ROSE: Role-based Signed Network Embedding

Amin Javari
University of Illinois at Urbana-Champaign
javari2@illinois.edu

Tyler Derr
Michigan State University
derrtyle@msu.edu

Pouya Esmailian
Sharif University of Technology
pouya.esmailian@cs.sharif.edu

Jiliang Tang
Michigan State University
tangjili@msu.edu

Kevin Chen-Chuan Chang
University of Illinois at Urbana-Champaign
kcchang@illinois.edu

ABSTRACT
In real-world networks, nodes might have more than one type of relationship. Signed networks are an important class of such networks consisting of two types of relations: positive and negative. Recently, embedding signed networks has attracted increasing attention and is more challenging than classic networks since nodes are connected by paths with multi-types of links. Existing works capture the complex relationships by relying on social theories. However, this approach has major drawbacks, including the incompleteness/inaccurateness of such theories. Thus, we propose network transformation based embedding to address these shortcomings. The core idea is that rather than directly finding the similarities of two nodes from the complex paths connecting them, we can obtain their similarities through simple paths connecting their different roles. We employ this idea to build our proposed embedding technique that can be described in three steps: (1) the input directed signed network is transformed into an unsigned bipartite network with each node mapped to a set of nodes we denote as role-nodes. Each role-node captures a certain role that a node in the original network plays; (2) the network of role-nodes is embedded; and (3) the original network is encoded by aggregating the embedding vectors of role-nodes. Our experiments show the novel proposed technique substantially outperforms existing models.

CCS CONCEPTS
- Information systems → World Wide Web; Information retrieval; Information systems applications;

KEYWORDS
Signed networks, Embedding, Network Transformation

1 INTRODUCTION
Most existing network embedding methods have been designed for networks with only a single edge type [5] and where relations between two nodes implies closeness. Hence, they primarily try to encode an unsigned network in a way that neighboring nodes are closer in the embedding space [5]. However, real-world networks might have more than one link type, i.e., a network can have \( K \) types of links where each type represents a different quality of relation between the nodes. Signed networks are an important class of such networks, having two types of links: positive and negative [6, 27].

A variety of social media sites, such as Amazon, Wikipedia, and Epinions can be represented as a signed network where positive signs represent trust, agreement or friendship while negative ones may show distrust, disagreement or enmity. The underlying principles of signed networks can be quite different from those of classic networks due to having both positive and negative links. Therefore, network embedding for signed networks cannot be carried out by simply applying classic embedding models. While embedding of signed networks is challenging, it has the potential to greatly advance network analysis tasks such as link/sign prediction [32].

Recently, signed network embedding has attracted increasing attention [4, 8, 19, 31]. Similar to many embedding models for unsigned networks, these models try to embed the network through finding similarities between nodes assuming connecting paths represent closeness. However, in signed networks path-based similarities are challenging since signed paths can indicate either closeness or distantness. Existing works [4, 8, 19] solve this challenge by relying on two signed social theories, namely balance theory [3, 13] and status theory [23]. However, the general architecture on the way they define similarity between two nodes and the use of social theories is associated with two major challenges. First, social theories are incomplete/inaccurate in explaining signed network structure, so models built on them are affected and result in lower quality
embeddings. For example, it has been shown that only around 70% and 65% of triads in real signed networks adhere to status or balance theories [16, 22]. Second, classic embedding models aim to capture presence/absence of links while existing signed embedding models only use two of the possible interaction states: positive and negative links. Thus, since they ignore link absence (the third interaction state), they can not reconstruct the presence/absence of links well, resulting in low performance in link prediction.

