

RESEARCH

Open Access



Towards modeling and analysis of longitudinal social networks

Jens Dörpinghaus^{1,2,5*}, Vera Weil³ and Martin W. Sommer⁴

*Correspondence:
doerpinghaus@uni-koblenz.de

¹ Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany

² University of Koblenz, Koblenz, Germany

³ Department of Mathematics and Computer Science, University of Cologne, Cologne, Germany

⁴ Argelander-Institut für Astronomie, University of Bonn, Bonn, Germany

⁵ Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden

Abstract

There are various methods for handling longitudinal data in graphs and social networks, all of which have an impact on the algorithms used in data analysis. This article provides an overview of limitations, potential solutions, and unanswered questions regarding different temporal data schemas in social networks that are comparable to existing techniques. Restricting algorithms to a specific time point or layer has no effect on the results. However, when applying these approaches to a network with multiple time points, adjusted algorithms or reinterpretation becomes necessary. Therefore, using a generic definition of temporal networks as one graph, we aim to explore how we could analyze longitudinal social networks with centrality measures. Additionally, we introduce two new measures, “importance” and “change”, to identify nodes with specific behaviors. We provide case studies featuring three different real-world networks exhibiting both limitations and benefits of the novel approach. Furthermore, we present techniques to estimate variations in importance and degree centrality over time.

Keywords: Social network analysis, Longitudinal networks, Importance, Change

Introduction

Network analysis provides a powerful framework to study interrelated data, for example social networks. However, there are several dimensions which do not naturally fit into network structures and can only hardly be represented. This includes spatial information, but also in particular temporal aspects. Thus, studying temporal or longitudinal networks lead to the development of additional visualization and analysis methods. The easiest approach is working with snapshots, but going ‘beyond the snapshot’ (Ryan and D’Angelo 2018) was early identified as research problem. Proposing a generic approach which fits into the multitude of different domains working with network approaches and the diverse set of applications is a gap in research which we will focus on.

This paper builds upon a generic framework for analyzing centrality measures in longitudinal social networks introduced in Dörpinghaus et al. (2023). The framework, including visual analysis tools, which was proposed there is the basis for our research. Here, we will extend and evaluate it in terms of its suitability for longitudinal network analysis. To this end, we add novel analysis and compare groups of actors, provide improved bounds for specific use cases of social networks, and present three complex example use cases

on real-world data. Additionally, we present experimental results on random graphs. As a further argument in favor of our approach, we present mathematical bounds for centrality measures. The most important addition to our previous paper is, however, the introduction of two additional measures: importance and change.

Background

Social network analysis (SNA) is an important part of the social sciences and has been used in both theory and practice for several decades. It is important to understand social interactions and networks and how they affect society. In the last few years, there has been a growing interest in the use of social networks in the historical sciences. In religious studies, especially narrative studies and theology, social networks have recently received considerable attention.

Scholars have always seen SNA as part of the humanities, and in recent years there has been a rapid increase in the use of methods from the digital humanities, which includes the humanities and computer science. Most works indicate that the data and source problems are one of the greatest hurdles to establish a network (Leidwanger et al. 2014). Although some preliminary work on how missing data influences a network has been carried out (Valeriola 2021), there are still several open questions regarding the stability of social networks with respect to missing and additional data. The main question is: Can we still use the same algorithms, if we know that the data are incomplete? The need to work with temporal data makes an answer to this question even more urgent.

Research questions

The primary objective of this paper is to investigate the potential for extending the methods introduced in Dörpinghaus et al. (2023) to model longitudinal social networks. To this end, we will address the following sub-research questions: (a) How can we analyze and compare groups of actors? (b) How can we approximate the change of centrality measures over time using more specific properties of social networks? (c) How can we apply these methods to real-world data including visual analysis?

Outline

This paper is divided into five sections. After this introduction, we give an overview of related work and the background of this research. Here, one focus is on historical network analysis (HNA), because it helps to highlight the challenges and is the natural habitat for longitudinal networks.

Our methodological approach is described in the “[Method](#)” section, where we discuss the modeling of longitudinal social networks, and their analysis. In the “[Experimental results](#)” section we describe the data to carry out our experiments. The “[Discussion and outlook](#)” section is dedicated to the experimental results. Our conclusions are presented in the final section.

Related work

Modeling temporal or longitudinal data in SNA is a well-known problem (Holme and Saramäki 2019). Temporal data lead to complex network structures and Lemerrier stated in 2015: “There is no one best way for the analysis or even description of such

multidimensional data” (Lemerrier 2015). There are several modeling challenges, for example with synchronous and asynchronous events or relations, see (Lehmann 2019).

Several methods, usually introducing a novel graph structure, have been proposed, for example, modeling with

- stream graphs (Latapy et al. 2018, 2019),
- Markov chains (Peixoto and Rosvall 2019; Scholtes et al. 2014),
- with network snapshots (Xu et al. 2013), or with a discrete set of time points that may contain snapshots.

