—of 1.0.

The polarization and social balance for the real-world graphs along with the two synthetic ones are shown in Figure 4, with the Markov time t selected based on signed link prediction performance (details in Section 5.1.3). As we see, most real-world graphs are at the same level of polarization as LFR-polarized. On the other hand, their social balance is not directly related to polarization, contrary to as one might think [11]. The only dataset that falls behind in polarization is Wiki-RfA, which might be due to the more collaborative nature of its inter-community interactions. As we will discuss in the next section, effective link prediction for those polarized graphs requires a novel embedding scheme that satisfies *polarized similarity consistency*.

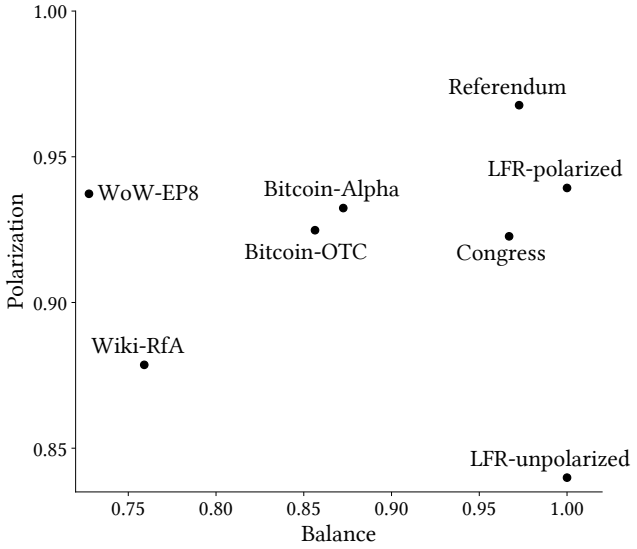


Figure 4: Polarization and social balance of real-world graphs, with reference to synthetic ones. Most real-world graphs are as polarized as the synthetic polarized one.

4 POLARIZED EMBEDDING FOR NETWORKS

Now that we have shown that many real-world graphs are polarized, here, we propose a novel embedding method for effectively predicting signed links in polarized graphs. We first demonstrate that existing embedding methods are incapable of this task due to their weak similarity consistency (see Section 2.3). We then introduce polarized embedding based on autocovariance and matrix factorization that addresses the limitations of existing approaches.

4.1 Limitation of Existing Methods

The objective of signed link prediction [1, 57] is to predict both the existence and the signs of future links given the observed graph. Most of existing signed embedding methods focus on the second half of the task, sign prediction, which is specific to signed graphs. The first half, link prediction, is a common downstream task for unsigned embedding methods. Combining them seems like a viable solution for the task. However, this approach fails to predict negative links in polarized graphs because they are sparse and mostly appear as inter-community connections. They are hard, if not impossible, to be predicted by unsigned link prediction algorithms based on topology only. Without the existence of links known a priori, sign prediction itself is incapable of predicting negative links.

We identify the key weakness of existing embedding methods as that they only preserve weak similarity consistency. In particular, signed embedding methods [4, 21, 22, 26, 54, 57] just ensure *signed similarity consistency*—positively related pairs are separable from negatively related pairs—for sign prediction. And similarly, unsigned embedding methods [14, 20, 42] only guarantee the *topological similarity consistency*—topologically close node pairs are separable from distant pairs—for link prediction. While positive links are detectable as the intersection of positive and topologically

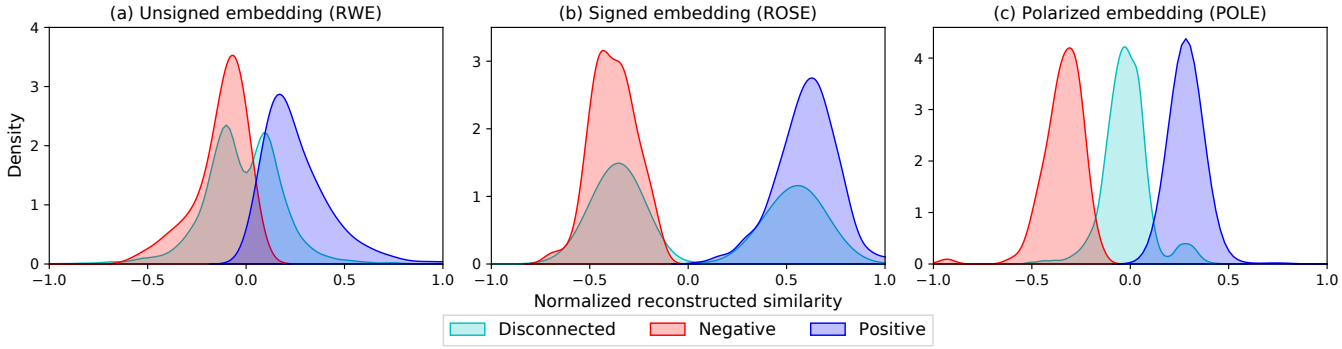


Figure 5: Distributions of the reconstructed similarity for different types of node pairs in LFR-POLARIZED. Polarized embedding (c) enables separation of negatively connected pairs from the others while both (a) unsigned embedding (RWE [20]) and (b) signed embedding (ROSE [22]) fail to do so.

close pairs, negative links in polarized graphs would remain hidden with other topologically distant pairs. Figure 5 (a) and Figure 5 (b) illustrate this idea. They show the distributions of reconstructed similarity of positively connected pairs, negatively connected pairs, and disconnected pairs for LFR-POLARIZED via dot products of unsigned and signed embedding, respectively. While both embedding methods are able to capture their respective similarity consistency, negative links are always inseparable from disconnected pairs. This motivates us to design an embedding method with a stronger similarity consistency that enables the separation of negative pairs from the others. We describe such an embedding method next.