To address these shortcomings, we lay out a new perspective for network embedding denoted as network transformation based embedding: if embedding the original network is challenging, it can be transformed into another network for which the embedding task has lower complexity. The transformation can be done by mapping each node in the original network to multiple nodes in the transformed network. Next, the transformed network can be embedded. Finally, the embedding vectors obtained from the transformed network can be aggregated to encode the original network. More specifically, to embed signed networks, we introduce a ROle based Signed network Embedding (ROSE) that bypasses the aforementioned challenges. The underlying idea is to transform the signed network into a bipartite network where each node takes both “user” and “item” roles for which they are the giver and receiver of signed links, respectively. Therefore, each node of a signed network can be modeled by a set of roles, denoted as role-nodes, where the relations between role-nodes can be fully captured using unsigned links. Then, ultimately this transformed network can utilize the state of the art unsigned embedding technique. Fig. 1 is a toy example of the transformation process.

Each role-node captures a certain aspect of a node in the original network. Hence, a comprehensive embedding of a node can be obtained by aggregating the embeddings of their role-nodes. We introduce two aggregation methods, denoted as fixed aggregation and target-aware aggregation. The fixed aggregation simply concatenates all the role-node embeddings together. The target-aware aggregation is based on a recent deep learning based recommendation model that introduced a model for target dependent encodings of users [34]. Based on this idea, we propose an attention mechanism based model to aggregate the embeddings of role-nodes in which attendance weights are obtained with respect to the target entity. To the best of our knowledge, this is the first work to build target aware embeddings of nodes in a signed network.

We evaluated the proposed embedding model based on sign prediction as well as link prediction tasks. The evaluations were carried out on three real-world datasets: Epinions, Wikipedia, and Slashdot. The results of the experiments confirm that the proposed model significantly outperforms all of the existing methods. Additionally, we show that ROSE has higher flexibility and generalizability compared to the existing models. In summary, the major contribution of the proposed embedding framework is three folds:

1) We introduce for the first time the general idea of network transformation based network embedding.

2) We propose a novel dedicated network transformation methodology to embed signed networks by transforming them into a bipartite unsigned network that can then utilize any existing state of the art unsigned embedding method to then ultimately obtain node representations for the original signed network.

3) We present the first target aware signed network embedding through our proposed attention mechanisms.

2 PRELIMINARIES AND RELATED WORKS

2.0.1 Problem Formulation. Let a graph be defined as $G(V, E)$ with a link type mapping function $\phi : E \rightarrow A$ where $V$ represents the nodes, $E$ represents the links, and each link $e \in E$ belongs to a link type $\phi(e) \in A$. In unsigned networks $A$ has only one values, but signed networks have two values: positive and negative. Given the graph $G$, the task of node encoding is to learn a function $f : V \rightarrow |d|$ that maps each node $v$ to a $d$-dimensional embedding vector which can be parametrized by the Matrix $W$ with size $|V| \times d$.

2.0.2 Unsigned network Embedding. Embedding models can be described as an encoding-decoding framework [12] having four components: 1) A pairwise node similarity function. 2) An encoder function to create embeddings from the similarity function. 3) A decoding function to recover the pairwise node similarities from their embeddings. 4) A loss function that evaluates the reconstructed similarity values. The primary difference in the literature is how the embedding methods define node similarity. However, the shared principle in unsigned similarity functions is that an unsigned path between two nodes indicates their closeness (e.g., as in [10, 26, 28]).

2.0.3 Signed network Embedding. The unsigned similarity functions cannot be directly applied to signed networks because negative edges do not represent closeness. Thus, the challenge in embedding signed networks is how to involve negative edges without hindering positive proximity. Existing methods have sought to capture node similarity using the paths between them at length one, two, or higher order paths.

Single-length paths: A trivial approach to embed signed networks while involving negative links is to embed a node based on its immediate neighbors [14, 29]. This, however, has limited effectiveness because it cannot capture the higher order proximities between nodes. However, capturing global structures in signed networks is challenging, e.g., given a path containing positive links and negative links, how can the similarity of the nodes be defined? Does the path indicate closeness or distantness between the nodes? To aid this previous works used the two signed social theories, namely balance and status, which we define below.

