

gl2vec: Learning Feature Representation Using Graphlets for Directed Networks



Kun Tu, Jian Li, Don Towsley (University of Massachusetts Amherst), Dave Braines (IBM UK), Liam Turner (Cardiff University).

gl2vec

We propose a novel, flexible and scalable network embedding methodology, gl2vec, for network classification in both static and temporal directed network:

- gl2vec constructs vectors for feature representation using a static or temporal graphlet distribution and a null model for comparison against random graphs
- Numerically characterize the impact of different null models on the performance of network classification in static directed networks
- gl2vec outperforms state-of-the-art methods in network type classification and subgraph identification in several real-world static and temporal datasets
- When gl2vec is concatenated with state-of-the-art methods, it provides significant improvements for these methods on classification accuracy in real-world application from several domains. This indicates that gl2vec provides more important and new features that are not captured in state-of-the-art methods
- The network profiling nature of gl2vec could be useful towards determining if a given terrorist network is similar to a typical terrorist network

SRPs and Null Model

We consider different null models for static networks: (i) random graphs with the same number of nodes and edges; (ii) random graphs with the same numbers of mutual, asymmetric and null edges; and (iii) random graphs with the same bi degree sequence.

We consider ensembles of randomized time-shuffled data as a temporal null model.

Subgraph ratio profile (SRP) for a graphlet i is defined as

$$SRP_i = \frac{\Delta_i}{\sqrt{\sum \Delta_i^2}},$$

where Δ_i is a normalized term that measures the difference between the count of graphlet i in an empirical network (denoted as $N_{observed_i}$) and the average count in random networks in a null model (denoted as $\langle N_{random_i} \rangle$):

$$\Delta_i = \frac{N_{observed_i} - \langle N_{random_i} \rangle}{N_{observed_i} + \langle N_{random_i} \rangle + \epsilon},$$

Experiments

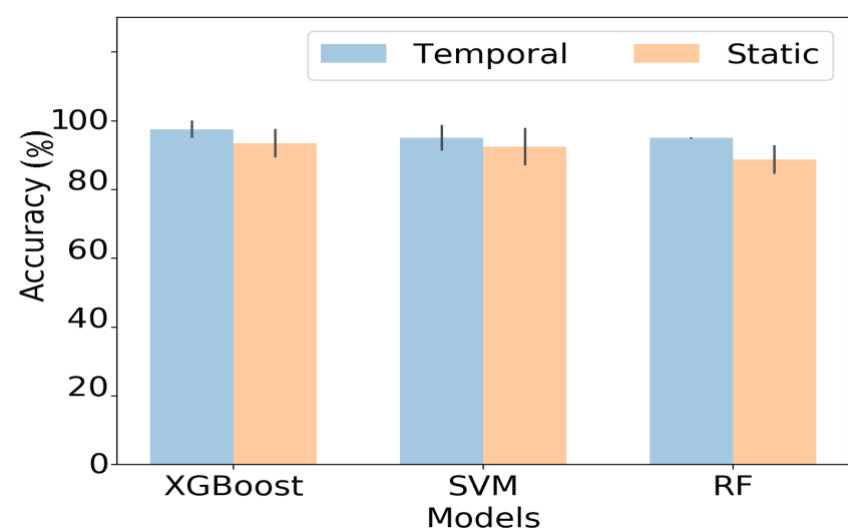


Figure 1: Classifying email and SwitchApp temporal networks.

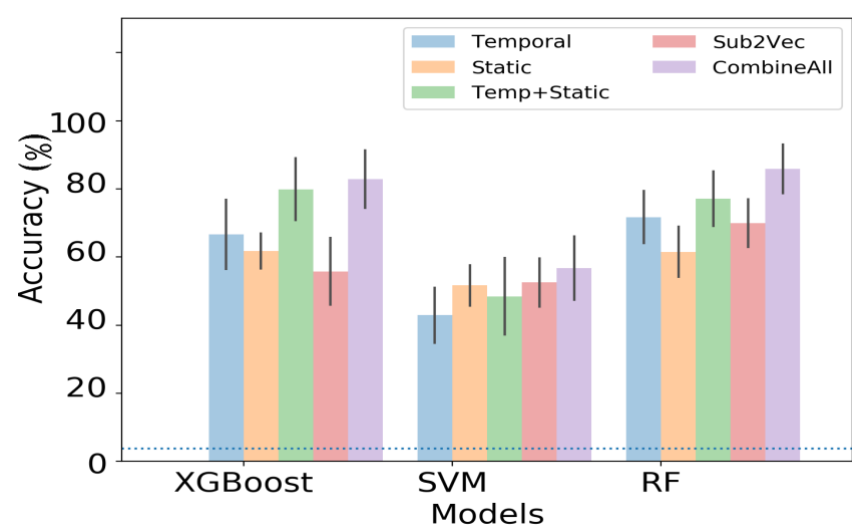


Figure 2: Department Identification in EmailEU dataset.