Predicting polarisation of dynamic social networks via graph auto-encoders

Keywords: Dynamic social networks, Network heterogeneity, Polarisation prediction, Signed graph clustering, Graph auto-encoders.

Introduction

Signed graphs have been used to model interactions in social networks, where positive edges represent agreement between users and negative edges indicate disagreement. Finding polarised communities in a social network can be formulated as the problem of signed graph clustering (SGC) where intra-cluster edges are mostly positive and inter-cluster edges mostly negative [1]. Once the communities are identified, various measurements, e.g., balance-based measure [2], can be applied to calculate the polarity score (PS) of the signed graph. However, current models only incorporate the graph structure in the computation of the PS and overlook another key input information, namely potential node features (e.g., user characteristics, or users' comments). Additionally, existing studies mostly focus on analyzing static graphs and do not take into account the dynamic nature of online discussions, such as the involvement of new users or changing intensity of discussions, both of which may lead to change in the PS score overtime. Predicting the future PS would provide valuable insights into public opinions and sentiment towards certain topics or events, which is in turn essential for early intervention against harmful trends in online social communities. In light of the aforementioned motivations, we propose to predict the PS of the signed dynamic social graphs using graph auto-encoders (GAE). The heterogeneity between graphs in terms of both graph structure and node features has a direct impact on the PS and should be considered when forecasting future PS. To quantify the heterogeneity, we propose to train a GAE on G_t (the graph observed at time instant t) via a synthetic reconstruction task, and apply the trained model to reconstruct G_{t+1}. The difference (d_{t+1}) between the reconstruction accuracy of G_t and G_{t+1} can then serve as an indicator of the heterogeneity between graphs. Finally, we predict the future PS by considering both the historical PS and the heterogeneity between graphs.

Methodology

There are three key steps in our framework as shown in Fig.1: 1) calculating the PS of signed graphs using SGC, 2) computing the difference in reconstruction accuracy (d_{t+1}) between the trained graph (G_t) at time t and test graph (G_{t+1}) at $t+1$ with GAE, and 3) using the historical PS and reconstruction accuracy difference to predict the future PS.

SGC aims to find a partition of nodes into two (or more) clusters, with the goal of maximising the number of positive edges within clusters and negative edges between them. In real-world applications, we may expect two polarised communities on a topic while other users remain neutral. Thus, we apply Eigensign [2] to discover polarised communities among other neutral nodes. This can be achieved by finding a partition $\{S_0, S_1, S_2\}$ of the node set V (represented by $\mathbf{c} \in \{0, -1, 1\}^N$) that maximises the PS of G, i.e., $\frac{\mathbf{c}^T\mathbf{A}\mathbf{c}}{\mathbf{c}^T\mathbf{c}}$ where $\mathbf{A}$ is the adjacency matrix.

GAE [3] is an unsupervised graph representation learning framework with an encoder and decoder as shown in Fig.2. The encoder maps the input graph structure $\mathbf{A}$ and node feature matrix $\mathbf{X}$ to the latent embedding $\mathbf{Z}$ using a relational graph convolutional network (RGCN) [5] which applies different transformation mapping matrices depending on the type of edges. The decoder is then a 3-class edge classifier: given embeddings of a node pair $\mathbf{z}_i$ and $\mathbf{z}_j$, we concatenate them and feed it to a multilayer perceptron (MLP) to predict the edge type. Training is done on a subset of training edges using the cross-entropy loss, and the model is then applied to predict/reconstruct the edges in a test set.

Given a fix-length time window, we first construct the graph G_t, and then train the GAE on a subset of edges in G_t. Once trained, the GAE can be applied to reconstruct the remaining edges in G_t, or edges in the unobserved graph G_{t+1}. The reconstruction accuracy will be high if the graph structure is spatially homogeneous or similar between the training and test sets. Consequently, the decrease in GAE testing accuracy (d_{t+1}) between G_t and G_{t+1} can be viewed as a measure of increase in graph heterogeneity or dissimilarity. In this work, we assume that such change would influence the polarisation trend in the social network and therefore can be useful in predicting the future PS at $t+2$.

Predicting PS. We propose to use an order-p autoregressive (AR) model to predict the future PS. Specifically, the PS at time $t+2$ is:

$$ps_{t+2} = \sum_{i=1}^{p} \phi_i ps_{t+2-i} + \alpha_i d_{t+1}, \tag{1}$$

where ϕ_i and α_i are the parameters of the model.