Balance theory [3, 13] defines a cycle to be balanced if there exist an even number of negative links, and unbalanced otherwise. Two such examples in triangles are that of, “A friend of my friend is my friend.” and “An enemy of my friend is my enemy.” Status theory [23] in social science states that “the person respected by me should have higher status than me”. Thus, in a signed network, a positive link from node $u$ to $v$ represents the higher status of $u$ than $v$, while a negative link shows the higher status of $u$.

Paths of length two: As the above social theories are formed on triangle structures, existing methods [4, 32] have used them to determine whether the signed path is representing closeness or distantness between two nodes. However, these methods do not go beyond paths of length two. As such, they have limited power in capturing global structures.

Longer paths: More recent work has tried to capture longer cycle paths in their embedding process mainly by relying on the extended version of social theories. [19, 35] both run a random walk
on signed networks similar to node2vec algorithm, [9] applied to node relevance and personalized ranking [18] using random walks. Then, a graph convolutional network method for embedding signed networks has been introduced which relies on balance theory [8].

In all, the shared strategy of all these works is that they embed nodes by analyzing the paths between them. If a path indicates closeness, they embed the nodes closer and distance them otherwise. However, to interpret if a path indicates closeness or distantness, they exploit some strong assumptions which naturally induce noise to the embedding process. Also, this strategy does not use a principled way to distant nodes based on the absence of links/paths between them, i.e., it only focuses on capturing positive/negative paths.

3 PROPOSED FRAMEWORK: ROSE

In this section, we describe the structure of ROSE. Based on the drawbacks of the previous works, we outline the following requirements: an effective universal network embedding model should be able to 1) capture the higher order connectivity between nodes, 2) take into account the link labels as well as the link structures (presence or absence of links), and 3) do not make assumptions about the origin of the network.

Network transformation based embedding. To address the requirements, we introduce the general notion of network transformation based embedding. Rather than directly finding the similarities of the nodes in the input network, it can be transformed to another network in which we do not encounter the embedding challenges present in the input network. One possible way to do transformation is to define different roles for a node, denoted as “role-nodes”, and build a network of role-nodes in a way that the similarities between role-nodes can be determined by adopting the classic similarity functions. Since each role-node captures a certain aspect of an original node, the embedding vector of a target node can be derived by aggregating the embeddings of the corresponding role-node. In sum, a network transformation based embedding model can be described in three main steps: 1) Network transformation, 2) Embedding the transformed network, 3) Embedding the original network by aggregating the embeddings of the transformed network. By relying on the general idea of network transformation, we propose ROSE. In the following, ROSE is described based on the aforementioned three-step architecture. We then illustrate how ROSE addresses the requirements of the problem.

3.1 Network transformation

The way the role nodes are defined is fundamental to the effectiveness of ROSE. We aim to define the transformation such that similarities between role-nodes can be obtained using classical methods. Note that there can be multiple ways to define the transformation process. Various embedding techniques have been introduced based on different similarity measures. Similarly, different embedding methods can be developed based on the idea of transformation based embedding by creating different transformations. Our transformation idea is inspired by recommender systems. Traditionally, user-item interactions in recommender systems are modeled by a user-item bipartite network. A signed all-to-all connected network can also be viewed as a bipartite network where each node plays a “user” role for the links it receives and plays an “item” role for the links it receives. Based on this analogy, we capture user/item roles of a node separately through a transformation process.

Step 1: Transformation to a bipartite network. Based on user-item analogy, each node is mapped to two role-nodes, i.e., the node u is mapped to the role-nodes u_out and u_in where a link from u to v in the original network is modeled as an undirected link between u_out and v_in. As it can be seen in Fig. 2(a), the input network is transformed to a signed bipartite network [7] with two types of nodes “in” and “out”. However, applying a classic similarity measure on the transformed network is still a challenge due to presence of positive and negative links.

