A Generative Model for Networks Grounded in Bayesian Modeling Principles

Federica Zoe Ricci, Swarthmore College, PA

Graphs play a crucial role in many applications, ranging from drug discovery to social network modeling and astronomy. We propose a generative model for graphs that matches or exceeds the state-of-the-art, but can scale to regimes where their training fails, and is orders of magnitude faster at generation. Our approach is inspired by the Aldous-Hoover theorem, a classic representation theorem that characterizes any probability distribution of graphs that is invariant to permutations of node indices (i.e., exchangeable). This theorem represents the foundation of many classic Bayesian models for networks, and establishes that three ingredients are needed: a graph-level variable, a set of node-specific variables, and a so-called graphon function that maps graph- and node-level variables to the probability of an edge between any node pair. Given a training set of graphs, our exchangeable Variational Graph AutoEncoder (xVGAE) learns an approximation to their underlying graphons, as well as graph- and node-level latent variables. We show that our xVGAE encodes interpretable representations of graph- and node-level properties, and can (quickly, in one shot) generate new graphs that closely match key training set statistics.

Bayesian Nonparametric Dependent Edge-Exchangeable Network Models for Criminal Data

Louise Alamichel, Bocconi University

Criminal networks are complex interaction systems characterized by sparsity, hidden structures, and incomplete information, posing significant challenges for statistical modeling. In this work, we propose a Bayesian nonparametric framework for modeling interaction data using dependent edge-exchangeable network models, built on the Product Centred Dirichlet Process (PCDP). To account for community structures commonly observed in criminal organizations, we further extend the model to a nested formulation, yielding a block edge-exchangeable model in the spirit of recent developments in node-level community detection within interaction processes. The proposed framework provides a principled and flexible approach for modeling bipartite criminal networks while incorporating dependence and hierarchical block structure.

Latent Network Models to Account for Noisy, Multiply Reported Social Network Data

Martina Contisciani, Center for Critical Computational Studies, Goethe University Frankfurt

Social network data are often constructed by incorporating reports from multiple individuals. However, it is not obvious how to reconcile discordant responses from individuals. There may be particular risks with multiply reported data if people's responses reflect normative expectations' such as an expectation of balanced, reciprocal relationships. In this talk, I will present a probabilistic model that incorporates ties reported by multiple individuals to estimate the unobserved network structure. In addition to estimating a parameter for each reporter that is related to their tendency of over- or under-reporting relationships, the model explicitly incorporates a term for mutuality, the tendency to report ties in both directions involving the same alter. Our model's algorithmic implementation is based on variational inference, which makes it efficient and scalable to large systems. We apply our model to data from a Nicaraguan community collected with a roster-based design and 75 Indian villages collected with a name-generator design. We observe strong evidence of mutuality in both datasets, and find that this value varies by relationship type. Consequently, our model estimates networks with reciprocity values that are substantially different than those resulting from standard deterministic aggregation approaches, demonstrating the need to consider such issues when gathering, constructing, and analysing survey-based network data.

Bayesian Poisson-Randomized Gamma Tensor Factorization with Application to International Trade Flows

Jie Jian, University of Chicago

We study sparse semi-continuous tensor data with excess zeros, heavy right tails, and slice-specific dispersion. Such features arise naturally in monetary-valued multi-way data, such as international trade, where most exporter--importer--product--year cells are zero while positive values are continuous and highly variable. To model these data, we propose a Bayesian hierarchical tensor factorization model that places a low-rank CP structure on a latent Poisson rate tensor and couples it with a conditional Gamma model for positive outcomes, with rate parameters that can vary across slices within a mode. The model therefore separates the occurrence and magnitude of positive observations while borrowing strength across all tensor dimensions through a shared low-rank latent structure. To scale posterior inference to large arrays, we develop a hybrid variational--Monte Carlo algorithm that combines efficient coordinate ascent updates with a partially collapsed augmented-data sampler. Applied to approximately 60 million trade flows, the method surfaces multiway dependence across exporters, importers, products, and years that is difficult to recover from gravity-type or pairwise network analyses, which do not jointly model the product and temporal dimensions.