Most of these approaches are equivalent (Holme and Saramäki 2012; Dörpinghaus et al. 2023). However, a single graph-theoretic definition covering all these approaches was only introduced in Dörpinghaus et al. (2023), although only tested on random graphs. Consequently, whereas other methodologies necessitate the introduction of novel graph structures or even the conversion of a ‘traditional’ social network or knowledge graph into an alternative graph, a generic solution merely incorporates the requisite data into the network. This represents a gap in the existing research that requires further investigation.

Longitudinal studies represent a central objective in both network science and sociology. In sociology, authors have developed mixed methods approaches to account for the integration of thick qualitative insights with quantitative network analysis. These approaches have been applied by Lazega and Snijders (2015), Lazega (2017), Lazega (2016) and Bellotti (2014). However, these approaches lack generality. For instance, in Espinosa-Rada et al. (2024), the authors defined micro–macro linkage as another method. Consequently, these approaches do not contribute to a generic approach; rather, they describe specific additional methods that must be integrated into a generic approach.

Scientists are not only careful about how to model temporal networks, but also how to analyze them: “Traditional analyses of temporal networks have addressed mostly pairwise interactions, where links describe dyadic connections among individuals” Cencetti et al. (2021). Concetti et al. thus introduced “temporal hypergraphs” to address this challenge. Other researchers proposed visual analysis (Yi et al. 2010), pattern search (Franzke et al. 2018), or probabilistic discrete temporal models (Hanneke et al. 2010). As previously stated, it is evident that these methodologies do not possess the capacity for generalization; rather, they are typically contingent upon the specific structural characteristics of the graphs in question.

Centrality measures, widely used in SNA, are also challenging in temporal networks. Some researchers have proposed definitions of temporal closeness, betweenness, and eigenvector centrality, see (Pan and Saramäki 2011; Taylor et al. 2017; Sizemore and Bassett 2018). However, these definitions remain limited to the underlying graph topology, e.g. Sizemore and Bassett (2018) work with a contact sequence where nodes remain static; (Naima et al. 2023) propose temporal walks. In addition, the natural extension of centrality to groups and classes (Everett and Borgatti 1999; Rasti and Vogiatzis 2022) is usually omitted. Other authors propose MLI based on network embedding and machine learning (ML) (Yu et al. 2020). In general, ML approaches are widely used in dynamic

networks, not only in temporal networks, see (Cinaglia and Cannataro 2022). However, these approaches – although providing significant insights on the networks – are usually not comparable to the results of centrality measures, which makes them difficult to reproduce. Thus, directly related to a generic definition of temporal networks is a second gap: How can algorithms track and use this temporal data, and how does this affect the analysis of networks, e.g., with centrality measures? In this paper, we will try to bridge this gap with the approaches introduced in Dörpinghaus et al. (2023) and apply them to real-world networks.

These issues may be due to the fact that several aspects of knowledge graphs and the semantic web are not widely perceived in the SNA community. They have only recently been brought together (Dörpinghaus et al. 2022). Barats et al. conclude in 2020: FAIR data, a topic directly related to knowledge graphs, “remains a theoretical discussion rather than a shared practice in the field of humanities and social sciences.” Barats et al. (2020) As has been demonstrated previously, social networks and knowledge graphs share not only a common data structure – both are graphs – but knowledge graphs can also be regarded as a generalization of social networks. Categories for nodes and edges can be modeled as (RDF) classes (Dörpinghaus et al. 2022). Consequently, our objective is to address the research questions using knowledge graphs, with the aim of establishing a connection between the two domains.

Method

We will use a definition of a knowledge graph¹ that combines the approaches of Franzke et al. (2018), Dörpinghaus et al. (2022):

Definition 1 (Temporal social network) A *Social Network* is a graph $G = (V, E, \mathcal{T})$ with vertices (nodes) $v \in V$, edges (relations) $e \in E$ and a time domain $\mathcal{T} = \{t_0, \dots, t_k\}$ where $t_i \in \mathbb{R}$ and \mathcal{T} is sorted ascending, which means $t_i < t_j \forall i < j$.

Every node and edge may exist at one or multiple intervals of time-points $[t_s, t_e] = \{x \in \mathcal{T} : t_s \leq x \leq t_e; t_s, t_e \in \mathcal{T}\}$ denoted by $t(v)$ and $t(e)$. Thus, $t : V \cup E \rightarrow I \subseteq \mathbb{R}$. We denote the graph G at time t by

$$G^t = (V^t, E^t), \text{ where} \\ V^t = \{v \in G \mid t \in t(v)\}, E^t = \{e \in E \mid t \in t(e)\}$$

so that

$$\bigcup_{t \in \mathcal{T}} G^t = G.$$

Both edges and vertices are part of previously well-defined categories C_1, \dots, C_n , which means $V \subseteq C_1 \cup C_2 \cup \dots \cup C_n$ and $E \subseteq R_1 \cup R_2 \cup \dots \cup R_m$.