Step 2: Transformation to an unsigned network. We transform the network into an unsigned network by defining new role-nodes. A role-node of type “in” u_in is mapped into two role-nodes: v_in^+ (or v_in^-) representing its role when positive (or negative) links point toward it. Accordingly, a link from u_out to v_in with label l is modeled as an unlabeled, undirected link between u_out and l_in^l. This enables us to use the well-established similarity functions to determine the similarities of role-nodes.

Step 3: Augmenting the network. Our strategy is to encode the original network by embedding the transformed network. However, some of the role-nodes may have a very low degree. In particular, role-nodes of type “in” tend to have a very low degree due to the fact that the number of negative links is often under-represented compared to positive links. According to our results, this can dramatically hinder the accurate embedding of such role-nodes. To solve this, we leverage implicit knowledge about the problem domain. If node u_out has connections towards both v_in and w_in, not only it reflects the adjacency between these two role-nodes but also it implies dependence between their opposite role-nodes: v_in and w_in. To bring this knowledge into our embedding process, which can attenuate the sparsity problem, we augment our unsigned network with a set of dummy nodes of type “out”, i.e., for each node of type “out” in the unsigned network with the set of connections \( \{v_1^{l_1}, v_2^{l_2}, ..., v_m^{l_m}\} \), we add a node of type “out-dummy” with the set of connections \( \{v_1^{l_1}, v_2^{l_2}, ..., v_m^{l_m}\} \) where \( l' \) is the inverse of \( l \), i.e., if \( l = + \), \( l' = - \) and vice versa.

Summary of transformation. In sum, \( G(V, E) \) is transformed to a bipartite unsigned graph \( G_u(V_u, E_u) \) where \( |V_u| = 4|V|, |E_u| = 2|E| \), i.e., a node \( u \in V \) is mapped to four role-nodes in \( V_u \): 1) u_out which initiates a link, 2) u_out^- which initiates a dummy link, 3) u_in which receives a positive link, and 4) u_in^- which receives a negative link. And a link \( e_{uv} \) with the label \( l \) is transformed to links \( e_{u_out,v_in}^{l}, e_{u_out,v_in^-}^{l}, e_{u_in,v_in}^{l}, e_{u_in,v_in^-}^{l} \). Fig. 2 depicts a toy example describing the process. It should be noted that the transformation is lossless, i.e., \( G(V, E) \) can be fully reconstructed from \( G_u(V_u, E_u) \).

3.2 Embedding the original network

Analogous to unsigned networks, the links between role-nodes indicates their closeness. Hence, a classic embedding model can be used to embed role-nodes. We employ node2vec [10]. Note that more advanced embedding models [5, 33] can be used/designed for this purpose. However, in this paper, our focus is on introducing the structure of ROSE. Once having the embeddings of role-nodes, a node’s embedding is created by aggregating the embeddings of
the corresponding role-nodes. In the following, we introduce two different aggregation models.

**Fixed Aggregation.** Each of the roles represents a certain perspective/role of a node in the original network. In general, if there are multiple representations of an entity, we can concatenate them or linearly combine them to build a unified representation. Accordingly, a straightforward way to build a comprehensive and unified embedding of a node is to concatenate the embedding vectors of the corresponding role-nodes. As such, the fixed representation of node \( u \) can be defined as \( W_u = W_{\text{in}}u || W_{\text{out}}u \) in which \( || \) represents concatenation. Note that dummy role-nodes are not used in the aggregation process. In fact, "out-dummy" role nodes are inverses of the role-nodes of type "out" and do not add extra knowledge about the representations of the original nodes.

**Target Aware Aggregation.** One important application of graph embedding is to use the embedding vectors to predict the pairwise interactions of nodes, e.g., link prediction. Intuitively, for such tasks, it is more accurate to encode the given initiator node with respect to the target entity. In fact, the idea of target-aware profiling is the basis for most of the recommendation models. For example, in item-based collaborative filtering, to predict the rating of a user towards an item, her previous ratings are aggregated in a weighted way because not all of her interactions are equally important in reflecting her taste towards the item [36]. Typically, the weight of a rating is determined based on the similarity of the corresponding item to the target item [34].