4.2 The Solution: Polarized Embedding

Our solution to the separability of negative pairs is polarized embedding, a novel embedding scheme that captures *polarized similarity consistency*, the strongest consistency among the three. This consistency guarantees that negatively related pairs are more dissimilar than unrelated topologically distant pairs (thus "polarized"), which are more dissimilar than positively related pairs. In this way, the negatively related pairs stand out at the other end of the similarity spectrum, easily separable from others. In the following subsections, we introduce our polarized embedding method, POLE, that applies signed autocovariance similarity and matrix factorization.

4.2.1 Signed autocovariance similarity. POLE is based on the signed random-walk introduced in Section 2.2. As discussed in Section 2.3, both the entries and the dot products of columns of the signed random-walk transition matrix $M(t)$ can be viewed as similarity metrics that satisfy *polarized similarity consistency*. Therefore, one can directly use a low-rank representation of them—for example, the matrix factorization of $M(t)$ or $M(t)^T M(t)$—as embeddings. However, those embeddings may not be effective for link-level inference. Instead, we propose to take advantage of the random-walk based similarity metrics that have been well understood in unsigned embedding. In particular, it has been shown that autocovariance similarity [7] enables state-of-the-art performance in unsigned link prediction by incorporating node degree information [20]. This coincides with our goal of predicting signed links in polarized graphs. Thus, we first extend autocovariance similarity to signed graphs.

The unsigned autocovariance similarity is built upon the co-visiting probability for node pairs in a walk. However, as defined in our signed random-walk, $M(t)$ is not a stochastic matrix and does not encode probabilities. Instead, we resort to another interpretation of autocovariance—a centered dynamic similarity metric [45]. The dynamic similarity based on $M(t)$ is

$$R(t) = M(t)^T W M(t) \tag{7}$$

where $W \in \mathbb{R}^{n \times n}$ is a weight matrix. In the unsigned case, selecting $W = \Pi - \pi\pi^T$ makes $R(t)$ equivalent to the autocovariance similarity, where $\pi \in \mathbb{R}^n$ is the stationary distribution of the unsigned random-walk and $\Pi = \text{diag}(\pi)$. For the signed case, $M(t)$ is singular if and only if the graph is perfectly balanced [29]. This means that there is no "stationary distribution" for signed random-walks on most real-world graphs. However, as [31] points out, the role of stationary distribution in the similarity formulation is to provide a centrality measure—in the unsigned case, the degree centrality. Thus, we can also use node degrees to construct the weight matrix:

$$W = \frac{1}{\text{vol}(G)} D - \frac{1}{\text{vol}(G)^2} d d^T \tag{8}$$

where $\text{vol}(G) = \sum_u d_u$. And substituting this weight matrix into Equation 7 leads to the signed autocovariance similarity matrix. To the best our knowledge, we are the first to extend random-walk based similarity metrics to signed graphs in a principled manner.

4.2.2 Matrix factorization. We are now ready to introduce how to generate embedding based on signed autocovariance. Let $\mathbf{u}_u \in \mathbb{R}^k$ be the embedding of node u and $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n)^T \in \mathbb{R}^{n \times k}$ be the embedding matrix. We use the dot product in the embedding space to preserve the signed autocovariance similarity R:

$$\begin{aligned} U^* &= \arg\min_U \sum_{u,v} (\mathbf{u}_u^T \mathbf{u}_v - R_{uv})^2 \\ &= \arg\min_U \|UU^T - R\|_F^2 \end{aligned} \tag{9}$$

This leads to a straightforward matrix factorization algorithm to find the optimal embedding. Specifically, $U^* = Q_k \sqrt{\Lambda_k}$—where $R = Q\Lambda Q^T$ is the Singular Value Decomposition (SVD) of R—is the optimal solution of U under the constraint $\text{rank}(UU^T) = k$ [8].

Figure 5 (c) shows distributions of reconstructed similarity of different types of node pairs for polarized embedding. It is clear that negative pairs can be effectively separated from unrelated and positive pairs. In addition, with negative pairs staying at the negative end of the similarity spectrum, positive pairs are also more separable. This demonstrates the strength of our polarized embedding that captures *polarized similarity consistency*.

5 EXPERIMENTS

In this section, we demonstrate the effectiveness of POLE in signed link prediction using the six datasets in Table 1. Our code is available at https://github.com/zexihuang/POLE.

5.1 Experimental Settings

5.1.1 Baselines. We consider the following baselines:

- SiNE [54]: Combines an objective function based on social balance theory with a learned pairwise similarity function.
- SIGNet [21]: Attempts to capture higher-order balance structure by computing embeddings for which dot product approximates the signed proximity for links in the graph.
- SIDE [26]: Extends random-walk based embedding to signed graphs. Embeddings are learned based on maximum likelihood using pairwise proximities that are sensitive to link signs.
- BESIDE [4]: Applies balance and status theory to model signed triangles and "bridge" links, which are not included in a triangle.
- SLF [57]: Decomposes node embeddings into two types of latent factors (positive and negative). These factors, which are learned via coordinate descent, are applied to generate four types of scores corresponding to positive, negative, neutral, and no link.
- ROSE [22]: Transforms the signed network into an unsigned bipartite one by representing each node multiple times. Embeddings for the unsigned network are generated using a random-walk based approach and combined into signed embeddings.

5.1.2 Downstream task. We focus on signed link prediction, which consists of predicting of both link existence and signs. We randomly remove 20% of links while ensuring that the residual graph is connected and embed the residual graph. Then, for POLE, we rank all disconnected node pairs based on reconstructed dot product similarity and predict the most similar/dissimilar pairs as positive/negative links. For baselines, we follow [57] and train two logistic regression classifiers on concatenated node embeddings for positive/negative pairs vs disconnected pairs, respectively. While ranking by classifier scores is adopted by most baselines, we also conduct experiments with baselines using the same dot product similarity ranking as for POLE. The results are similar and included in Appendix A.2. We report *precision@k* [35] for positive/negative links respectively, where k is the number of top pairs in terms of the ratio of removed positive/negative links, ranging from 10% to 100%.