It is important to notice, that – in contrast to other definitions, e.g. Santoro and Sarpe (2022) – both edges and nodes are temporal. Unless otherwise noted, we assume that G

¹ Consequently, in this paper, the terms “social networks” and “knowledge graphs” are used synonymously. In this definition, both share a common notation and a common toolbox.

is an undirected graph. We will now present examples of the notation introduced above. It should be noted that the desired properties of this network have not yet been defined. However, we will subsequently examine the impact of very unstable or link-stream style networks.

Each vertex $v \in V$ has a lifetime $t(v)$. In general, any edge connected to v may only exist for times $t \in t(v)$. But this rule is not strict. For example, we can define categories for successors T_s and predecessors T_p , so that these edges can indicate a predecessor of a certain position at any time. To illustrate, a company (denoted as a) may be acquired by another company (denoted as b), which suggests that a can be considered a predecessor of b . Similarly, a historical figure may succeed another in a distinct position. The modeling of these relationships relies on the specific context and research questions.

For these edges we set $t(e) = \emptyset$, they are ‘timeless’. The necessity of these nodes is evident in specific modeling scenarios, such as the representation of locations. Additionally, the construction of a compatible model with certain other approaches necessitates their inclusion. In addition, v can be part of several categories, e.g., it can be an actor $v \in C_a$ and a politician $v \in C_p$. Thus, our approach can combine static and temporal information.

It was shown in Dörpinghaus et al. (2023) that this definition is equivalent to stream graphs. Here, the only difficulties are those edges and vertices that are ‘timeless’. However, extending their interval to \mathcal{T} models their behaviour in the intended way. It is quite easy to see that both approaches are also equivalent to models using snapshots of time points (Yu et al. 2020). For a detailed overview we refer to Holme and Saramäki (2012).

Thus, Definition 1 is well aligned with other approaches. However, it is also compatible with semantic web approaches and makes it easier to integrate analysis approaches. We will now move on to modelling longitudinal social networks with semantic web technologies.

Modelling longitudinal social networks

The initial definition of a social network in Dörpinghaus et al. (2022) corresponds to the definition of a knowledge graph. In particular, the categories for nodes C_1, \dots, C_n and edges R_1, \dots, R_m can be modelled using RDF classes. So we need to add time intervals to nodes and edges. To do this, Hobbs and Pan introduced the time ontology, see (Hobbs and Pan 2004; Grüninger 2011). Here they use a function *duration*: Intervals \times TemporalUnits to express intervals. We can set $duration(v) = t(v)$ and $duration(e) = t(e)$ for any node $v \in V$ and edge $e \in E$.

Thus, any social network according to the knowledge graph definition in Dörpinghaus et al. (2022) can be easily transformed into a temporal social network, where time is modelled as a property of nodes and edges.

Example 1 Consider the graph $G = (V, E, \mathcal{T})$ in Fig. 1 with $V = \{v_1, v_2, v_3\}$ and $E = \{e_1, e_2\}$ and a set of time intervals $t(v_1) = [1, 6]$, $t(v_2) = [2, 4]$, $t(v_3) = [4, 6]$, $t(e_1) = [3, 4]$ and $t(e_2) = [4, 4]$. They also provide a visualisation according to Sizemore and Bassett (2018): We visualise time by plotting a sequence of edges on a time scale. However, we extend the latter approach by adding information about the lifetime of nodes. In this case, each lifetime can be mapped according to the temporal duration.

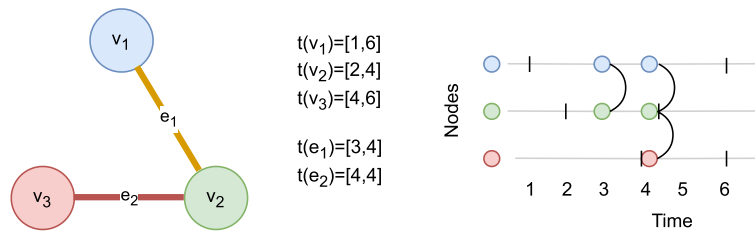


Fig. 1 Illustration of the graph in Example 1 with a definition of lifetimes in the middle and a visualisation of the lifetime of edges and the sequence of edges over time (right)

It is worth noting that the general knowledge graph definition of a social network is open to adding a variety of additional data while maintaining the general graph structure. Thus, it is useful for modelling not only temporal social networks, but also any other temporal data, e.g., disease models.

All graph structures naturally extend to temporal graphs, e.g., the degree $d(v)$ of a node v at timepoint t can be denoted by $d^t(v)$. See Appendix B for details.

Analysing networks

For a detailed overview of centrality measures, we first follow (Dörpinghaus et al. 2023). We can consider the series of a particular measure, e.g., a generic c (centrality) which could refer, for example, to closeness – cc – or betweenness centrality – bc), which is basically a vector in $\mathbb{R}^{|\mathcal{T}|}$:

$$\tilde{c}(v) = (c^{t_1}(v), \dots, c^{t_{|\mathcal{T}|}}(v)). \tag{1}$$

Recall that $\mathcal{T} = \{t_1, \dots, t_{|\mathcal{T}|}\}$ is the time domain. Note that $c^{t_i}(v) = \emptyset$ if $t_i \notin t(v)$.