Inspired by recommender systems, we introduce a target-aware embedding technique by proposing a target-aware aggregation model. To the best of our knowledge, the existing techniques build only fixed embeddings. In our framework, intuitively predicting the pairwise interaction from \( u \) to \( v \) depends on the "out" role node of \( u \) and "in" role nodes of \( v \). We propose that "out" role-node of \( u \) can be embedded according to \( v \) which is denoted by \( W_{\text{out}}u \). Having \( W_{\text{in}}u \), the target dependent embedding of \( u \) w.r.t. \( v \) is defined as \( W_u = W_{\text{in}}u || W_{\text{out}}u \). Indeed, we concatenate the fixed embedding of \( u \) with a component that depends on the target entity to build its target-aware embedding. To build \( W_{\text{out}}u \), we design an attention mechanism. We suggest that \( W_{\text{out}}u \) can be obtained by attending to the neighbors of \( u \) based on their relevancy to the target entity \( v \). More formally, \( W_{\text{out}}u \) is defined as:

\[
W_{\text{out}}u = \sum_{s_{\text{in}}' \in N(u)} [e(s_{\text{in}}', v)] W_{s_{\text{in}}'}
\]

where \( e(s_{\text{in}}', v) \) is the importance weight of \( s_{\text{in}} \) w.r.t. \( v \) and \( N(u) \) is the set of role-nodes connected to \( u \). To estimate the importance weights, we introduce an unsupervised attention model. The intuition behind the model is that the "in" role-nodes of two nodes are more related if the are more tightly connected in the network. Note that we do not take into account the labels of connections to find the relevancy of two nodes. To systematically implement the idea, given the target signed network, we assume the links are unsigned and transform it to a bipartite network where the obtained network has two types of role-nodes: "in" and "out". Next, the transformed network is embedded using node2vec. Finally, \( \text{att}(s_{\text{in}}', v) \) is defined as follows:

\[
\text{att}(s_{\text{in}}', v) = \sigma(W_{s_{\text{in}}}, W_{v}) = \frac{1}{1 + \exp(-W_{s_{\text{in}}}, W_{v})}
\]

In fact, this weight determines how tightly \( s \) and \( v \) are connected in terms of the nodes that rated them regardless of the rating values.

### 3.3 Model Justification

We outline three main requirements for an effective signed embedding technique. ROSE fulfills the requirements. 1) Unlike existing models, ROSE does not rely on any assumption about the origin of the network. 2) To obtain the embedding of role-nodes, we use a random walk based model to ensure the obtained embeddings capture the higher order proximities. 3) The model preserves both link labels and link structures. To address the sign/link prediction tasks the embeddings from ROSE can be fed to a nonlinear function trained by a method like MLP to determine the target interaction state. In fact, the embeddings of role-nodes contain major patterns that can aid to fully reconstruct the graph. We encode the role-nodes in a way that if a link with label \( l \) exists from \( u \) to \( v \), \( W_{\text{in}}u \) has higher proximity to \( W_{l_{\text{in}}}v \) than \( W_{l_{\text{out}}}v \). And if there is no link from \( u \) to \( v \), \( W_{\text{out}}u \) is expected to have low proximities to both \( W_{l_{\text{in}}}v \) and \( W_{l_{\text{out}}}v \). As such, having a function to find the similarities of embeddings, the label of the link from \( u \) to \( v \) is expected to be \( l \) if the proximity of \( W_{\text{in}}u \) to \( W_{l_{\text{in}}}v \) is greater than it’s proximity to \( W_{l_{\text{out}}}v \). And if \( W_{\text{out}}u \) has low similarity to both \( W_{l_{\text{in}}}v \) and \( W_{l_{\text{out}}}v \), it indicates absence of link. Moreover, we observe other interesting patterns in our experiments, which will be discussed in the experiments section. In addition to addressing the requirements of the problem, the proposed framework creates an avenue to make a connection between the recommender systems and signed networks contexts. Lastly, the proposed model is quite generalizable. Since the model
does not rely on any assumption specific for signed networks, it can be generalized for networks with multi-type of links.