Since the baselines are designed to only capture *signed similarity consistency* and thus may perform poorly without link existence information, we also consider supplementing them with an unsigned embedding method. This also allows us to analyze the interaction between unsigned similarity and the signed similarity captured by our polarized embedding. Specifically, we apply RWE [20] with unsigned autocovariance and continuous random-walk and compute the reconstructed similarity based on embeddings for the unsigned residual graph. We then train logistic regression classifiers to combine the unsigned similarity (encoding link existence information) and either reconstructed similarity (for POLE) or classifier scores (for baselines) for final ranking and report *precision@k* scores.

5.1.3 Parameters. We set the number of embedding dimensions k to 40 for all methods (except ROSE which requires a multiple of 3, we set it to 42). The only other parameter in RWE and POLE is the Markov time t, which is selected from $\{10^{0.0}, 10^{0.1}, \ldots, 10^{1.0}\}$ based on signed/unsigned link prediction. Other parameters for baselines are set as recommended in the original papers.

5.2 Results

5.2.1 Signed link prediction. Figure 6 shows the signed link prediction performance for different methods without supplementing the unsigned link existence information. We first note that the proposed method, POLE, outperforms all baselines in almost all datasets for both positive and negative links. The average gains in *precision@k* over the best baseline for each dataset are 629.2%/58.8%/115.7% /108.3% /142.5%/468.8% on positive links and 220.0%/27.6%/261.8% /1539.4%/4.3%/-10.9% on negative links. This shows that POLE enables state-of-the-art signed link prediction performance.

While the main motivation for POLE is predicting negative links, its superior performance in positive link prediction is also consistent with our previous similarity analysis (see Figure 5 (b) and Figure 5 (c)). The only scenario where POLE underperforms the best baseline is the negative link prediction on Wiki-RfA. This is actually not surprising as Wiki-RfA is the least polarized network among the six, as shown in Figure 4. When the graph is less polarized (e.g., Figure 1 (b)), negative links are formed independently from community structure, and therefore the *polarized similarity consistency* may not be the best criterion to embed nodes.

5.2.2 Signed link prediction with link existence information. We now supplement each method with unsigned link existence information. Results are shown in Figure 7. Adding unsigned information narrows down the average performance gains of POLE over baselines on positive links to 18.0%/5.1%/4.4%/-6.6%/11.2%/3.5%. But our method still exhibits significant gains on negative links, outperforming the best baseline by 300.0%/93.9%/70.4%/243.8%/29.6%/79.7% for all datasets including the least polarized Wiki-RfA. Even with unsigned link existence information, our polarized embedding method has an edge over existing approaches, especially for negative links.

The notable improvement of baselines after adding unsigned link existence information is consistent with our expectation. For positive links, combining the signed/unsigned embeddings enables the separation of positive pairs from negative/disconnected pairs, respectively, as illustrated in Figure 5 (a) and Figure 5 (b). This boosts baselines to a performance comparable with ours. For negative links, while all baselines still underperform POLE by a large margin, their predictions are improved significantly compared to the scenario without link existence information. The most prominent improvement is observed for Bitcoin-OTC, from negligible precision scores to notable ones of around 0.2 at $k = 100\%$. This is reasonable since none of the networks are fully polarized, and not all negative links in the networks are inter-community. Those that are intra-community can be correctly predicted by unsigned

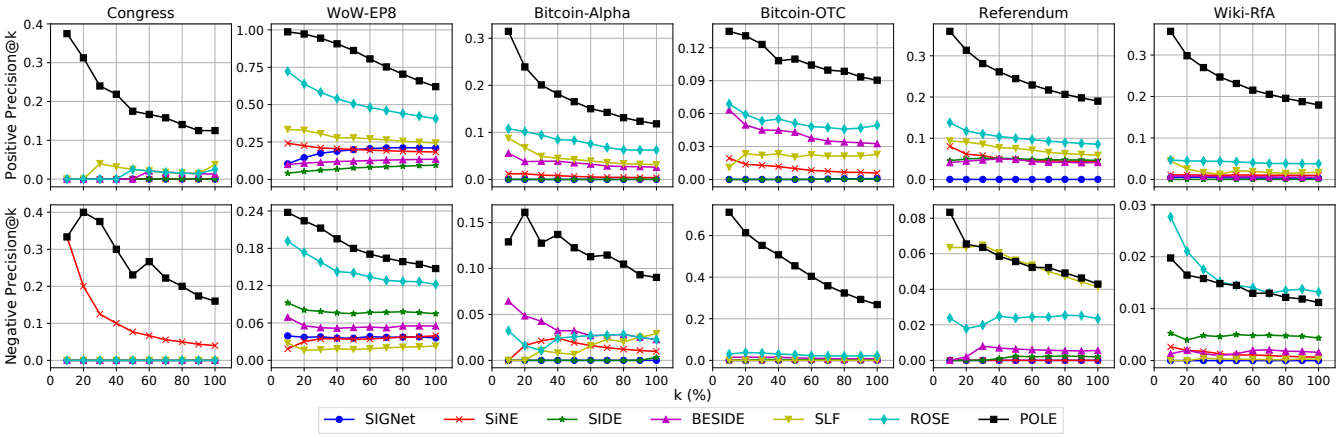


Figure 6: Comparison of performance of signed link prediction between POLE and baselines. POLE outperforms all baselines in all datasets on both positive and negative link prediction, except for negative links in Wiki-RfA, the least polarized network.

