Towards Learning Structural Node Embeddings using Personalized PageRank

Felix Borutta, Julian Busch, Evgeniy Faerman, Matthias Schubert

Ludwig-Maximilians-Universität München

In this abstract we present our work in progress on embeddings for nodes in graphs representing their structural similarities. Intuitively, the more similar the structural representations of the respective neighbors of two nodes, the more structurally similar they are. An exemplary application is given by the task of identifying the role of each node in the graph, which might for instance correspond to the function of a protein within a protein-protein interaction network. Modeling and developing formal definitions for concepts such as structural similarity, structural identity and roles remain challenging problems. Existing works either learn embeddings which are not able to appropriately model structural similarity or identify roles based on hand-crafted features rather than learning important features directly from the input graph. Recent work[1] proposes a framework for learning a structural embedding vector for each node based on the degree sequences of it's neighbors within a k-hop distance. Two major drawbacks of this approach are the cubic time complexity and the fact that different k-hop neighborhoods are considered equally important. Note that in most cases the closer neighbors are considered more important. In contrast, we compute Personalized PageRank (PPR) for each node in the graph and use the resulting vectors as node representations. PPRs are very fast to compute and effectively capture the probability distribution over the corresponding neighborhoods. The preliminary results in Table 1 prove our intuition: we simply sorted the raw PPR vector of each node and used them as representations. However, the experiments used in[1] are of small practical relevance, as they use strongly artificial datasets. Therefore, we also work on an appropriate evaluation framework.

Network	complete graph				removed 30% of edges			
	struc2vec		ppr		struc2vec		ppr	
	corresp.	all	corresp.	all	corresp.	all	corresp.	all
Karate	$.012 \pm .007$	$.517 \pm .275$	$.0 \pm .0$	$.082 \pm .048$	$.381 \pm .208$	$.532 \pm .245$	$.102 \pm .098$	$.209 \pm .149$
Barbell	$.008 \pm .005$	$.171 \pm .178$	$.0 \pm .0$	$.066 \pm .069$	$.012 \pm .008$	$.166 \pm .212$	$.041 \pm .055$	$.159 \pm .136$
PPI	–	–	$.0 \pm .0$	$.402 \pm .225$	–	–	$.081 \pm .098$	$.405 \pm .229$

Table 1. Experimental results on three mirrored networks with each having one bridging edge; the graph setting and setup for struc2vec is the same as in [1]. 'corresp.' shows the mean cosine distances $\pm$ the standard deviation between the node embeddings of the corresponding nodes, 'all' shows the corresponding values when the distance is measured pairwise between all embeddings. Note that struc2vec did not terminate after 3 weeks for the mirrored PPI network (7780 nodes, 77479 edges).

[1] struc2vec: Learning Node Representations from Structural Identity. Figueiredo, Daniel R and Ribeiro, Leonardo FR and Saverese, Pedro HP. accepted for KDD'17

In: M. Leyer (Ed.): Proceedings of the LWDA 2017 Workshops: KDML, FGWM, IR, and FGDB.
Rostock, Germany, 11.-13. September 2017, published at http://ceur-ws.org