# Generative models for social network data

Kevin S. Xu and TBA

## Abstract

A social network is a structure of actors or vertices, representing people or groups of people, and ties or edges, representing relations between actors. Due in part to the ubiquity of social network data today, interest in social network analysis has spread beyond its traditional home in the social sciences to many other disciplines including physics, computer science, statistics, and engineering. A topic of significant interest in social network analysis is the creation of statistical models for social network data. Many of these models share a common underlying structure: edges are generated with probability conditional on a set of latent or hidden attributes for vertices. Differences between models generally stem from different philosophical choices about how to learn from data or different empirically-motivated goals.

In this tutorial, we cover three main classes of *generative models* for social network data, under which many of the commonly used statistical network models fall:

- *Latent space models*, which generally assume latent continuous attributes for vertices where the probability of an edge between two vertices is given by a distance function applied to the attributes of the vertices.
- *Block models*, which divide vertices into one of $k$ latent classes where the probability of an edge between two vertices depends only on the classes of the vertices.
- *Latent feature models*, which allow vertices to have arbitrarily many unique (typically binary) features, where the probability of an edge between two vertices is given by a weighted sum of the elements of their feature vectors.

We discuss some of the challenges when it comes to applying these types of generative models on social network data, including

- Interpretability of model parameters, model structure, and their relationship to social network structure.
- Optimization methods to fit these generative models, which involves estimating the latent attributes of the vertices, in an optimal or near-optimal manner.
- Model selection and model checking to validate a particular fit to a social network model.

We conclude with an overview of recent developments on generative models for social network data, including models for a *collection* of networks, which can be used to represent relations at different times (dynamic networks) or different types of relations (multi-layer networks).

## Expected audience

This tutorial should be applicable to attendees with interests in social network analysis from a statistical perspective and backgrounds in any of the topic areas covered by SBP-BRiMS, including behavioral and social sciences, public health, and computer and information sciences.

## Biosketch

Kevin S. Xu received the B.A.Sc. degree in Electrical Engineering from the University of Waterloo in 2007 and the M.S.E. and Ph.D. degrees in Electrical Engineering: Systems from the University of Michigan in 2009 and 2012, respectively. He was a recipient of the Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Master's and Doctorate Scholarships. He is currently an assistant professor in the EECS Department at the University of Toledo and has previously held industry research positions at Technicolor and 3M. His main research interests are in machine learning and statistical signal processing with applications to network science and human dynamics.

Second presenter TBA