4 EXPERIMENTS

We conducted experiments to verify the effectiveness of the proposed framework and the ideas behind the model. The experiments are focused on answering two key questions:

- **RQ1:** How do the proposed embedding frameworks perform when compared to the state of the art models in terms of link-label prediction and link prediction tasks?
- **RQ2:** What is the interpretation of the embeddings obtained from the network of role-nodes?

**Datasets:** Three real-world datasets were used in the experiments: Epinions [11], WikiElection [2], and Slashdot [21] which have been used in previous works [27]. WikiElection dataset is the signed network obtained from users’ votes for elections of administrators. Epinions was an online product review site. Users can express positive or negative votes to other users regarding the trustworthiness of their reviews; this dataset is from the positive/negative votes between users. Slashdot: Slashdot dataset is also obtained from an online service (technology news website) where users can share comments and flag each other as friend or foe. The flags indicate approval or disapproval of comments. Analogous to Epinions, Slashdot dataset models the interactions of users using a signed network. Statistics of the datasets are given in Tab. 1.

### 4.1 Performance of the proposed model (RQ1)

In this experiment, we compared the performance of ROSE with four recently introduced signed network embedding models on two tasks: sign prediction and link prediction. Moreover, we compared the target dependent variant of ROSE denoted as ROSE-UAT with the fixed variant represented by ROSE on both of the tasks. AUC (Area Under the Curve) was used as the evaluation metric.

The following is the list of the models used in this experiment.

- **SIDE** is a random walk based approach that aims to capture global structures in the embedding process [19]. BESIDE aims to use both balance and status theories in a complementary manner to encode signed networks[4]. SiNE is a deep learning based framework that performs based on undirected networks. The main principle behind the model is that “users should sit closer to their friends than their foes” [32]. SIGNet is also a random-walk based model that maintains structural balance using targeted negative sampling. [14].

**Evaluation:** In our datasets, the number of negative edges is much smaller than the positive links. Thus, comparing methods based on their original test set accuracy could be misleading, especially for sign prediction. Thus, analogous to previous works [4, 22], we balanced the datasets by randomly removing positive links and used 5-fold cross-validation for our experiments. The baselines were evaluated based on the source-codes released by their authors. The embedding dimension for all of the models was set to 30.

**Sign Prediction:** Sign prediction [11, 15] is the major task that has been used to evaluate encoding models in previous works. The Tab. 2 shows the AUC of the models on three datasets. First, we notice ROSE-UAT and ROSE outperform all the baselines. For example, ROSE-UAT outperforms BESIDE by 1.6%, 0.8%, and 2% in terms of AUC on WikiElection, Slashdot, and Epinions datasets, respectively. The higher accuracy of ROSE can be attributed to its effectiveness in addressing the requirements of the problem. Additionally, we observe ROSE-UAT perform better than ROSE. In fact, encoding the nodes with respect to a target entity helps to better analyze the interactions of the node and the entity.

**Link Prediction:** Although link prediction is an important task in network mining [24], previous works have not evaluated their models based on link prediction task. To evaluate the models for the link prediction task, we first fed the training graph to the models and obtained the node encodings. Next, we created training and test sets. Each data instance in the training/test set is the concatenation of the encoding vectors of a node pair \((W_u, W_v)\) and the label of the instance is 1 if there is a link from \(u\) to \(v\) and 0 otherwise. In both training and test sets, 50% of instances have label 1 and 50% of them have label 0. The node pairs with 0 label were randomly selected. The training set obtained from each embedding method was fed to a multi-layer perceptron classifier, and the AUC of the trained model was obtained based on the test set. Tab. 2 shows the results of the experiments. As it can be seen, ROSE has superior performance than the baseline models on all but one of the datasets since the model systematically differentiates the three different interaction-states between nodes in its embedding process.