Experimental results and conclusion

Data. We use data Brexit subReddit, as part of the Debagreement dataset [4] as shown in Table 1. Interactions (Edges) between users (nodes) form a multi-edge, temporal graph. For simplicity, we consider one-hot encoded user identity as node features. At time t, we construct the graph G_t using data from months $t-2$ to t, and G_{t+1} by shifting the time window ahead by one month. We use a random subset of edges in G_t for training the GAE, and the remaining in G_t as well as the full edge set in G_{t+1} for testing. To predict the future PS, we set $p=3$ and use data 2018.05-2020.08 (first 70%) for training, and 2020.08-2021.03 (the last 30%) for testing.

Results. We first explore the relationship between d_{t+1} and ps_{t+2}, as shown in Fig.3. We can see that d_{t+1} appears to co-move with ps_{t+2}. Next, we look at the effectiveness of d_{t+1} in predicting ps_{t+2}. We evaluate the model with mean squared error (MSE). Figure 4 shows that adding d_{t+1} in the AR model significantly improves the prediction performance.

Conclusion. This paper proposed a novel framework to predict the future polarisation level of a social network by taking into account both historical polarity scores and the heterogeneity in dynamic social graphs. Unlike previous literature, such heterogeneity is measured by taking into account both the graph structure and the node features. Our results support the hypothesis that the heterogeneity between dynamic social graphs could contain useful information for predicting future polarity score.

References

[1] Nir Ailon et al. Aggregating inconsistent information: ranking and clustering. *JACM*, 2008.
[2] Francesco Bonchi et al. Discovering polarized communities in signed networks. In *CIKM*, 2019.
[3] Thomas N Kipf et al. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.
[4] John Pougué-Biyong et al. Debagreement: A comment-reply dataset for (dis) agreement detection in online debates. In *NeurIPS*, 2021.
[5] Michael Schlichtkrull et al. Modeling relational data with graph convolutional networks. In *ESWC 2018*.

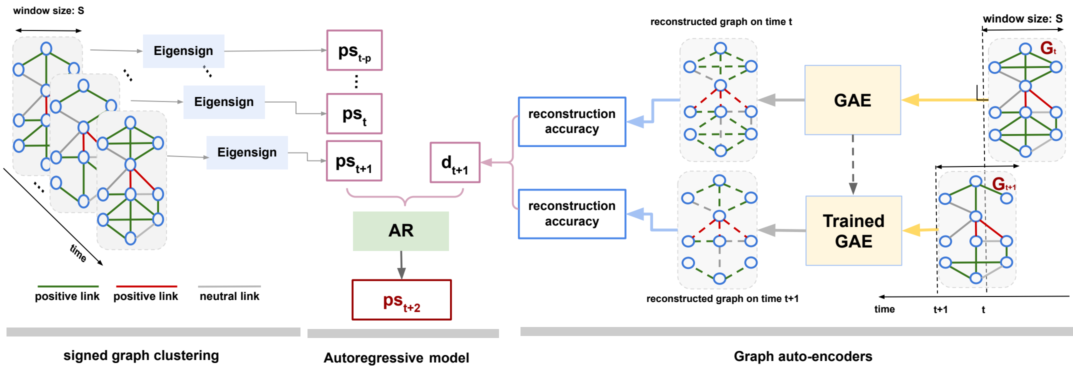


Figure 1: Our proposed framework contains three key steps. Signed graph clustering module aims to calculate polarity scores (ps_{t-p}, ..., ps_t, ps_{t+1}) of signed graphs using the Eigensign method. Graph auto-encoder module trains the GAE model on a subset of edges in G_t and tests the trained GAE model on the remaining edges in G_t and unobserved graph G_{t+1} to get the difference d_{t+1} between the reconstruction accuracies. The Autoregressive model utilises historical polarity scores (ps_{t-p}, ..., ps_t, ps_{t+1}) and the difference in reconstruction accuracy (d_{t+1}) to predict future polarity score ps_{t+2} of the social network.

