gl2vec: Learning Feature Representation Using Graphlets for Directed Networks

Kun Tu (Umass), **Jian Li** (Umass), **Don Towsley** (Umass),
Dave Braines (IBM UK), **Liam Turner** (Cardiff).

Objectives

- We propose gl2vec, a new network embedding methodology for network classification in both static and temporal directed networks.
- gl2vec constructs vectors for feature representation using a static or temporal graphlet distribution (Figure 1a and Figure 1b) and a null model for comparison against random graphs.

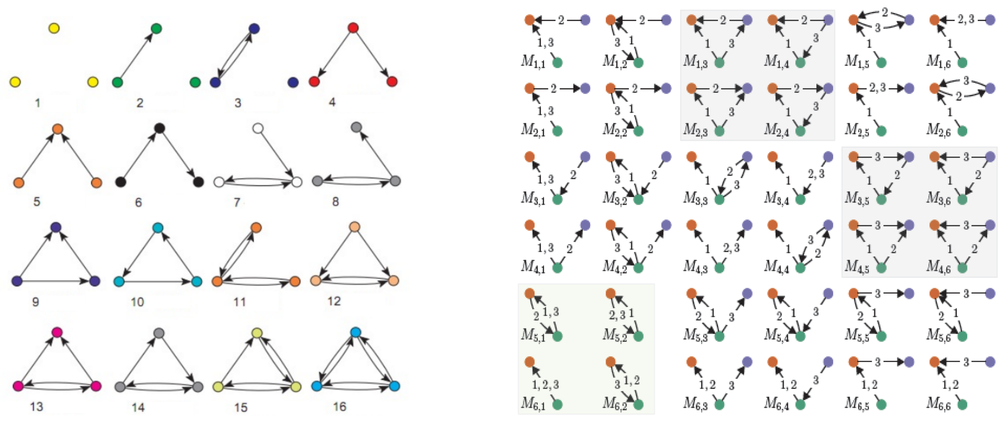

Figure 1: Types of static triadic graphlets (left) and 3-node temporal graphlets (right)

Approaches

- We use subgraph ratio profiles (SRP) as the basis for the vector representation. Where for each graphlet, i, SRP_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 $< N_{random_i} >$):

$$\Delta_i = \frac{N_{observed_i} - < N_{random_i} >}{N_{observed_i} + < N_{random_i} > + \epsilon},$$

- 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. For temporal networks, we consider ensembles of randomized time-shuffled data as the null model.

Military & Coalition Relevance

We present a general-purpose method for embedding networks into small vector representations.

Results

- We apply gl2vec to two network classification tasks, i) Network type/genre identification and ii) Specific network identification, and compare the results against state-of-the-art baselines.

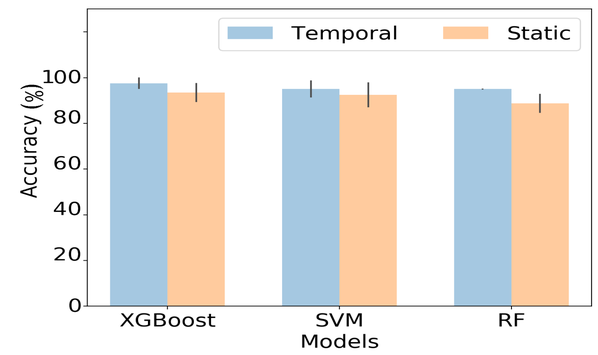

Figure 2: Classifying email networks and smartphone app switching networks

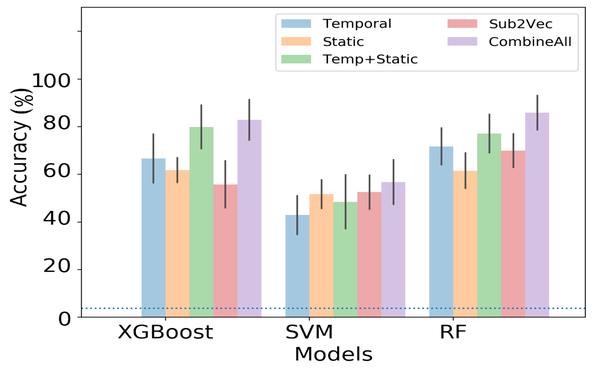

Figure 3: Department Identification in (temporal) EmailEU dataset. Dash line represents the accuracy of a random selection model

Summary & Future Work

- gl2vec comparable to state-of-the-art methods. When combined with state-of the art methods, gl2vec exhibits significant improvement.
- Results suggest local subgraph structure captures information not captured by other features; will explore this with other ML techniques.
- Study benefit of hierarchical decomposition graph coupled with gl2vec to classification tasks.

Publication(s) & Impact

- K. Tu, J. Li, D. Towsley, D. Brains, and L. Turner, "**gl2vec: Learning Feature Representation Using Graphlet for Directed Networks**", IEEE/ACM ASONAM 2019.
- K. Tu, J. Li, D. Towsley, D. Brains, and L. Turner, "**Network Classification in Temporal Networks**", ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, 2018.