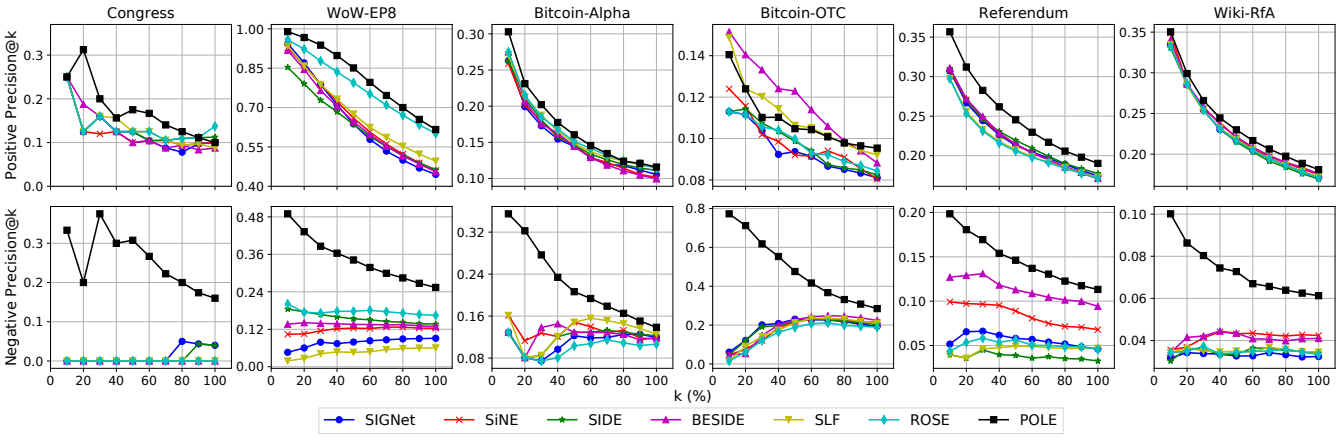


Figure 7: Comparison of performance of signed link prediction with link existence information between POLE and baselines. While adding unsigned similarity narrows the performance gap between POLE and baselines on positive link prediction, it significantly improves POLE on negative link prediction and keeps its edge.

similarity first and then correct signs can be discovered using the baselines. The inter-community negative links, however, remain exclusively predictable via *polarized similarity consistency*, for which our polarized embedding method is designed.

At this point, it is important to analyze the interaction between unsigned similarity and our polarized embedding. While the performance for negative links more than doubles on average across the datasets, precision for positive links decreases by an average of 2.9%. To understand this phenomenon, we draw scatter plots of the reconstructed signed and unsigned similarity for different node pairs in signed link prediction, along with the decision boundaries based on each similarity and a combination of both (via the classifier). The results for Referendum and Wiki-RfA are shown in Figure 8, with the rest in Figure 11 in Appendix A.3. For positive pairs, signed and unsigned similarities are highly correlated, and signed similarity alone is predictive enough. This is consistent with the fact that regions above the signed and combined decision boundaries highly overlap. On the other hand, prediction for negative links is clearly improved by combining signed and unsigned similarities. The decision boundaries of the learned classifiers imply that negative pairs should have smaller signed similarity than unsigned similarity. This rule is especially useful as it filters out a large number of disconnected pairs with large negative signed similarity but even larger negative unsigned similarity. Finally, this plot is another illustration of how our method captures polarization. Referendum, the most polarized graph (see Figure 4) of all, shows a clearer separation between positive and negative pairs compared to Wiki-RfA, which is the least polarized.

6 RELATED WORK

Measuring polarization. Social polarization has received great attention with the increasing popularity in online social media [15].

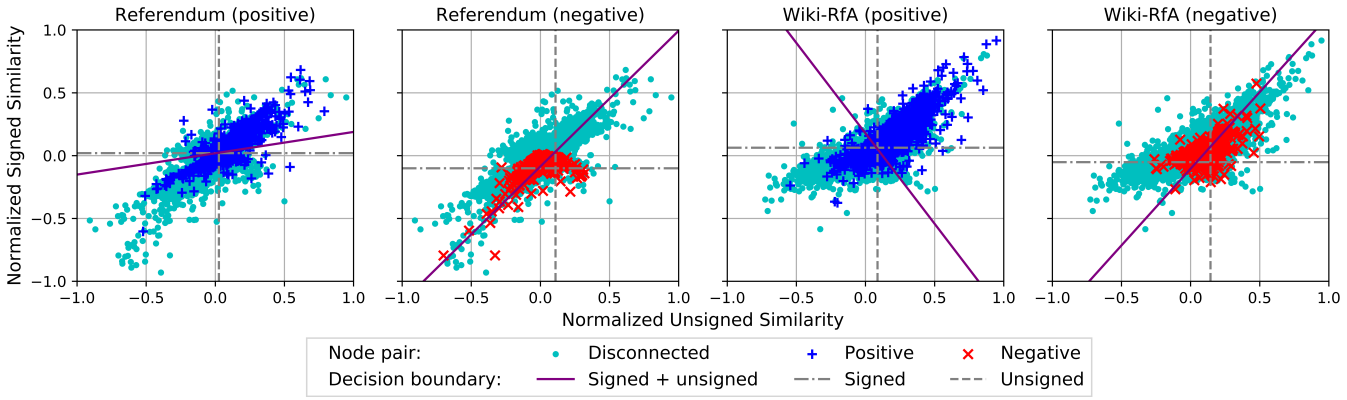


Figure 8: Scatter plot of the reconstructed signed and unsigned similarity for different node pairs in signed link prediction, along with the decision boundaries based on each similarity and a combination of both (via the classifier). Combining signed and unsigned similarity improves prediction for negative links but has a negligible effect on predicting positive links.

