Learning a Degree-Augmented Distance Metric From a Network

Bert Huang*
Computer Science Dept.
University of Maryland
College Park, MD 20742
bert@cs.umd.edu

Blake Shaw
Foursquare
36 Cooper Sq., Fl. 6
New York, NY 10003
blake@foursquare.com

Tony Jebara
Computer Science Dept.
Columbia University
New York, NY 10027
jebara@cs.columbia.edu

In many naturally occurring networks, connected nodes tend to have empirical similarities [1], which is a phenomenon commonly referred to as *homophily*. It is useful to learn how to relate the network homophily to the measurable features from data. However, due to the inherent structural nature of networks, we should not expect the similarity between connected nodes to behave in a purely pairwise independent manner. In an attempt to address the structural nature of networks, we model similarity between nodes with an added structural component: the structural measure *node degree*. We present a new algorithm that learns a similarity metric and a set of degree-based score functions that together provide a structure-aware, distance-based method for link prediction.

This method, *degree distributional metric learning* (DDML) is an extension of *structure preserving metric learning* (SPML) [4], both of which, given a set of points in feature space and a graph connecting these points, learn a linear feature transform matrix (analogous to a Mahalanobis matrix) such that a given connectivity algorithm, such as k-nearest neighbors, outputs the original graph. Such a metric is considered *structure preserving*, borrowing the idea from *structure preserving embedding* [5], which embeds nodes without observable feature information, and applying it to the setting where nodes already have associated features. Many variants of structure preserving embedding are defined by using different connectivity algorithms.

Degree distributional metric learning can be viewed as one such variant, which uses a connectivity algorithm that maximizes total metric similarity between connected points plus additional feature-dependent *degree preference functions* for each node. Maximizing with degree preference functions generalizes fixed-degree connectivity algorithms such as k-nearest neighbors or b-matching. Following the intuition that the popularity of a node is either reflected or caused by its features, we allow the features of the node to determine the number of neighbors it prefers to have. The added complexity of the degree preference functions allows further separation of structural behavior from pairwise distance-based behavior, and should thus lead to better learned metrics when degree preference behavior is accurately learned.

Given as input an adjacency matrix $\mathbf{A} \in \mathbb{B}^{n \times n}$, and node features $\mathbf{X} \in \mathbb{R}^{d \times n}$, DDML learns a similarity metric, and a degree preference function $g : \{\mathbb{R}^D, \mathbb{N}\} \mapsto \mathbb{R}$, which takes a node descriptor and a candidate degree d and outputs a real valued preference score for that node having degree d. The degree distributional metric is parameterized by matrices $\mathbf{M} \in \mathbb{R}^{D \times D}$ and $\mathbf{S} \in \mathbb{R}^{n \times D}$. The distance between two points under the metric is defined as $D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)$. Using the notation that $\mathbf{s}_c$ is the $1 \times D$ dimensional c'th row of $\mathbf{S}$, the degree preference function is $g(\mathbf{x}_i, d; \mathbf{S}) = \sum_{c=1}^{d} \mathbf{s}_c \mathbf{x}_i$. The linearity of this parameterization for the degree preference function is useful for efficient learning, as discussed below. A graph is predicted by finding a connectivity that maximizes the sum of the similarity between neighbors and the degree preference functions for each node. The prediction operation computes

$$\underset{\mathbf{A}}{\operatorname{argmax}} \quad - \sum_{ij|A_{ij}=1} D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j; \mathbf{M}) + \sum_i g\left(\mathbf{x}_i, \sum_j A_{ij}; \mathbf{S}\right). \tag{1}$$

*Most of the work described in this abstract was done while all three authors were at Columbia University.

Table 1: Wikipedia (top), Facebook (bottom) dataset and experiment information. Shown below: number of nodes n, number of edges $|E|$, dimensionality d, and AUC performance. The small Wikipedia categories (*) are ranked by DDML using degree information from the training graph, and the other data sets are ranked using only the learned distance.

	n	$\|E\|$	d	Euclid.	RTM	SVM	SPML	DDML
Graph Theory	223	917	6695	0.624	0.591	0.610	**0.722**	0.691*
Philosophy Concepts	303	921	6695	0.705	0.571	0.708	0.707	**0.746***
Search Engines	269	332	6695	0.662	0.487	0.611	**0.742**	0.725*
Philosophy Crawl	100k	4m	7702	0.547	–	–	**0.601**	0.562
Harvard	1937	48k	193	0.764	0.562	0.839	**0.854**	0.848
MIT	2128	95k	173	0.702	0.494	0.784	**0.801**	0.797
Stanford	3014	147k	270	0.718	0.532	0.784	0.808	**0.810**
Columbia	3050	118k	251	0.717	0.519	0.796	0.818	**0.821**

Prediction function (1) can be efficiently optimized by an augmented maximum weight b-matching [2]. This, along with the linearity of degree preference function g, yields a cutting-plane algorithm for learning the parameters based on *structured support vector machines* (SVMs), but such an algorithm requires iteratively solving a maximum weight b-matching and a quadratic program, which becomes cumbersome with large input sizes. Instead, the learning objective can be posed as another form of SVM, which can be efficiently optimized using stochastic subgradient descent, with theoretical guarantees that convergence does not depend on the input size [3]. Additionally, limiting the maximum degree of any node to a fixed constant yields a learning algorithm whose overall running time is independent of the size of the input.

Extending the experiments done for SPML [4], we compare DDML to a variety of methods for predicting links from node features: Euclidean distances, *relational topic models* (RTM) , traditional *support vector machines* (SVM), and SPML. Table 1 summarizes the experimental results. DDML's performance is always comparable to SPML, and often better.

In summary, DDML is an extension of SPML that learns degree preference functions, which are used in addition to the learned distances to predict a graph. DDML aims to learn a richer model than SPML, yet uses a comparable learning algorithm which also can learn from large-scale input. Large-scale prediction leveraging the full power of the model remains an open problem, since the degree preference functions introduce dependencies between all nodes, requiring a combinatorial optimization to find the maximum-scoring graph. Nevertheless, using only the learned metric or a simple heuristic for prediction based on the full model yields prediction performance that is comparable to or better than SPML on some data.

References

[1] Nicholas A. Christakis and James H. Fowler. The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, 357(4):370–379, July 2007.

[2] B. Huang and T. Jebara. Exact graph structure estimation with degree priors. In M. Wani, M. Kantardzic, V. Palade, L. Kurgan, and Y. Qi, editors, *International Conference on Machine Learning and Applications*. IEEE Computer Society, 2009.

[3] S. Shalev-Shwartz, Y. Singer, and N. Srebro. Pegasos: Primal estimated sub-gradient solver for SVM. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 807–814, New York, NY, USA, 2007. ACM.

[4] B. Shaw, B. Huang, and T. Jebara. Learning a distance metric from a network. In *Advances in Neural Information Processing Systems 25*, To appear.

[5] B. Shaw and T. Jebara. Structure preserving embedding. In A. Danyluk, L. Bottou, and M. Littman, editors, *International Conference on Machine Learning*, volume 382 of *ACM International Conference Proceeding Series*. ACM, 2009.