RESEARCH STATEMENT

Cornelius Fritz
Department of Statistics ⋄ Pennsylvania State University
Email: cfritz@psu.edu⋄ Website: corneliusfritz.com

Smart devices collecting interpersonal data surround us at every move and facilitate novel ways of measuring and understanding social behavior. The collected data provide planetary-scale views of online interpersonal relations, allowing a more nuanced look at bias in information diffusion, polarization, and echo chamber effects. In my research, I use statistics to learn from such network data to answer questions posed within the social sciences in uncertain and changing environments.

My research mainly originates from multidisciplinary collaborations with social scientists approaching me with data and questions revolving around networks. As a statistician, I operate in two worlds: the *real* world, which encompasses observed data with all its imperfections and substantive knowledge of the subject matter, and the *model* world, which is an artificial representation of the *real* world characterized by a stochastic model. I develop novel data analysis techniques by combining statistical and machine learning with substantive theory to bridge the gap between the *real* and *model* world.

PAST RESEARCH

Political Science: In collaboration with political scientists, I studied networks observed in discrete and continuous time. Building on the theorization of signed networks, I introduced the Signed Exponential Random Graph Model (SERGM) for studying dynamic signed networks observed at discrete points in time. Based on hypotheses derived from structural balance theory, we formulate interpretable signed network statistics, capturing dynamics such as "*the enemy of my enemy is my friend*". In the empirical application, we find evidence for structural balance in the network of cooperation and conflict between countries (Fritz et al., 2023b). In another paper, we studied combat events during the Syrian civil war, which are instantaneous and continuously observed interactions. Since automated or human-coded events often suffer from non-negligible false-discovery rates in event identification, I proposed the Relational Event Model for Spurious Events (REMSE) as a flexible solution to control for false-positive events (Fritz et al., 2023a) which can lead to biased estimates.

Economics: To gain insights into innovation in the workplace, I examined the co-inventorship of all electrical engineering patents over 20 years as a bipartite network of inverter-to-patent relations (Fritz et al., 2023). To account for the sheer size of the population network, I suggested a temporal decomposition of the data into multiple smaller networks accounting for the natural mortality of inventors. Our results corroborate that inventor characteristics (gender, seniority, and spatial proximity) and team formation are essential to the dynamics of invention. Moreover, in De Nicola et al. (2023), I co-authored a survey geared towards economists showcasing different ways dependence can be accounted for in the analysis of networks.

Epidemiology: Driven by the pressing issues during the COVID-19 pandemic, I was an active CODAG (COVID-19 Data Analysis Group) member at the Ludwig Maximilian University Munich, Germany. With data from Facebook's *Data for Good* program, I investigated how regional mobility and social connectivity affect COVID-19 infections in Germany (Fritz and Kauermann, 2022). I extended this work to yield better predictions by leveraging recent advances from graph neural networks (Fritz et al., 2022). Summarizing my work in the CODAG group in Fritz et al. (2022), I show how Generalized Additive Models have been successfully employed on numerous occasions to obtain data-driven insights during the COVID-19 pandemic. Our work as the CODAG group was repeatedly covered in the German national press, demonstrating how our work as scientists can impact policy decisions and the general society.

CURRENT RESEARCH

Forecasting of State-based Conflict: Relying on ideas from conflict research, I proposed an interpretable hierarchical hurdle model for predicting fatalities in Africa that separately models conflict on the national and local levels and produces realistic forecasts (Fritz et al., 2022). In collaboration with political scientists from the University of Leeds and the University of York, we supplement the point predictions of this hierarchical model with probabilistic forecasts and provide estimates for 2024. The predictions are now entire distributions to convey the uncertainty underlying conflict forecasts. Initial assessments indicate that our projections are more accurate and reasonable than many complex and uninterpretable neuronal network architectures and gradient boosting.

Learning from Discrete and Dependent Attribute and Network Data: The world of the twenty-first century is interconnected and interdependent, as demonstrated by recent events that started as local problems and turned into global crises (e.g., pandemics or political and military conflicts). Often, such events are unique and cannot be replicated. To learn from dependent events involving attributes and networks, a statistical platform is needed for studying non-causal or causal relationships of discrete attributes under network interference, with the potential to appeal to a wide range of practitioners. Therefore, I introduce a scalable statistical platform for studying non-causal or causal relationships of discrete attributes under network interference. We apply the model to political discourse among U.S. state legislators on the social media platform X.