### 4.2 Interpretation of the encodings of role-nodes (RQ2)

Given nodes \(u\) as initiator and \(v\) as receiver, three interaction-states can be considered between them: absence of link, positive link and negative link. We introduced the major pattern extracted from the distance/similarity of the encoding vectors of \(u\) and \(v\) that can aid to determine the interaction type between them. We investigated the existence of such patterns. Fig. 3 shows the average distance of different encoding components of a node pair \((u, v)\) as a function of the interaction-state between them. For example, in Slashdot dataset if the link is positive the average distance of \(W_{\text{out}}\) from \(W_{\text{ip}}\) denoted as \(d_{\text{avg}}(u_{\text{out}}, v_{\text{ip}})\) is 1.2. The average distances of the components are consistent with the introduced major patterns. If the link from \(u\) to \(v\) is positive, \(d(u_{\text{out}}, v_{\text{in}})\) is expected to be smaller than \(d(u_{\text{out}}, v_{\text{ip}})\). For example, for a positive link in Epinions, \(d_{\text{avg}}(u_{\text{out}}, v_{\text{ip}})\) is 1.1 while \(d_{\text{avg}}(u_{\text{out}}, v_{\text{in}})\) is 1.7. Moreover, we observe that if there is a link from \(u\) to \(v\), \(d_{\text{avg}}(u_{\text{out}}, v_{\text{in}}) + d_{\text{avg}}(u_{\text{out}}, v_{\text{ip}})\) is smaller than the case when there is no link. For example, in Slashdot, \(d_{\text{avg}}(u_{\text{out}}, v_{\text{ip}}) + d_{\text{avg}}(u_{\text{out}}, v_{\text{in}})\) is 3.3 if there is no link from \(u\) to \(v\) and it is 2.75 if there is a link. In addition to the major pattern, we observe four other patterns. We name these patterns as implicit patterns because our model has not targeted to extract them.

**First,** if the sign of the link from \(u\) to \(v\) is positive, similar nodes rate them similarly and if it is negative, similar nodes rate them with the help of their mutual friends.

### Table 1: Datasets statistics.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Nodes</th>
<th>Edges</th>
<th>+Edges</th>
<th>±Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>WikiElection</td>
<td>7118</td>
<td>10347</td>
<td>78.7%</td>
<td>21.2%</td>
</tr>
<tr>
<td>Epinions</td>
<td>119217</td>
<td>841200</td>
<td>85.0%</td>
<td>15.0%</td>
</tr>
<tr>
<td>Slashdot</td>
<td>82144</td>
<td>549202</td>
<td>77.4%</td>
<td>22.6%</td>
</tr>
</tbody>
</table>
Table 2: AUC of the proposed model (ROSE) and the baseline methods on the WikiElection, Slashdot and Epinions datasets.

<table>
<thead>
<tr>
<th>Model</th>
<th>WikiElection</th>
<th>Slashdot</th>
<th>Epinions</th>
<th>WikiElection</th>
<th>Slashdot</th>
<th>Epinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIDE</td>
<td>0.7986</td>
<td>0.8815</td>
<td>0.8672</td>
<td>0.9184</td>
<td>0.9342</td>
<td>0.9314</td>
</tr>
<tr>
<td>BESIDE</td>
<td>0.8953</td>
<td>0.9012</td>
<td>0.9342</td>
<td>0.9092</td>
<td>0.9265</td>
<td>0.9397</td>
</tr>
<tr>
<td>SINE</td>
<td>0.8632</td>
<td>0.8680</td>
<td>0.8543</td>
<td>0.5833</td>
<td>0.5983</td>
<td>0.6488</td>
</tr>
<tr>
<td>SIGNET</td>
<td>0.8943</td>
<td>0.8997</td>
<td>0.9181</td>
<td>0.9099</td>
<td>0.8862</td>
<td>0.9205</td>
</tr>
<tr>
<td>ROSE</td>
<td>0.9091</td>
<td>0.9082</td>
<td>0.9533</td>
<td>0.7418</td>
<td>0.9357</td>
<td>0.9403</td>
</tr>
<tr>
<td>ROSE-UAT</td>
<td>0.9116</td>
<td>0.9095</td>
<td>0.9547</td>
<td>0.9426</td>
<td>0.9391</td>
<td>0.9444</td>
</tr>
</tbody>
</table>