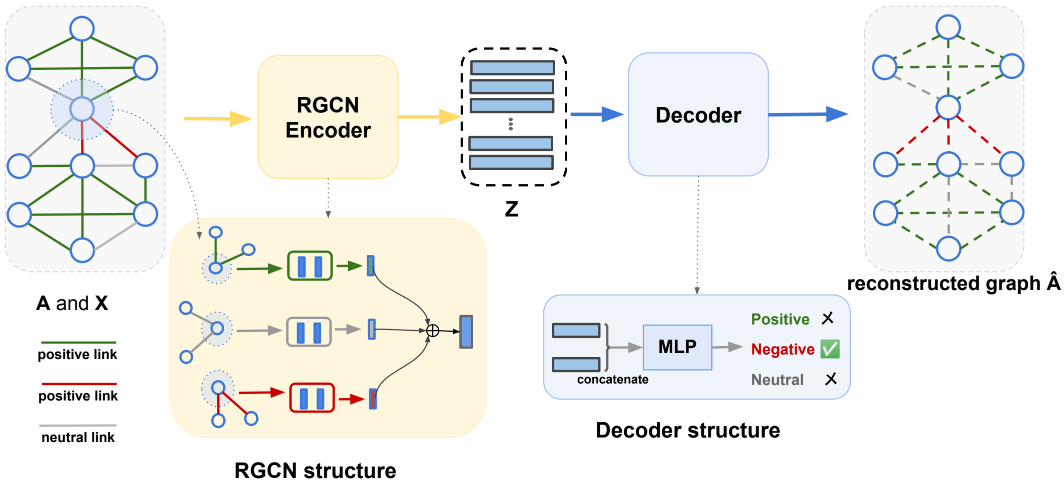
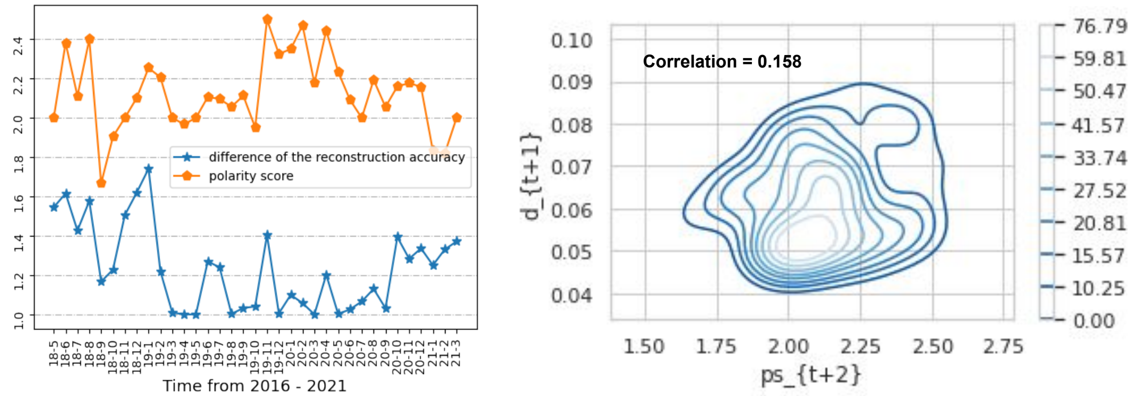


Figure 2: Graph auto-encoders. The encoder maps the input graph structure **A** and node features **X** to the latent embedding **Z** by relational graph convolutional network (RGCN) that applies different transformation mapping matrices depending on the type of edges (positive, negative or neutral). The decoder is a 3-class edge classifier.

Table 1: Overview of the Brexit subReddit dataset.

Start date	End date	# Nodes	# Edges	Positive edge	Neutral edge	Negative edge
June 2016	May 2021	722	15,745	29%	29%	42%



(a) Time series of ps_{t+2} and d_{t+1}.

(b) Correlation between ps_{t+2} and d_{t+1} is 0.158.

Figure 3: Correlation analysis of the polarity score ps_{t+2} and the difference between reconstruction accuracies d_{t+1}.

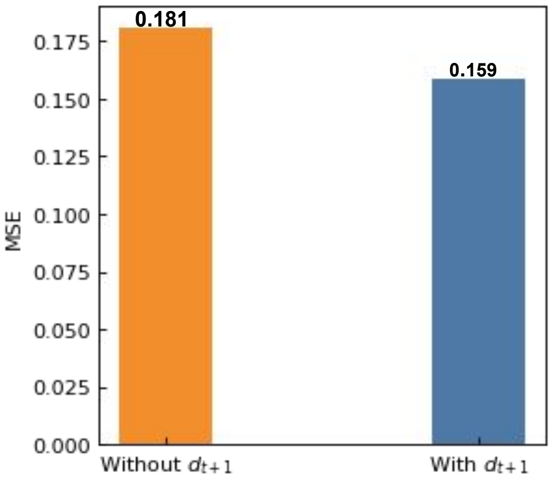


Figure 4: MSE in predicting polarity scores (the lower the better). Without d_{t+1}: the AR model only uses historical polarity scores for prediction; With d_{t+1}: the AR model uses both historical polarity scores and the difference in reconstruction accuracy (d_{t+1}) for prediction.