In the context of networks/graphs, polarization is often defined as the cohesiveness of communities. Several community quality measures, such as conductance [25] and modularity [41], have been used for analyzing polarization in the U.S. congress [55] and online social media platforms [6]. Later works consider other measures based on the community structure. For example, [15] measures polarization based on the concentration of high-degree nodes in the community boundaries, and [10] extends it with betweenness and inter-community random-walk transitions. Likewise, polarization in signed graphs is often related to the objective function of signed community detection. In [3], the authors define polarization as the size-normalized signed cut between two conflicting communities, which is later extended to multiple communities in [52]. The main distinction of our polarization measure compared with existing methods is that it leverages the signed random-walk to capture the community structure at multiple scales and does not require specific community assignments as input.

Signed link prediction. Link prediction is a common inference task for graphs [35, 38]. Link signs in signed graphs add another dimension to the problem, leading to the different tasks of (1) link sign prediction [23, 33, 47], (2) link existence prediction [22, 61], and (3) signed link prediction [5, 19, 53]. Signed link prediction is the most challenging one among them, requiring both existence and signs of links to be inferred. Existing signed link prediction methods are based on feature engineering [5], matrix factorization [19, 58], and graph embedding [57]. Due to sparsity of links, several methods also consider side information, including user-item ratings [1], user-user opinions [1, 2], and user-review ratings [48]. Our work focuses on signed link prediction in polarized graphs, which is an even harder problem as negative links mostly appear as sparse inter-community connections. Yet by incorporating polarization in the embedding, our method enables state-of-the-art signed link prediction without using additional information.

Signed graph embedding. Signed graph embedding enables the application of classical machine learning algorithms to graph-based downstream tasks, such as signed link prediction [5, 19], node classification [21], and polarized community detection [3, 52]. Existing methods are often based on extending unsigned embedding methods (such as those based on random-walks [14, 42]) and incorporating social theories (such as social balance theory [18] and social status theory [17]). Both SNE [60] and SIDE [26] extend random-walk based objective functions to signed embedding. SiNE [54] incorporates social balance theory and uses a learned pairwise similarity function modeled as a multi-layer neural network. SIGNet [21] captures higher-order balance structure with efficient sampling algorithms. BESIDE [4] combines social balance theory and social status theory to model triangles and "bridge" links for embedding. SLF [57] models each node with four signed latent factors in order to capture positive, negative, neutral, and the absence of a relationship between them. ROSE [22] transforms the signed network into a bipartite unsigned network and leverages an existing random-walk based method [14] for embedding. While these methods capture the *signed similarity consistency* needed for sign prediction (see Figure 5), they are unable to predict negative links in polarized graphs even when combined with unsigned embedding [20]. By comparison, our polarized embedding preserves *polarized similarity consistency* and can effectively predict both positive and negative links in real-world signed graphs.

7 CONCLUSION

We have introduced several analytical tools for understanding and combating polarization in signed graphs. They include a partition-agnostic polarization measure for both nodes and graphs and an embedding algorithm that enables state-of-the-art signed link prediction, especially for the hostile links in polarized graphs. We hope that our work will be an important step towards making social media an environment for healthy and constructive communication among individuals with diverse viewpoints.

ACKNOWLEDGMENTS

This work is partially funded by NSF via grant IIS 1817046 and DTRA via grant HDTRA1-19-1-0017.

REFERENCES

[1] Ghazaleh Beigi, Suhas Ranganath, and Huan Liu. 2019. Signed link prediction with sparse data: The role of personality information. In *WebConf.*

[2] Ghazaleh Beigi, Jiliang Tang, Suhang Wang, and Huan Liu. 2016. Exploiting emotional information for trust/distrust prediction. In *SDM.*

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

[4] Yiqi Chen, Tieyun Qian, Huan Liu, and Ke Sun. 2018. "Bridge" Enhanced Signed Directed Network Embedding. In *CIKM.*

[5] Kai-Yang Chiang, Nagarajan Natarajan, Ambuj Tewari, and Inderjit S Dhillon. 2011. Exploiting longer cycles for link prediction in signed networks. In *CIKM.*

[6] Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *ICWSM.*

[7] J-C Delvenne, Sophia N Yaliraki, and Mauricio Barahona. 2010. Stability of graph communities across time scales. *PNAS* 107, 29 (2010), 12755–12760.

[8] Carl Eckart and Gale Young. 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1, 3 (1936), 211–218.

[9] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. Reducing controversy by connecting opposing views. In *WSDM.*

[10] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *TSC* 1, 1 (2018), 1–27.

[11] Venkata Rama Kiran Garimella and Ingmar Weber. 2017. A long-term analysis of polarization on Twitter. In *ICWSM.*

[12] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. 2018. Me, my echo chamber, and I: introspection on social media polarization. In *WebConf.*

[13] Palash Goyal and Emilio Ferrara. 2018. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems* 151 (2018), 78–94.

[14] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *SIGKDD.* 855–864.

[15] Pedro Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A measure of polarization on social media networks based on community boundaries. In *ICWSM.*

[16] Andrew Guess, Brendan Nyhan, and Jason Reifler. 2018. Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. *European Research Council* 9, 3 (2018), 4.

[17] Ramanthan Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2004. Propagation of trust and distrust. In *WebConf.*

[18] Fritz Heider. 1946. Attitudes and cognitive organization. *The Journal of psychology* 21, 1 (1946), 107–112.

[19] Cho-Jui Hsieh, Kai-Yang Chiang, and Inderjit S Dhillon. 2012. Low rank modeling of signed networks. In *SIGKDD.*

[20] Zexi Huang, Arlei Silva, and Ambuj Singh. 2021. A Broader Picture of Random-walk Based Graph Embedding. In *SIGKDD.*

[21] Mohammad Raihanul Islam, B Aditya Prakash, and Naren Ramakrishnan. 2018. Signet: Scalable embeddings for signed networks. In *PAKDD.*

[22] Amin Javari, Tyler Derr, Pouya Esmailian, Jiliang Tang, and Kevin Chen-Chuan Chang. 2020. ROSE: Role-based Signed Network Embedding. In *WebConf.*

[23] Amin Javari and Mahdi Jalili. 2014. Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links. *TIST* 5, 2 (2014), 1–19.

