
The Deep Latent Position Topic Model for Graph Clustering and Representation in Networks with Textual Edges

Rémi Boutin^{*1}, Charles Bouveyron , and Pierre Latouche

¹Mathématiques Appliquées Paris 5 – Institut National des Sciences Mathématiques et de leurs Interactions : UMR8145, Centre National de la Recherche Scientifique : UMR8145, Université de Paris : UMR8145 – France

Abstract

Numerical interactions between individuals often result in the creation or exchange of texts. For instance, on Twitter, any user can retweet content published from another user or directly mention another Twitter account when writing a tweet. This type of data is naturally represented by a network where the individuals are associated with the nodes and the exchanged texts with the edges. To understand those complex data structures, clustering nodes into homogeneous groups is crucial. Moreover, being able to render a comprehensible visualisation of the data and of the results is mandatory. Therefore, we propose Deep-LPTM, a model-based clustering strategy relying on a variational graph autoencoder approach as well as a probabilistic model to characterise the topics of discussion. Deep-LPTM allows to build a joint representation of the nodes and of the edges in two embeddings spaces. The parameters are inferred using a variational inference algorithm. We also introduce IC2L, a model selection criterion specifically designed to choose models with relevant clustering and visualisation properties. An extensive benchmark study on synthetic data is provided. In particular, we find that Deep-LPTM better recovers the partitions of the nodes than the state-of-the art ETSBM and STBM. Eventually, the emails of the Enron company are analysed and visualisations of the results are presented.

Keywords: graph convolutional network, embedded topic model, deep latent position model, unsupervised learning

^{*}Speaker