FUTURE RESEARCH

Dynamic Learning from Sampled Attribute and Network Data: Since it can be argued that static networks are, in many cases, snapshots of dynamic networks taken at a particular time, it will be helpful to extend the aforementioned framework for studying the relationships of discrete attributes under network interference to dynamic settings. Continuous temporal information is often available when measurements are based on digital trace data. Likewise, versions of our model are possible for observations at only discrete points in time. In political science, these models will allow us to analyze how, for example, fake news spreads in social networks, while in psychology, they can shed light on the interplay between mental health and communication networks. Additionally, in many cases, only a subset of actors can be observed, and not the entire population, leading to sampled data. Therefore, additional work is needed to assess under what circumstances inferential conclusions from sampled data to the population are valid and how this affects estimation.

Local Dependence in Large Event Data: Driven by the conjecture that individuals can only maintain up to 150 relationships, known as Dunbar's number, I want to propose realistic models for large-scale event data. Given the ease at which massive data can be collected and how profoundly online behavior influences global issues, such as political polarization, there is a real need for novel scalable models and scalable methods to use large-scale relational event data. We relax the unreasonable assumption of global dependence between events made in state-of-the-art relational event models by relying on the concept of local dependence. Local dependence helps to translate the idea of Dunbar's number into a mathematical model by assuming that events only affect one another if the involved actors belong to the same neighborhood. I plan to investigate three ways to leverage different forms of local dependence for event data, where local neighborhoods are non-overlapping, overlapping, or nested. Collaborations with sociologists and political scientists are planned to incorporate domain knowledge in the definitions of these neighborhoods.

Analyzing Durational Event Data: While relational event data are often characterized as instantaneous events in contrast to durable ties studied in classic social network analysis, even durable ties often come with time stamps nowadays. For instance, phone or Zoom calls are an example where we can associate a duration with each observed event. The corresponding type of event data, for which I coined the term "durational event data," encodes the start and end of an interaction. The complexity of the observed events, such as an actor being able to participate in only one call at a time and the intricate interdependencies between the two types of events, presents significant challenges, necessitating a novel model. I have modified competing risk regression models to study durational events to overcome these challenges. Since most theories in analytical sociology on networks are based on durable ties, this class of models provides a convenient way to test theories in practice without being required to aggregate the temporal information, which is currently the standard practice. Applications include data from Bluetooth proximity sensors and phone interactions, while extensions of the model to use it for clustering actors are possible.

SELECTED PAPERS

De Nicola, G., C. Fritz, M. Mehrl, and G. Kauermann (2023). Dependence matters: Statistical models to identify the drivers of tie formation in economic networks. *Journal of Economic Behavior & Organization 215*, 351–363.

Fritz, C., G. De Nicola, S. Kevorg, D. Harhoff, and G. Kauermann (2023). Modelling the large and dynamically growing bipartite network of german patents and inventors. *Journal of the Royal Statistical Society. Series A (Statistics in Society) 186*(3), 557–576.

Fritz, C., G. De Nicola, M. Rave, M. Weigert, U. Berger, H. Küchenhoff, and G. Kauermann (2022). Statistical modelling of COVID-19 data: Putting generalised additive models to work. *Statistical Modelling (OnlineFirst)*.

Fritz, C., E. Dorigatti, and D. Rügamer (2022). Combining graph neural networks and spatio-temporal disease models to predict COVID-19 cases in Germany. *Scientific Reports 3930*(12), 1–18.

Fritz, C. and G. Kauermann (2022). On the interplay of regional mobility, social connectedness, and the spread of COVID-19 in Germany. *Journal of the Royal Statistical Society. Series A (Statistics in Society) 185*(1), 400–424.

Fritz, C., M. Mehrl, P. W. Thurner, and G. Kauermann (2022). The role of governmental weapons procurements in forecasting monthly fatalities in intrastate conflicts: A semiparametric hierarchical hurdle model. *International Interactions 48*(4), 778–799.

Fritz, C., M. Mehrl, P. W. Thurner, and G. Kauermann (2023a). All that glitters is not gold: Relational events models with spurious events. *Network Science 11*(Special Issue 2).

Fritz, C., M. Mehrl, P. W. Thurner, and G. Kauermann (2023+b). Exponential random graph models for dynamic signed networks: An application to international relations. *arXiv*. Minor Revision at Political Analysis.