[24] Jinhong Jung, Woojeong Jin, Lee Sael, and U Kang. 2016. Personalized ranking in signed networks using signed random walk with restart. In *ICDM.*

[25] Ravi Kannan, Santosh Vempala, and Adrian Vetta. 2004. On clusterings: Good, bad and spectral. *JACM* 51, 3 (2004), 497–515.

[26] Junghwan Kim, Haekyu Park, Ji-Eun Lee, and U Kang. 2018. Side: representation learning in signed directed networks. In *WebConf.*

[27] Victor Kristof, Matthias Grossglauser, and Patrick Thiran. 2020. War of words: The competitive dynamics of legislative processes. In *WebConf.*

[28] Srijan Kumar, Francesca Spezzano, VS Subrahmanian, and Christos Faloutsos. 2016. Edge weight prediction in weighted signed networks. In *ICDM.*

[29] Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto W De Luca, and Sahin Albayrak. 2010. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM.*

[30] Mirko Lai, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso. 2018. Stance evolution and twitter interactions in an italian political debate. In *NLDB.*

[31] Renaud Lambiotte, Jean-Charles Delvenne, and Mauricio Barahona. 2014. Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE TSNE* 1, 2 (2014), 76–90.

[32] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. 2008. Benchmark graphs for testing community detection algorithms. *Physical review E* 78, 4 (2008), 046110.

[33] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Predicting positive and negative links in online social networks. In *WebConf.*

[34] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed networks in social media. In *SIGCHI.*

[35] Linyuan Lü and Tao Zhou. 2011. Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications* 390, 6 (2011), 1150–1170.

[36] Jacqueline M. Otala, Gillian Kurtic, Isabella Grasso, Yu Liu, Jeanna Matthews, and Golshan Madraki. 2021. Political Polarization and Platform Migration: A Study of Parler and Twitter Usage by United States of America Congress Members. In *Fairness, Accountability, Transparency, Ethics and Society on the Web.* 224–231.

[37] Clare Malone. 2017. A Q&A with the house Democrat who's voted with Trump 75 percent of the time. *FiveThirtyEight* (2017).

[38] Víctor Martínez, Fernando Berzal, and Juan-Carlos Cubero. 2016. A survey of link prediction in complex networks. *CSUR* 49, 4 (2016), 1–33.

[39] David McCumber. 2014. From the house on the hill: congressman looks back at his life. *Laredo Morning Times* (2014).

[40] Salman Bin Naeem, Rubina Bhatti, and Aqsa Khan. 2021. An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal* 38, 2 (2021), 143–149.

[41] Mark EJ Newman. 2006. Modularity and community structure in networks. *PNAS* 103, 23 (2006), 8577–8582.

[42] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *SIGKDD.* 701–710.

[43] Ben Pershing. 2009. Pelosi, Harman have long history. *The Washington Post* (2009).

[44] Avinash Prabhu, Dipanwita Guhathakurta, Mallika Subramanian, Manvith Reddy, Shradha Sehgal, Tanvi Karandikar, Amogh Gulati, Udit Arora, Rajiv Ratn Shah, Ponnurangam Kumaraguru, et al. 2021. Capitol (Pat) riots: A comparative study of Twitter and Parler. *arXiv preprint arXiv:2101.06914* (2021).

[45] Michael T Schaub, Jean-Charles Delvenne, Renaud Lambiotte, and Mauricio Barahona. 2019. Multiscale dynamical embeddings of complex networks. *Physical Review E* 99, 6 (2019), 062308.

[46] George Skelton. 1998. California and the west: in the ring, with contenders for governor. *Los Angeles Times* (1998).

[47] Dongjin Song and David A Meyer. 2015. Link sign prediction and ranking in signed directed social networks. *Social network analysis and mining* 5, 1 (2015), 1–14.

[48] Jiliang Tang, Xia Hu, Yi Chang, and Huan Liu. 2014. Predictability of distrust with interaction data. In *CIKM.*

[49] Sean M Theriault. 2006. Party polarization in the US Congress: Member replacement and member adaptation. *Party Politics* 12, 4 (2006), 483–503.

[50] Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: determining support or opposition from congressional floor-debate transcripts. In *EMNLP.*

[51] Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. *Social media, political polarization, and political disinformation: A review of the scientific literature.* Technical Report. The William and Flora Hewlett Foundation.

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

[53] Jing Wang, Jie Shen, Ping Li, and Huan Xu. 2017. Online matrix completion for signed link prediction. In *WSDM.*

[54] Suhang Wang, Jiliang Tang, Charu Aggarwal, Yi Chang, and Huan Liu. 2017. Signed network embedding in social media. In *SDM.*

[55] Andrew Scott Waugh, Liuyi Pei, James H Fowler, Peter J Mucha, and Mason Alexander Porter. 2011. Party polarization in congress: a network science approach. *arXiv preprint arXiv:0907.3509* (2011).

[56] Robert West, Hristo S Paskov, Jure Leskovec, and Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *TACL* 2 (2014), 297–310.

[57] Pinghua Xu, Wenbin Hu, Jia Wu, and Bo Du. 2019. Link prediction with signed latent factors in signed social networks. In *SIGKDD.*

[58] Jihang Ye, Hong Cheng, Zhe Zhu, and Minghua Chen. 2013. Predicting positive and negative links in signed social networks by transfer learning. In *WebConf.*

[59] Xiaoyan Yin, Xiao Hu, Yanjiao Chen, Xu Yuan, and Baochun Li. 2021. SignedPageRank: An Efficient Influence Maximization Framework for Signed Social Networks. *IEEE TKDE* 33, 5 (2021), 2208–2222.