Definition 2 (Lifespan) For $x \in V$ or $x \in E$ we set

$$l(x) = \sum_{i \in \{1, \dots, |\mathcal{T}|\}, c^{t_i}(x) \neq \emptyset} l(t_{i-1}, t_i),$$

where $l(t_{i-1}, t_i)$ defines the length of time elapsed between two times t_{i-1} and t_i as the lifespan of x .

However, if all times are equally distributed, this simplifies to

$$l(v) = |\mathcal{T}| - |\{x \in \tilde{c}(v) \mid x = \emptyset\}|. \tag{2}$$

This allows us to calculate the *average temporal centrality* of a node v over its lifetime which is normalized by the lifespan of a node:

Definition 3 (Average temporal centrality) The average temporal centrality of a node v is defined by

$$\bar{c}(v) = \frac{1}{l(v)} \sum_{t \in \mathcal{T}, c^t(v) \neq \emptyset} c^t(v).$$

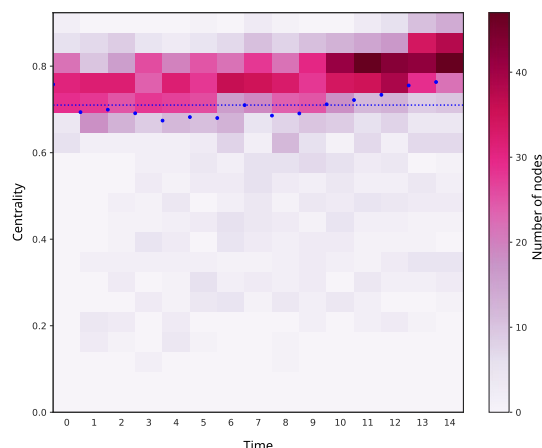


Fig. 2 Illustration of the distribution of a centrality measure over time, grouped into 20 bins between 0 and 1, as a heatmap. The blue horizontal line refers to the overall average centrality, while the blue dots refer to the average degree at a given time. This illustrates the degree centrality for $\mathcal{G}^s(100, 15, 0.1)$, see “Approximating the changes over time” section

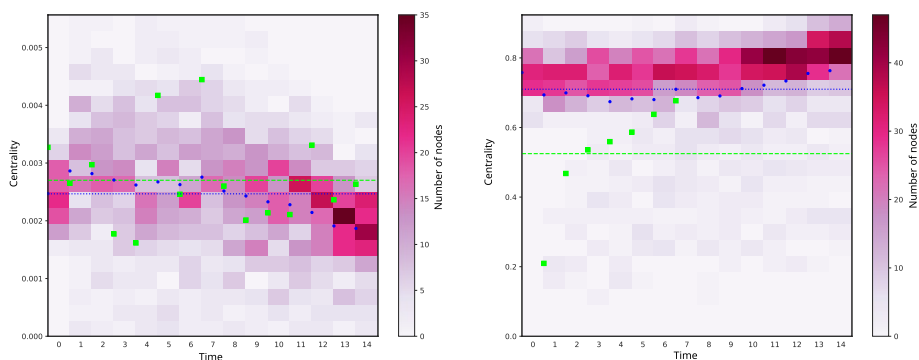


Fig. 3 Illustration of the distribution of a centrality measure over time, grouped into 20 bins between 0 and 1, as a heatmap. The blue horizontal line refers to the overall average centrality, while the blue dots refer to the average degree at one point in time. Both figures show $\tilde{c}(v)$ and $\bar{c}(v)$ (green dots and horizontal line, respectively) for two different nodes. Left: This node exists over all 15 time points and usually shows that the betweenness centrality varies a lot. Right: This node exists from time 1 to 7 and has an increasing degree centrality value. The network is based on $\mathcal{G}^s(100, 15, 0.1)$

In “[Experimental results](#)” section we will discuss several working examples and offer an interpretation of these values in light of the current state of research on degree and betweenness centrality.

First, we consider how a centrality measure evolves over time. Since we need to plot this for n nodes, we consider a heatmap visualisation that bins the number of nodes in a given interval. Next, we can plot the average centrality measure at a particular time and the average centrality over all time points, as we show in Fig. 2. This and the following two figures are illustrative and based on random graphs; for details, see “[Random graphs](#)” section.

This figure gives us a good overview of how many nodes are below and above the average centrality at a given time, and whether the network at a given time is special for the scenario. To analyse and compare a particular node with this overall picture, we can plot $\tilde{c}(v)$ and $\bar{c}(v)$, as we show in Fig. 3.

Some scholars like (Taylor et al. 2017) considered calculating and plotting $\tilde{c}(v)$, others added probabilities (Pan and Saramäki 2011). Thus, in addition to the classical approach (e.g. Sizemore and Bassett (2018)), $\tilde{c}(v)$ and $\bar{c}(v)$ allow the study of static centrality measures at a time $t \in \mathcal{T}$, comparing the individual centrality value of a particular node with the average node degree and the distribution of node degrees. In addition, by plotting the series of centrality over time, we can compare the temporal centrality measures within a given interval or across the entire timeline. While some general measures, such as average temporal centrality, have been studied previously (Holme and Saramäki 2019), their interpretation remains vague. If networks change significantly over time, this value is not comparable.

