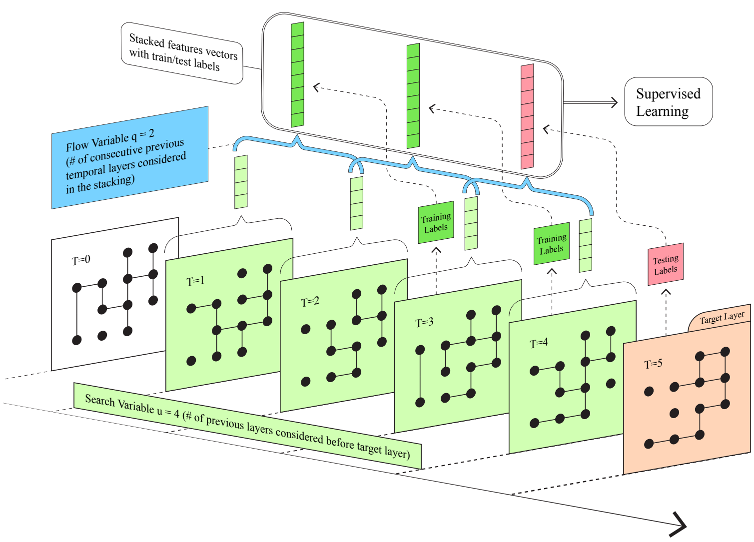
A Sequential Stacking Link Prediction for Temporal Networks

Xie He[*], Amir Ghasemian[‡], Eun Lee[♠], Aaron Clauset[†], and Peter J. Mucha[*]

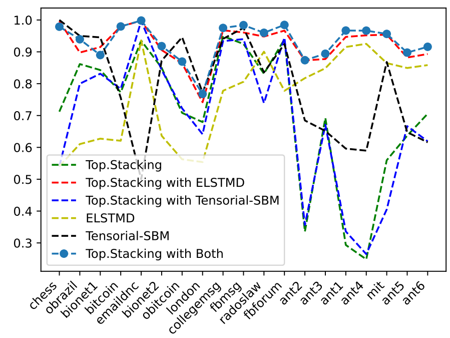
[*]Dartmouth College; [‡]Yale University;
[♠]Pukyong National University; [†]University of Colorado, Boulder

Link prediction has become an essential tool for speeding up the collection of network data and filling in incomplete network data. However, links change over time. With increased availability of temporal network data and improved methods for analyzing these data, there is a crucial need to predict missing links in temporal networks. We find that various temporal extensions of topological features, in addition to having high computational cost, fail to improve link prediction accuracy compared to using sequentially stacked static features. We construct a sequential stacking link prediction method using 41 static features to avoid verbose feature engineering, demonstrating our approach works well under a variety of settings.

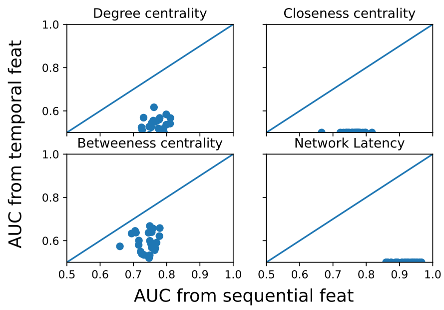
A diagrammatic explanation of the sequential stacking approach for link prediction in temporal networks is shown in Fig. 1a. We use q consecutive temporal layers in a stacked feature vector to predict the target layer: features are generated for sampled dyads in each layer and stacked across q consecutive layers. We use supervised learning to generate predicted links in the target layer, training the prediction using u layers before the target ($u > q$) with edge presence/absence labels in the layer following the sequentially stacked features (green for training and red for testing on the target). The sequential stacking method is demonstrably applicable to a variety of situations, obtaining near-oracle-level performance on two different temporal stochastic block models (not shown here) and yielding high accuracy compared to other methods on real-world temporal networks (see area under the ROC curve [AUC] scores in Fig. 1b). Indeed, sequential stacking provides much higher accuracy than direct use of temporally-extended network features (Fig. 1c) while being much less costly to compute (not shown here). Importantly, the method is immediately applicable whether the target layer is partially observed or completely missing (wheras some methods only work for partially-observed target layers). The high accuracy and computational efficiency of sequential stacking provides a highly functional method for a wide range of applications in different temporal link prediction tasks.



(a) Model



(b) Prediction AUC for Different Methods



(c) Temporal Features Comparison

Figure 1: Models and Results