[60] Shuhan Yuan, Xintao Wu, and Yang Xiang. 2017. Sne: signed network embedding. In *PAKDD.*

[61] Weiwei Yuan, Kangya He, Donghai Guan, Li Zhou, and Chenliang Li. 2019. Graph kernel based link prediction for signed social networks. *Information Fusion* 46 (2019), 1–10.

A APPENDIX

A.1 Polarization Scores for the U.S. Congresspeople

Table 2 and Table 3 (both referred in Section 3.2.1) show the polarization scores for the top 20 least and most U.S. congresspeople.

Table 2: Top 20 least polarized members of the U.S. Congress in terms of our random-walk based polarization measure.

Congressperson	State	Party	Score
Henry Cuellar	Texas	D	-0.6542
Jane Harman	California	D	-0.5376
Curt Weldon	Pennsylvania	R	-0.4381
Dutch Ruppersberger	Maryland	D	-0.4318
Jim Moran	Virginia	D	-0.3832
Dave Obey	Wisconsin	D	-0.3588
Wayne Gilchrest	Maryland	R	-0.3503
Duke Cunningham	California	R	-0.3248
Al Edwards	Texas	D	-0.3063
Lincoln Davis	Tennessee	D	-0.2901
Joe Barton	Texas	R	-0.1492
Charles Bass	New Hampshire	R	-0.1381
Mary Bono	California	R	-0.1381
Charlie Dent	Pennsylvania	R	-0.1381
Nick Rahall	West Virginia	D	-0.1357
Bill Young	Florida	R	-0.1250
Roger Wicker	Mississippi	R	-0.1189
Charles Rangel	New York	D	-0.0926
Candice Miller	Michigan	R	-0.0743
Jim Kolbe	Arizona	R	-0.0716

Table 3: Top 20 most polarized members of the U.S. Congress in terms of our random-walk based polarization measure.

Congressperson	State	Party	Score
Steve Buyer	Indiana	R	0.9897
James T. Walsh	New York	R	0.9877
Charles H. Taylor	North Carolina	R	0.9851
Zach Wamp	Tennessee	R	0.9776
Tom Cole	Oklahoma	R	0.9747
Geoff Davis	Kentucky	R	0.9261
Steve Israel	New York	D	0.9019
Marsha Blackburn	Tennessee	R	0.8875
Virginia Foxx	North Carolina	R	0.8814
Steve King	Iowa	R	0.8516
Joe Crowley	New York	D	0.8444
Virgil Goode	Virginia	R	0.8175
Robert E. Cramer	Alabama	D	0.8117
Rubén Hinojosa	Texas	D	0.8030
Carolyn McCarthy	New York	D	0.8030
David Dreier	California	R	0.7874
Christopher Cox	California	R	0.7868
Jo Ann Davis	Virginia	R	0.7867
John E. Peterson	Pennsylvania	R	0.7867
Ray LaHood	Illinois	R	0.7793

A.2 Signed Link Prediction for Baselines with Reconstructed Similarity Ranking

Figure 9 and Figure 10 (both referred in Section 5.1.2) show the signed link prediction performance comparison between POLE and baselines with reconstructed similarity ranking via embedding dot products.

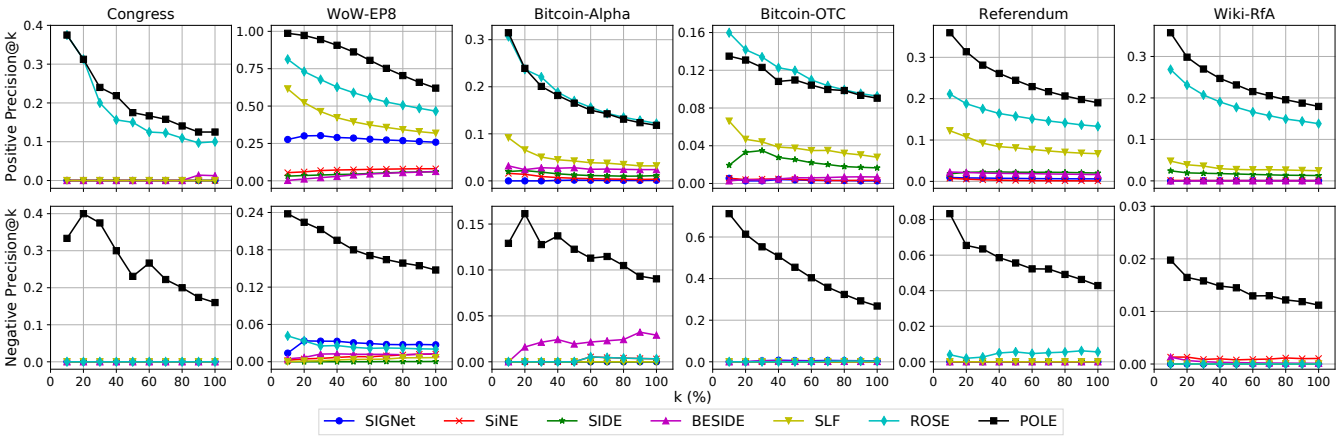


Figure 9: Comparison of performance of signed link prediction between POLE and baselines with reconstructed similarity ranking. POLE outperforms all baselines in almost all datasets on both positive and negative link prediction.