For large networks, however, we also want to measure the difference between individual actors and the entire network, e.g., which actors have a higher or lower centrality measure. For this we can compute the *importance* of a node v for a given centrality measure c at a certain time t , which is basically the distance between the centrality value of a node and the average of all nodes:

$$importance_c^t(v) = \frac{1}{l(t)} \left(c^t(v) - \frac{1}{|\{u \in V, c^t(u) \neq \emptyset\}|} \sum_{u \in V, c^t(u) \neq \emptyset} c^t(u) \right). \tag{3}$$

We can naturally extend this measure over all timepoints to

$$importance_c(v) = \sum_{t \in \mathcal{T}, c^t(v) \neq \emptyset} importance_c^t(v). \tag{4}$$

This allows us to identify those actors in a network that have significantly higher $importance(v) > 0$ or lower $importance(v) < 0$. However, this does not help to identify actors that change over time, because the importance of an actor that starts with a very low importance and gets higher importance over time may sum up to zero. Here we can compute the *change* of an actor v given a vector $\hat{\mathcal{T}} = (t \in \mathcal{T}, c^t(v) \neq \emptyset)$ with all lifetimes of v :

$$change_c(v) = \frac{1}{|\hat{\mathcal{T}}|} \sum_{i \in \{1, \dots, |\hat{\mathcal{T}}|\}} |c^{\hat{\mathcal{T}}_i}(v) - c^{\hat{\mathcal{T}}_{i+1}}(v)| \tag{5}$$

With this we can identify those actors that change over time if $change(v) > 0$. We will now continue with some methods that help to evaluate the change of these measures over time.

Approximating the changes over time

Let $\mathfrak{G}^p = \{G_1, \dots, G_t\}$ be a series of graphs and $p \in \mathbb{R}$ with $0 \leq p \leq 1$ and

$$\begin{aligned} |(V(G_i) \cup V(G_i + 1)) \setminus (V(G_i) \cap V(G_i + 1))| &\leq p|V(G_i)|, \\ |(E(G_i) \cup E(G_i + 1)) \setminus (E(G_i) \cap E(G_i + 1))| &\leq p|E(G_i)|, \end{aligned}$$

for $i \in \{1, \dots, t - 1\}$. Thus, \mathfrak{G}^p is a series of graphs with a fixed set of differences and changes from one to the other. We will use this formal framework to make some

mathematical observations about the measures introduced above, but also to evaluate networks. Here, we can either use random networks; see Appendix C for several random graph models used in this manuscript. Not only that, but we can also use real networks. In this case, the parameter p is not fixed, but changes individually for each step.

Now we can approximate the changes over time, or the error in the centrality measures that can occur due to these changes. Unless otherwise noted, we will consider $\mathfrak{G}^p = \{G_1, \dots, G_t\}$.

In Dörpinghaus et al. (2023) two bound were introduced for betweenness centrality (bc) and degree centrality (dc):

Theorem 2 *Let $i \in \{1, \dots, t - 1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then it holds that*

$$dc^{i+1}(v) \geq \frac{d^i(v) - p|V(G_i)|}{|V(G_i)| - 1 + p|V(G_i)|},$$

$$dc^{i+1}(v) \leq \frac{d^i(v) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|}.$$

For betweenness centrality, they defined

$$\sigma = |N(G_i)|p,$$

$$\epsilon = \min \left\{ D(G_i)^2, 2|V(G_i)|p \right\},$$

where $D(G)$ is the diameter of G .

Theorem 3 *Let $i \in \{1, \dots, t - 1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then,*

$$bc^i(v)\epsilon \leq bc^{i+1}(v) \leq bc^i(v) \frac{1}{\sigma}$$

holds.

However, in several social networks, nodes are usually not removed. In citation networks, for example, it is a rare case that publications are retracted. In this case, we set $V(G_i + 1) \subseteq V(G_i)$. With this, we can make bound 2 more sharp:

Theorem 4 *Let $i \in \{1, \dots, t - 1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then it holds that*

$$dc^{i+1}(v) \geq \frac{d^i(v)}{|V(G_i)| - 1 + p|V(G_i)|},$$

$$dc^{i+1}(v) \leq \frac{d^i(v) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|}.$$

Proof We know that

$$dc^i(v) = \frac{d^i(v)}{|V(G_i)| - 1}.$$

However, due to the definition of \mathfrak{G}^p , we know that at most $p|V(G_i)|$ new connections from v to other nodes can exist in G_{i+1} . In addition, $d^i(v) \leq d^{i+1}(v)$. Thus, in G_{i+1} it holds that

$$d^i(v) \leq d^{i+1}(v) \leq d^i(v) + p|V(G_i)|.$$

In addition, we know that for G_{i+1}

$$|V(G_i)| - p|V(G_i)| \leq |V(G_{i+1})| \leq |V(G_i)| + p|V(G_i)|$$

holds. Hence the claim follows. □

This allows us to set some bounds for how actors will become more important and change in the future. Here we focus on degree centrality. For the following two lemmas, p does not describe a global change ratio, but the local change ratio to the future time. For Lemma 5 we assume that $V(G_i + 1) \subseteq V(G_i)$, while this restriction is not necessary for the Lemma 6.

