

Algorithmic challenges towards temporal data in social network analysis: A case study on the assemblies of Brethren in Germany

Jens Dörpinghaus^{1,2}

¹ Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany,
Email: jens.doerpinghaus@bibb.de, <https://orcid.org/0000-0003-0245-7752>

² University of Koblenz, Germany

Background

Understanding social interactions and networks and how they influence society are important issues for historical research. Most works show that the described data and source problem are one of the greatest hurdle¹. Although some preliminary work has been carried out on the question, how missing data influences a network², there are still several open questions regarding the stability of social networks for missing and additional data. The main question is: Can we still use the same algorithms, if we know that data is incomplete? However, the need to work with temporal data makes an answer to this question even more urgent. Our first research question is how longitudinal data influences a network. In other words: If a network G comprises data for different time points t_1, \dots, t_n , can we still apply analysis methods, e.g., centrality measures or community detection, that were initially developed for one particular time point? Or do we need to re-interpret their output or adjust the algorithms? However, this question cannot be answered without discussing the data schema for temporal data. The second research question is thus dedicated to the efficient storage of temporal data in a social network. While most entities like actors and locations have a given lifetime, organizations or functions may have predecessors and successors. In other words: If an entity is detached, which relations exist, and how can we manage the lifespan of them? How can algorithms track and use this temporal data? Answering these questions is key for understanding the algorithmic challenges towards temporal data in social network analysis.

Methods and data

We will illustrate our research with a dataset on the development of the assemblies of Brethren in Germany between 1849 and today. Their development in Germany is initially connected to Julius Anton von Poseck and Carl Brockhaus. The dataset shows several challenges: First, the data is sparse, because information are no longer available or were never published. Second, the movement soon split up in several subgroups, e.g., exclusive and open Brethren, and congregations may have changed their affiliation over time, which makes the data even more complex. Third, while this network is congregation-centric it also contains data on actors, some of them travelling preachers, schools and publishers. Thus, it also includes the challenge of multi-layer networks which store multiple data types³.

With this data, we will evaluate how existing methods from computer science for versioning of knowledge graphs and ontologies can be applied to social networks⁴. We will present a generic approach inspired by *temporal RDF*⁵ which covers different modelling schemas. While interdisciplinary approaches are at hand, applications from humanities and in particular historical networks research have a different perspective on data. For example, a closed organization may still have an influence or may be referred to later. However, with this approach, we will evaluate the

1 See Leidwanger et al., “A Manifesto for the Study of Ancient Mediterranean Maritime Networks.”

2 See de Valeriola, “Can Historians Trust Centrality?”; Dörpinghaus et al., “Centrality Measures in Multi-Layer Knowledge Graphs.”

3 Cf. Bornhofen and Düring, “Exploring Dynamic Multilayer Graphs for Digital Humanities”; Kivelä et al., “Multilayer Networks.”

4 For example, VocBench, Mobi, or other algorithmic approaches, see Stellato et al., “VocBench 3”; Dörpinghaus, Weil, and Binnewitt, “Analyzing Longitudinal Data in Knowledge Graphs Utilizing Shrinking Pseudo-Triangles.”

5 Gutierrez, Hurtado, and Vaisman, “Temporal RDF”; Zhang et al., “RDF for Temporal Data Management—a Survey.”

behavior of analysis methods like centrality measures and community detection and discuss limitations and questions for further research.

Findings

Several approaches exist to manage longitudinal data in networks. All of them bias the output of algorithms analyzing the data. We present an overview on limitations, possible solutions and open questions to different data schemas for temporal data in social networks based on a generic RDF-inspired approach. We illustrate our findings with the data based on the assemblies of Brethren in Germany. While limiting algorithms to one particular time point or layer does not influence the output, applying them to a network comprising multiple time points does either need adjusted algorithms or reinterpretation.

This interdisciplinary approach has several advantages. First, it allows the integration of further data from semantic web and is compatible with widely used technologies in computer science. Thus, second, it makes results and approaches directly available for social networks. There are still several open questions in other domains, and no single solutions exists. In addition, rewriting algorithms to analyze longitudinal social networks and their re-interpretation demands discussion between different domains. Therefore, our paper is also a plea for more interdisciplinary exchange.

Bibliography

- Bornhofen, Stefan, and Marten Düring. "Exploring Dynamic Multilayer Graphs for Digital Humanities." *Applied Network Science* 5, no. 1 (2020): 1–13.
- Dörpinghaus, Jens, Vera Weil, and Johanna Binnewitt. "Analyzing Longitudinal Data in Knowledge Graphs Utilizing Shrinking Pseudo-Triangles." In *2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS)*, 323–27. IEEE, 2022.
- Dörpinghaus, Jens, Vera Weil, Carsten Düing, and Martin W. Sommer. "Centrality Measures in Multi-Layer Knowledge Graphs." *Communication Papers of the 17th Conference on Computer Science and Intelligence Systems, 2022*, 163–70.
- Gutierrez, Claudio, Carlos Hurtado, and Alejandro Vaisman. "Temporal RDF." In *The Semantic Web: Research and Applications: Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29–June 1, 2005. Proceedings 2*, 93–107. Springer, 2005.
- Kivelä, Mikko, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. "Multilayer Networks." *Journal of Complex Networks* 2, no. 3 (2014): 203–71.
- Leidwanger, Justin, Carl Knappett, Pascal Arnaud, Paul Arthur, Emma Blake, Cyprian Broodbank, Tom Brughmans, et al. "A Manifesto for the Study of Ancient Mediterranean Maritime Networks." *Antiquity* 342 (2014).
- Stellato, Armando, Manuel Fiorelli, Andrea Turbati, Tiziano Lorenzetti, Willem Van Gemert, Denis Dechandon, Christine Laaboudi-Spoiden, Anikó Gerencsér, Anne Waniart, and Eugeniu Costetchi. "VocBench 3: A Collaborative Semantic Web Editor for Ontologies, Thesauri and Lexicons." *Semantic Web* 11, no. 5 (2020): 855–81.
- Valeriola, Sébastien de. "Can Historians Trust Centrality? Historical Network Analysis and Centrality Metrics Robustness." *Journal of Historical Network Research* 6, no. 1 (2021).
- Zhang, Fu, Zhiyin Li, Dunhong Peng, and Jingwei Cheng. "RDF for Temporal Data Management—a Survey." *Earth Science Informatics* 14 (2021): 563–99.