The Annals of Statistics
2023, Vol. 51, No. 6, 2506–2525
https://doi.org/10.1214/23-AOS2340

PROJECTIVE, SPARSE AND LEARNABLE LATENT POSITION NETWORK MODELS

BY NEIL A. SPENCER[1,a] AND COSMA ROHILLA SHALIZI[2,b]

[1]*Department of Statistics, University of Connecticut,* [a]*neil.spencer@uconn.edu*
[2]*Department of Statistics and Data Science, Carnegie Mellon University,* [b]*cshalizi@cmu.edu*

When modeling network data using a latent position model, it is typical to assume that the nodes' positions are independently and identically distributed. However, this assumption implies the average node degree grows linearly with the number of nodes, which is inappropriate when the graph is thought to be sparse. We propose an alternative assumption—that the latent positions are generated according to a Poisson point process—and show that it is compatible with various levels of sparsity. Unlike other notions of sparse latent position models in the literature, our framework also defines a projective sequence of probability models, thus ensuring consistency of statistical inference across networks of different sizes. We establish conditions for consistent estimation of the latent positions, and compare our results to existing frameworks for modeling sparse networks.

1. Introduction. Network data consist of relational information between entities, such as friendships between people or interactions between cell proteins. Often, these data take the form of binary measurements on dyads, indicating the presence or absence of a relationship between entities. Such network data can be modeled as a stochastic graph, with each individual dyad being a random edge. Stochastic graph models have been an active area of research for over 50 years across physics, sociology, mathematics, statistics, computer science and other disciplines [26].

Many leading stochastic graph models assume that the inhomogeneity in connection patterns across nodes is explained by node-level latent variables. The most tractable version of this assumption is that the dyads are conditionally independent given the latent variables. In this article, we focus on a subclass of these conditionally independent dyad models—the distance-based latent position network model (LPM) of Hoff, Raftery and Handcock [18].

In LPMs, each node is assumed to have a latent position in a continuous space. The edges follow independent Bernoulli distributions with probabilities given by a decreasing function of the distance between the nodes' latent positions. By the triangle inequality, LPMs exhibit edge transitivity; friends of friends are more likely to be friends. When the latent space is assumed to be $\mathbb{R}^2$ or $\mathbb{R}^3$, the inferred latent positions can provide an embedding with which to visualize and interpret the network.

Recently, there has been an effort to classify stochastic graph models into general unified frameworks. One notable success story has been that of the graphon for exchangeable networks [12]. The graphon characterizes all stochastic graphs invariant under isomorphism as latent variable models. LPMs can be placed within the graphon framework by assuming the latent positions are random effects drawn independently from the same (possibly unknown) probability distribution. However, graphons can be inappropriate for some modeling tasks, due to their asymptotic properties.

Received February 2020; revised September 2023.
MSC2020 subject classifications. 60G55, 62M09, 91D30.
Key words and phrases. Network models, latent position network model, latent space model, random geometric graph, projective family, network sparsity, consistent estimation.