Lemma 5 *Let $G = (V, E, \mathcal{T})$ be a social networks with a time domain $\mathcal{T} = \{t_0, \dots, t_k\}$. For any additional timepoint t_{k+1} and a node $v \in V$ it holds that*

$$\begin{aligned} importance_{dc}^+(v) \leq & importance_{dc}(v) + \frac{1}{l(t+1)(|V(G_i)| - 1 + p|V(G_i)|)} \\ & \left(d^i(v) - \sum_{u \in V, dc^t(u) \neq \emptyset} d^i(u) + p|V(G_i)| \right), \end{aligned}$$

here $importance_{dc}(v)$ denotes the importance in \mathcal{T} and $importance_{dc}^+(v)$ for $\mathcal{T} \cup t_{k+1}$.

Proof We know by Lemma 4 that

$$\begin{aligned} \sum_{u \in V, dc^t(u) \neq \emptyset} dc^t(u) & \leq \sum_{u \in V, dc^t(u) \neq \emptyset} \frac{d^i(u) + p|V(G_i)|}{|V(G_i)| - 1 + p|V(G_i)|} \\ c^t(v) & \geq \frac{d^i(v)}{|V(G_i)| - 1 + p|V(G_i)|} \end{aligned}$$

And with this we can find an upper bound for $importance_c^{t+1}(v)$:

$$\begin{aligned} importance_{dc}^{t+1}(v) & \leq \frac{1}{l(t+1)} \left(\frac{d^i(v)}{|V(G_i)| - 1 + p|V(G_i)|} - \sum_{u \in V, dc^t(u) \neq \emptyset} \frac{d^i(u) + p|V(G_i)|}{|V(G_i)| - 1 + p|V(G_i)|} \right) \\ & = \frac{1}{l(t+1)(|V(G_i)| - 1 + p|V(G_i)|)} \left(d^i(v) - \sum_{u \in V, dc^t(u) \neq \emptyset} d^i(u) + p|V(G_i)| \right) \end{aligned}$$

Using this bound for the sum in Formula 4 shows the assumption. □

We can show a similar bound for the change of a node in the network:

Lemma 6 *Let $G = (V, E, T)$ be a social networks with a time domain $T = \{t_0, \dots, t_k\}$. For any additional timepoint t_{k+1} and a node $v \in V$ it holds that*

$$\text{change}_{dc}^+(v) \leq \text{change}_{dc}(v) + \left| \frac{dc^{|\hat{T}|}(v)(|V(G_i)| - 2 - p|V(G_i)|) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|} \right|$$

here $\text{change}_{dc}(v)$ denotes the change in T and $\text{change}_{dc}^+(v)$ for $T \cup t_{k+1}$.

Proof We know by the Formula 5 and Lemma 2 that

$$\begin{aligned} \text{change}_{dc}^+(v) &= \text{change}_{dc}(v) + \left| dc^{|\hat{T}|}(v) - dc^{|\hat{T}|+1}(v) \right| \\ &\leq \text{change}_{dc}(v) + \left| dc^{|\hat{T}|}(v) - \frac{dc^{|\hat{T}|}(v) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|} \right| \\ &= \text{change}_{dc}(v) + \left| \frac{dc^{|\hat{T}|}(v)(|V(G_i)| - 1 - p|V(G_i)|)}{|V(G_i)| - 1 - p|V(G_i)|} - \frac{dc^{|\hat{T}|}(v) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|} \right| \\ &= \text{change}_{dc}(v) + \left| \frac{dc^{|\hat{T}|}(v)(|V(G_i)| - 2 - p|V(G_i)|) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|} \right| \end{aligned}$$

This shows the assumption. □

We will now continue with experimental results on real-world networks and experiments on random graphs showing the results of these bounds.

Experimental results

Before considering random graph scenarios, we will turn to the three real-world networks introduced in “Data” section and begin with Luke’s Gospel.

Data

This section will present the results of three distinct real-world graphs. These networks were selected for their ability to exemplify typical social network use cases while representing the diverse field of applications. The selected networks are comparable to other networks that describe socio-patterns, such as human contact networks. A general overview will be provided in this section, while a more detailed analysis follows.

The first network S_1 is a network of actors in Luke-Acts (Dörpinghaus 2022; Dörpinghaus and Stenschke 2021). It thus represents both a narrative and a historical network. The exegetical approach best suited to the questions that need to be answered for SNA – and which seeks to examine the social networks represented in biblical literature – is a literary approach and narrative criticism. In this analysis, however, we will focus on the Gospel of Luke and three different time periods: Before Jesus leaves Galilee (up to Lk 9:50, 17 nodes, 37 edges), before Jesus arrives in Jerusalem (Lk 19:27, 57 nodes, 287 edges), and the entire Gospel (57 nodes, 367 edges). Thus, it will provide a narrative evaluation in a simple network at the points where the narrative changes. The density of the whole network is $d(S_1) \approx 0.23$, the diameter is $\Delta(S_1) = 4$, and the average clustering coefficient is $C(S_1) \approx 0.26$.