Figure 3: The average pairwise distance of the encoding vectors of the role-nodes of a node pair \( (u, v) \) for different interaction-types between them: positive link, negative link, and absence of a link.

different signs. In fact, the smaller distance between the embeddings of the role-nodes of type “in+” and of type “in-” of two nodes means that they were rated by similar nodes similarly. For example, in Epinions dataset \( d_{\text{avg}}(u_{\text{in+}}, v_{\text{in+}}) \) and \( d_{\text{avg}}(u_{\text{in-}}, v_{\text{in-}}) \) are 1.1 and 1.3 respectively when the sign of the edge from \( u \) to \( v \) is positive while those distances are 1.7 and 1.6 respectively when the edge sign is negative. It can be said this pattern is aligned with balance theory, i.e., the triangle structures described in balance theory can be regarded as a special case of this pattern [20]. Second, \( u \) and \( v \) rate similar nodes more similarly when there is a positive link between them than when there is a negative a link connecting them. The smaller distance values between the embeddings of the role-nodes of type “out” of two nodes indicate that they have rated similar nodes similarly. As it can be seen, \( d_{\text{avg}}(u_{\text{out}}, v_{\text{out}}) \) is smaller when there is a positive link from \( u \) to \( v \). Again, balance theory can be regarded as the special case of this pattern. Third, the signs of the link between two nodes in different directions are correlated. \( d_{\text{avg}}(v_{\text{out}}, u_{\text{in+}}) \) is smaller when there is a positive link from \( u \) to \( v \) than when there is a negative link. This pattern is contradictory to status theory. Fourth, the average distance between the embeddings of the role-nodes of two nodes is quite larger when there is no link between them than when there is a link. A large distance between the embeddings of two nodes implies they are not tightly connected and belong to different clusters.

5 CONCLUSION

In general, existing models for embedding signed networks build their models upon path-based similarity measures where social theories are used to define such similarity measures. However, this structure to build embedding models faces major challenges. 1) Social theories do not accurately explain the structure of signed networks. 2) The models built based on this structure mainly focus on link labels and neglect capturing presence/absence of links. In order to address these challenges, we introduced a novel network transformation based embedding framework denoted as ROSE, which relies on transforming the original network to an unsigned bipartite network. To the best of our knowledge, this is the first paper to propose the idea of network transformation based embedding. Also, it is capable of encoding nodes with respect to a target entity. Our experiments confirmed that the model outperforms the state of the art models. Moreover, the model is generalizable to networks with more than two types of connections. We plan to investigate ideas of how to use network transformation for other graph types, e.g., knowledge graphs [1, 17, 25, 30], so that they can also benefit from the state of the art embedding methods being developed for simple graphs.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF) IIS 16-19302 and IIS 16-33755, Zhejiang University ZJU Research 083650, Futurewei Technologies HF2017060011 and 094013, UIUC OVCR CCIL Planning Grant 434534, UIUC CSBS Small Grant 434C8U, and Advanced Digital Sciences Center Faculty Grant. In addition, Tyler Derr and Jiliang Tang are supported by NSF IIS 1845081. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the funding agencies.
REFERENCES


