
# Challenges in modern statistical network analysis: Data collection and covariate effect assessment

Cornelius Fritz

(Pennsylvania State University)

13.07.2023, 16.00 (c.t.)

Department of Statistics, Ludwigstr. 33, Room 144
and online via Zoom (Link)
(Meeting-ID: 913-2473-4411; Password: StatsCol22)

Statistical network analysis has recently gained considerable traction due to the insufficiencies of common regression models to cope with their inherent violation of the iid assumption and the increasing availability of network-valued data. In this talk, I will showcase how two important areas of applied statistics, namely data collection and covariate effect assessment, are affected by this recent trend and propose novel methods to address them.

While relational event data are often characterized as instantaneous events contrasting to durable ties commonly studied in classic social network analysis, nowadays, even durable ties often come with time stamps. For instance, phone or Zoom calls are an example where one can associate a duration with each observed event. The corresponding type of event data, for which we coin the term "durational event data", encodes a start and end of an interaction. To use such data, I introduce durational event models in analogy to competing risk models and Relational Event Models for instantaneous events. Note that peculiarities of the type of studied events, e.g., that each actor can only be in one simultaneous call, should be accounted for in the model. Moreover, complex interdependencies between the two event types lead to novel sufficient statistics and preprocessing techniques. Applications include data from Bluetooth proximity sensors and phone interactions.

The primary interest in analyzing network data in many cases lies in assessing the effects of exogenous covariates on edge formation rather than understanding structural aspects of the observed network pattern. Yet most models for network data focus on the latter issue and, most importantly, impose specific structural assumptions. To formulate a generalized linear model framework that, on the one hand, allows for straightforward incorporating and interpreting covariate effects and, on the other hand, takes account of the complex dependency structure without encompassing too restrictive assumptions, we introduce graphon-structured residuals. More specifically, we develop an extension to the Graphon model where one can incorporate covariate information. We achieve this by extending the linear predictor by a Graphon term and using a suitable link function. Our approach heeds recent calls that network dependence can lead to spurious associations if not accounted for and that one should test a substantive theory with models with adequate predictive power. In three application cases to binary, count-valued, and weighted networks, we showcase the need to account for dependencies and the link-prediction abilities of our model.

**Biography:**
Cornelius Fritz is a postdoctoral fellow at Penn State working with Michael Schweinberger and David Hunter on network models under local dependence. Before that, he was the interim professor of Data Science at LMU. He obtained his Ph.D. in statistics under the supervision of Göran Kauermann. His research was partially funded by the DFG project "International Arms Trade: A Network Approach" and the Munich Center for Machine Learning (MCML). Cornelius' research mainly revolves around analyzing dynamic networks to answer questions posed within substantive sciences, e.g., Political Science and Sociology, through novel data analysis techniques that combine statistical and machine learning thinking. He was also an active member of the CODAG (COVID-19 Data Analysis Group), researching the interplay of mobility patterns and COVID-19 infections as well as supporting other projects.