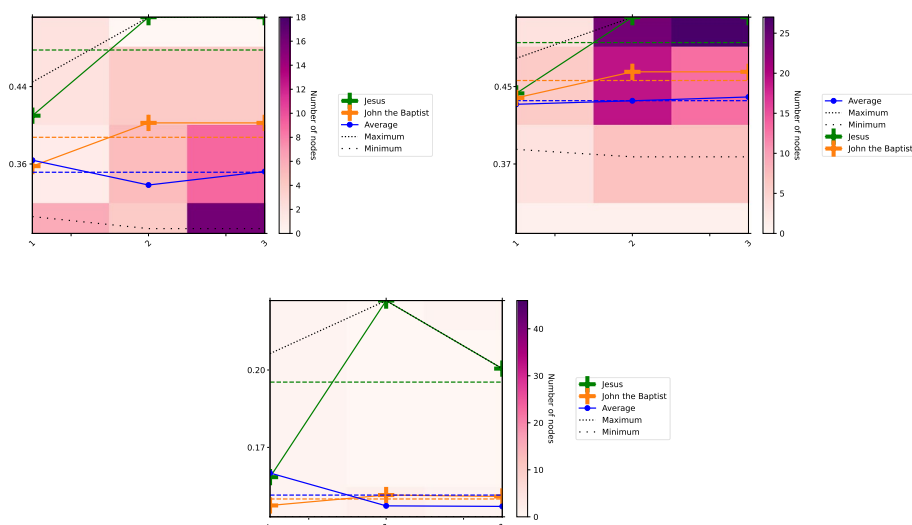


Fig. 4 Distribution of centrality measures for S_1 over time is illustrated in a heatmap. The data is grouped into 10 bins between 0 and 1. The three centrality measures, degree, closeness and Betweenness centrality are represented on the left, middle and right, respectively. We chose to show results for two actors, Jesus and John the Baptist, despite average, minimum and maximum values

The second network S_2 is a dataset on the development of the assemblies of Brethren in Germany between 2012 and 2023 (Dörpinghaus 2023). The history of the assemblies of the Brethren in Germany is constantly being researched, see for example (Riedel and Runkel 2015; Kessler 2022; Schafer 2004). This dataset contains data points for 939 unique congregations (a maximum of 656 nodes per time step), including changes of affiliation (e.g., from closed to open Brethren). It contains information for most years, but also contains some gaps (2013, 2019, 2022). The dataset presents two challenges that are important to our analysis: First, the data is sparse because information is no longer available or was never published. The relationships were computed using proximal point analysis, i.e. based only on geographic information. Second, the movement soon split into several subgroups, e.g., exclusive and open Brethren, and congregations may have changed their affiliation over time, making the data even more complex. It is therefore necessary to provide an analysis of the groups within the networks. The density of the entire network is $d(S_2) \approx 0.016$, the network is not connected, and the average clustering coefficient is $C(S_2) \approx 0.63$.

The third network S_3 is a high energy physics phenomenology citation network extracted from the arXiv e-print from January 1993 to April 2003, see (Gehrke et al. 2003; Leskovec et al. 2005). The network contains 34,546 papers and 421,578 citations within the network. While the data includes timestamps, we will analyze the information categorized by year. The density of the whole network is $d(S_3) \approx 0.0007$, the network is not connected, and the average clustering coefficient is $C(S_3) \approx 0.28$.

Our work considers very small (S_1) and very large (S_3) networks with very different topologies. While the density of S_1 is high, the other networks have a very low density. The average clustering coefficient is very high for S_2 ; only S_1 is connected.

Table 1 An overview of the top five and bottom five nodes in terms of importance and change value is provided for the network on Luke's Gospel S_1 considering degree centrality

Node	Importance
Jesus	1.362810
Elisabeth	0.612810
Zacharias	0.514595
Joseph	0.411654
John the Baptist	0.389595
...	
Simon (Pharisee)	0.018797
Maria	− 0.195489
Pharisees	− 0.231203
Teachers of the law	− 0.266917
Disciples of John	− 0.338346
Node	Change
Jesus	0.366071
John the Baptist	0.160714
Simon Peter	0.107143
Jerusalem	0.107143
Simon	0.107143
...	
Maria	0.000000
Joseph	0.000000
Simon (Pharisee)	0.000000
Simon's mother-in-law	0.000000
Anna	0.000000

Luke's Gospel

Within New Testament research, the relationship between Jesus, the main figure in Luke's Gospel, and John the Baptist and his disciples remains an open question. See (Brownlee 1955; Chauchot 2021) for further discussion. Figure 4 displays the degree, closeness, and betweenness centrality development of both figures. No surprises are apparent from the figure. Notably, Jesus exhibits a considerably high centrality value in all three measures. John the Baptist, however, shows minimal significance since he is only a peripheral character in the latter half of Luke's Gospel. Both actors receive above-average ratings.

