Conditional Network Embeddings

Bo Kang, Jefrey Lijffijt, and Tijl De Bie

Ghent University, Belgium
firstname.lastname@ugent.be

Introduction. Network Embeddings (NEs) map the nodes of a network into a d-dimensional vector space $\mathbb{R}^d$. Ideally, 'similar' nodes are mapped to nearby points, such that an NE can be used for purposes like link prediction and classification. This paper focuses on a difficulty faced by all existing methods: many networks are fundamentally hard to embed due to their structural properties, e.g., non-uniform degree distribution, assortativity, or (approximate) multipartiteness—the latter holds, e.g., for heterogeneous information networks.

To overcome this, we propose to *condition embeddings on an informative prior* that carries information about the network that we may want to filter out, e.g., because it is hard to embed. We achieve this by bestowing a probabilistic interpretation on the embedding of the network, such that *the embedding together with the prior yields a posterior distribution over all possible networks*.

Taking this view, we try to find an embedding that maximizes the probability of the network given the embedding. We introduce a conditional probability density function over the distance between all pairs of nodes, based on a half-Normal distribution of which the variance parameter is dependent on their connectivity. Using Bayes Rule and the prior over the network, we obtain the likelihood function of the embedding. We introduce a scalable algorithm to implement the maximization, based on block stochastic gradient descent. We name the overall approach *Conditional Network Embedding* (*CNE*).

In summary, we make the following contributions:

– Conceptual innovation of *conditioning* NEs, to factor out information that is known a priori or estimated from the given network.
– A novel method called *Conditional Network Embedding* that implements this concept with a specific density function and optimization procedure.
– Extensive experiments, comparing CNE with state-of-the-art methods on link prediction and multi-label node classification. We find that CNE has consistently superior performance in link prediction and performs on par or outperforms existing methods in node classification, notably while using substantially smaller dimensionalities. We illustrate also how CNE may be used for network visualization.
– Code and data is available: https://bitbucket.org/ghentdatascience/cne.

This paper is an abstract of [1].

Summary of Method. CNE aims to find an embedding $\mathbf{X} \in \mathbb{R}^{n \times d}$ that is maximally informative about the given network $G = (V, E)$, formalized as a Maximum Likelihood Estimation (MLE) problem: $\text{argmax}_{\mathbf{X}}\ P(G|\mathbf{X})$.

Table 1. AUC scores for link prediction [more results in full paper].

Algorithm	Facebook	PPI	arXiv	BlogCat.	Wikiped.	StudentDB	Gowalla
Adamic Adar	0.9751	0.7719	0.9427	0.9268	0.8634	0.4160	0.7719
Prefer. Attach.	0.8295	0.8892	0.8640	0.9519	0.9130	0.9106	0.5626
LINE	0.9525	0.7462	0.9771	0.7563	0.7077	0.8562	0.8173
node2vec	0.9881	0.6802	0.9721	0.7332	0.6720	0.8261	0.7984
metapath2vec++	0.7408	0.8516	0.8258	0.9125	0.8334	0.9244	0.7769
CNE (uniform)	0.9905	0.8908	0.9865	0.9190	0.8417	0.9300	0.9738
CNE (degree)	**0.9909**	**0.9115**	**0.9882**	**0.9636**	**0.9158**	**0.9439**	**0.9818**
CNE (block)	NA	NA	NA	NA	NA	**0.9830**	NA

Innovative about CNE is that we do not postulate the likelihood function $P(G|\mathbf{X})$ directly, as is common in MLE. Instead, we use a generic approach to derive prior distributions for the network $P(G)$, and we postulate the density function of the embedding (conditional on the network) as the product of two $\mu = 0$ half-normal distributions for all pairwise distances:

$$p(\mathbf{X}|G) = \prod_{\{i,j\}\in E} \mathcal{N}_+\left(d_{ij}|\sigma_1^2\right) \cdot \prod_{\{k,l\}\notin E} \mathcal{N}_+\left(d_{kl}|\sigma_2^2\right).$$

with pairwise distances $d_{ij} \triangleq \|\mathbf{x}_i - \mathbf{x}_j\|_2$ between points $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d$, standard deviation $\sigma_2 > \sigma_1 > 0$. This allows one to introduce any prior knowledge about the network into the formulation, through a simple application of Bayes rule: $P(G|\mathbf{X}) = \frac{p(\mathbf{X}|G)P(G)}{p(\mathbf{X})}$. The consequence is that the embedding will not need to represent any information that is already represented by the prior $P(G)$.

Summary of Experiments. Table 1 lists AUC scores on test data for CNE and baseline methods on seven datasets from several domains. More baselines are included in the full paper, but all best performing baselines are included in Table 1. Note the two top-most methods involve no learning. We observe CNE consistently outperforms the competition and including degree information of each node in the prior improves performance. Notably, encoding the relational DB schema of *StudentDB* into the prior provides a significant performance boost.

Experiments presented in the paper on multi-class classification on the BlogCatalog, PPI, and Wikipedia networks show conditioning can also improve classification performance and CNE performs on-par or better than existing methods.

Acknowledgements. This research has received funding from the ERC under the EU's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement no. 615517, from the FWO (project no. G091017N, G0F9816N), and from the EU's Horizon 2020 research and innovation programme and the FWO under the Marie Sklodowska-Curie Grant Agreement no. 665501.

References

1. Kang, B., Lijffijt, J., De Bie, T.: Conditional network embeddings. In: Proceedings of the International Conference on Learning Representations (2019)