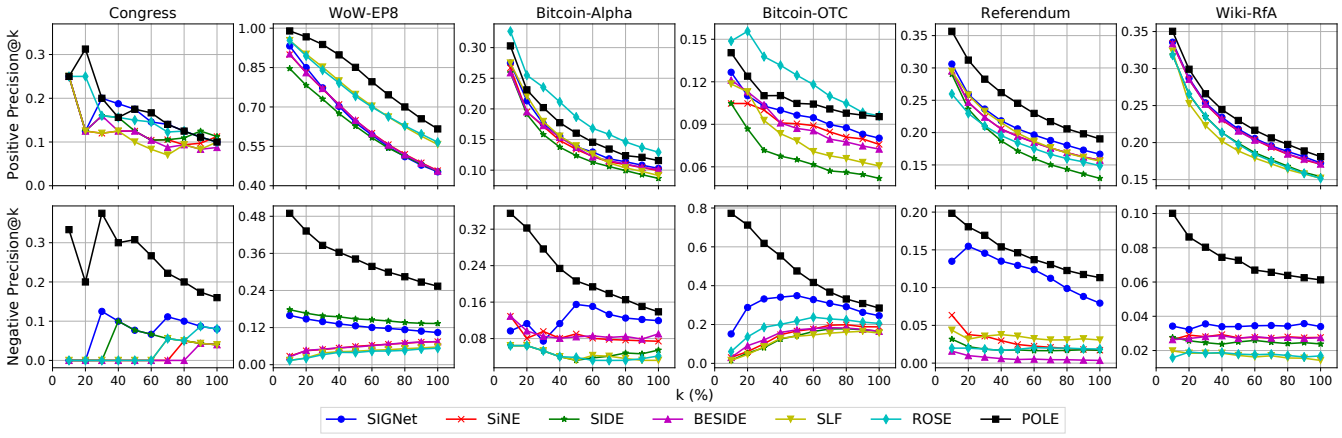


Figure 10: Comparison of performance of signed link prediction with link existence information between POLE and baselines with reconstructed similarity ranking. While adding unsigned similarity narrows the performance gap between POLE and baselines on positive link prediction, it significantly improves POLE on negative link prediction and keeps its edge.

A.3 Scatter Plots of Signed and Unsigned Similarities with Decision Boundaries

Figure 11 (referred in Section 5.2.2) shows the scatter plots of signed and unsigned similarity for different node pairs in signed link prediction along with the decision boundaries for Congress, WoW-EP8, Bitcoin-Alpha, and Bitcoin-OTC, respectively.

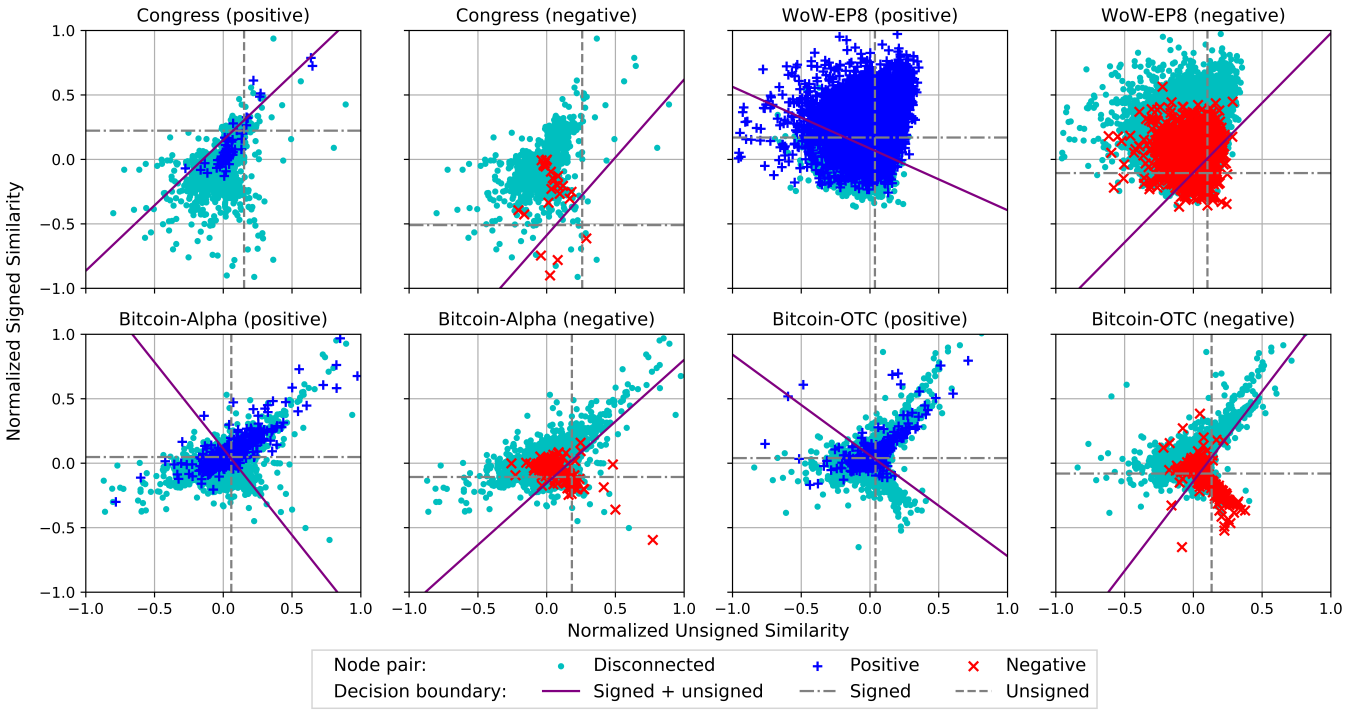


Figure 11: Scatter plot of the reconstructed signed and unsigned similarity for different node pairs in signed link prediction, along with the decision boundaries based on each similarity and a combination of both (via the classifier). Combining signed and unsigned similarity improves prediction for negative links but has a negligible effect on predicting positive links.