CS-Trust2Vec: Context-Sensitive Trust Network Embedding Model

Onat Özdemir | Advisors: Prof. Dr. Pınar Karagöz & Prof. Dr. İsmail Hakkı Toroslu

Abstract

Trust network embedding is a network modeling task aiming to map nodes in social trust networks to low dimensional vector space. Although there are trust network embedding models proposed in the literature, these models focus on the structural properties of the networks and disregard additional supportive information such as the context of trust. Context-specificity is an often-mentioned property of trust, and it has been studied well in sociological and psychological terms. However, it has not been elaborated on within computational trust modeling. In this paper, we propose a novel trust network embedding model (CS-Trust2Vec) that integrates the context information into the embedding process. Due to the lack of a dataset containing contextual information, we created a context-labeled trust network dataset by scraping a Turkish social media platform. We validate the efficiency and accuracy of our model on the link prediction task using the dataset.

Motivation

A context of trust defines on which topic a person trusts another person. Although the context of trust has been studied well in sociological terms, to the best of our knowledge, it has not been studied in computational trust modeling. In this project, our motivation is to solve the trust embedding problem using a context-sensitive trust embedding model and investigate the impacts of context to fill the mentioned gap in the literature.

Conclusion

In this study, we propose a novel trust network embedding model that integrates the context information of trust relationships into the embedding process and investigate the impacts of context information on the trust network embedding task. We validate the accuracy of our model on the link prediction task. Experiments indicate that using the context information in trust embedding models remarkably increases the accuracies of the models. In future work, we plan to extend our model for signed trust networks.

References

[1] Dorwin Cartwright and Frank Harary. Structural balance: a generalization of heider's theory. Psychological review, pages 277–93, 1956.

[2] Aditya Grover and Jure Leskovec. Node2vec: Scalable feature learning for networks. 2016. doi: 10.1145/2939672.2939754.

Proposed Model

Our model consists of three modules: Preprocessor, Encoder, and Link Predictor. The architecture diagram of the model can be seen in Figure 1.

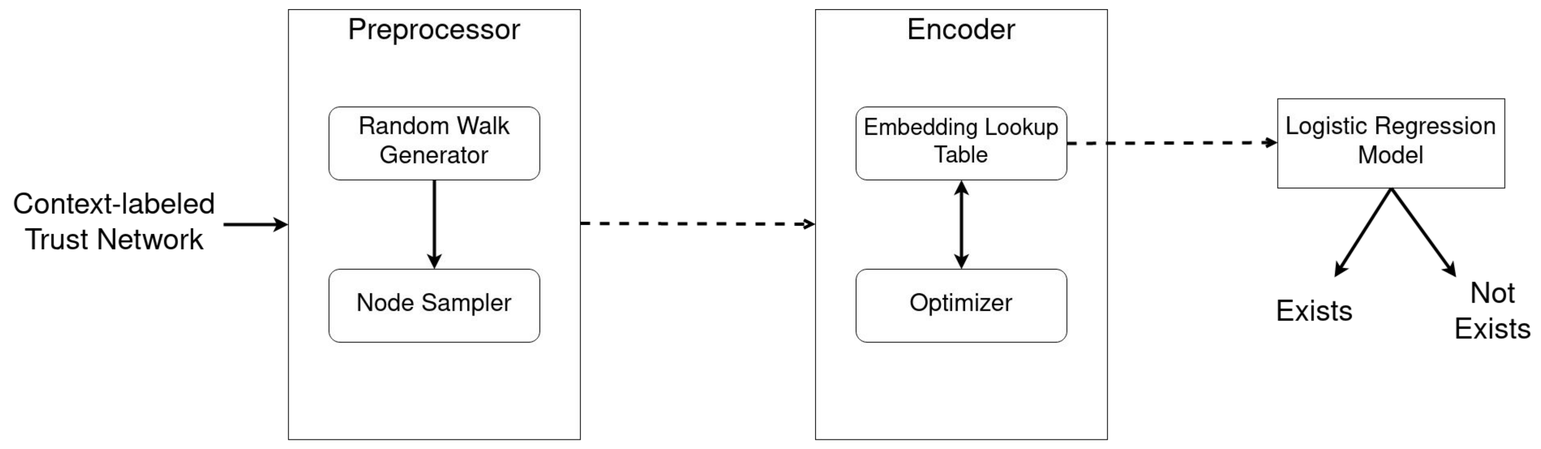

Figure 1: Architecture of the CS-Trust2Vec

Preprocessor

Random Walk Generator: To generate training samples, we propose a biased random walk strategy. The first edge in each walk is randomly chosen. The following edges are randomly selected from a non-uniform probability distribution whose weights are determined by the similarities between the context label of the previous edge and labels of the candidate edges.

Node Sampler: The Node Sampler generates positive and negative samples using random walks. Even if there is no edge between two nodes, if they are in the same random walk chain, we can claim that they are positively related using the Structural Balance Theorem [1]. For each random walk, a sliding context window moves along the random walk, one node at a time. In each step, the set of nodes covered by the context window is added to the positive sample set of the target node.

Encoder

Embedding Lookup Table: Embeddings of the nodes in the network are stored in a matrix named Embedding Lookup Table. Each node embedding consists of four different features;

$$L_u^{in} \;\; L_u^{out} \;\; CT_u^{in} \;\; CT_u^{out}$$

Optimizer: The model uses the function

$$f(u,v) = L_u^{out} \cdot L_v^{in} + (CT_u^{out} + CT_v^{in}) \cdot \phi_{uv}$$

to measure the node u's tendency to trust node v.

Objective function for the (u, v) pair is defined as,

$$max \; J_{(u,v)} = h(u,v) + \sum_{i=1}^{n_{neg}} h'(u, v'^{(i)})$$

where

$$h(u,v) = log\sigma(f(u,v))$$
$$h'(u,v') = log\sigma(-f(u,v'))$$

Link Predictor

The Link Predictor module contains a logistic regression model to predict whether there exists a trust relationship between the given node pair (u, v) or not. Formally, the Link Predictor can be seen as a function $f : \mathbb{R}^{2d} \rightarrow \{0, 1\}$. For a given node pair (u, v), $f(z_u \oplus z_v)$ is defined as,

$$f(z_u \oplus z_v) = \begin{cases} 0, & \text{predicts } (u,v) \notin \mathcal{E} \\ 1, & \text{predicts } (u,v) \in \mathcal{E} \end{cases}$$

where $\oplus$ is the concatenation operation.

Experiment Results

The following models are used as baselines for the experiment;

- **Node2vec:** Node2vec [2] is a skip-gram based graph embedding model. It uses a biased random walk strategy to generate node embeddings.
- **Trust2Vec:** Trust2Vec is a non context-sensitive variation of the CS-Trust2Vec. Trust2Vec generates random walks using uniform distribution and does not have CT features.

We test the models on the link prediction test using the Ekşi Sözlük dataset. Results can be seen in Table 3.

Dimension of L	96
Dimension of CT	56
Window length	5
Maximum walk length	10
# of walks	20
# of negative samples	3
Similarity measure	cosine similarity

Table 1: Parameter settings

# of Users	10069
# of Relationships	73468
# of Categories	28

Table 2: Statistics of the network

	Node2vec	CS-Trust2Vec	Trust2Vec
F1-Macro	0.615	**0.759**	0.744
AUC	0.654	**0.831**	0.820

Table 3: Link prediction results