This does not really help to understand the longitudinal development of actors in Luke's Gospel. It also remains an open question: Which actors would be interesting to study? For this, we present an overview of the top five and bottom five nodes in terms of importance and change value in Table 1. Here, we focus on degree centrality, although this analysis could be carried out for other centrality measures as well. For such a small network, however, we found out that the results are similar.

The overall importance is expected: Jesus and John the Baptist are key characters, while the disciples of John are merely supporting roles. It is noteworthy that this perspective overvalues characters who are important solely in the initial stages of the narrative. For instance, Elisabeth and Zacharias hold significance in the early parts of

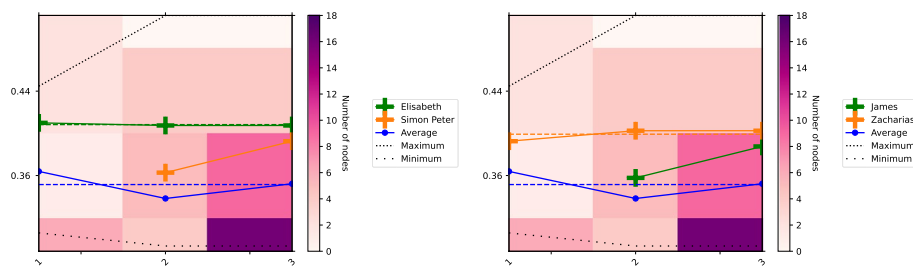


Fig. 5 Distribution of centrality measures for S_1 over time is illustrated in a heatmap. The data is grouped into 10 bins between 0 and 1. These two figures show degree centrality, for Elisabeth and Simon Peter (left) and Zacharias and James (right)

the narrative, but lose prominence as the story progresses. Nevertheless, they remain integral to the larger network.

This major disadvantage is crucial for the overall change: If figures such as Maria and Herod appear later, they cannot attain a higher value. Consequently, it is unsurprising that Jesus and John the Baptist possess a greater change value. In the left panel of Fig. 5, we can contrast Elisabeth, who holds a high importance value, and Simon Peter, who, based on New Testament research, is a significant character in Luke's Gospel and also boasts a high change value. In Fig. 5(left), we compare Zacharias, who has a high importance value, with James, a disciple of Jesus, who has a change value of 0.10. It's crucial to note that the network topology doesn't undergo significant changes from timestep 2 to 3. Therefore, marginal actors continue to have high values, and it's difficult but not impossible to identify the actors pushing the narrative.

Including more timepoints could have enhanced the results, as we will demonstrate in our next analysis. Thus, to conduct longitudinal analysis, it is essential to account for breaks in the narrative. While this study confirms alignment with New Testament research, there are no significant findings and they are difficult to interpret.

Figures 4 and 5 illustrate that the importance of nodes increases over time. This is due to the fact that more nodes have been observed by the end of the time interval. Consequently, it can be concluded that the notions presented in these plots are heavily biased in time. This is a common issue when dealing with temporal networks. However, these plots clearly demonstrate these biases by plotting the maximum, minimum, and average measures over time.

The assemblies of Brethren in Germany

The development of Brethren assemblies in Germany began in 1849, originating from Ireland. For more on this subject, see (Holthaus et al. 2003; Geldbach 2023). The Brethren movement later divided into subgroups, such as the Exclusive Brethren and the Open Brethren. The international history of the Brethren movement is intricately complex and closely associated with the Baptist church in Germany, as outlined in Liese (2007), Geldbach (2023). In Germany, there are a variety of Brethren churches, including those affiliated with the BEFG, Open Brethren, Exclusive Brethren, and Brethren congregations that have separated themselves from the Exclusive Brethren, known as the "blockfree" or "blockfreie Gemeinden". The conservative Raven Brethren

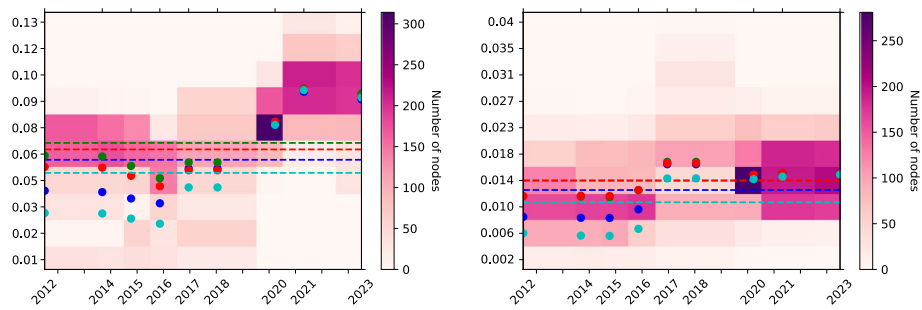


Fig. 6 Distribution of centrality measures for S_2 over time is illustrated in a heatmap. The data is grouped into 10 bins between 0 and 1. The two centrality measures, closeness and degree centrality are represented on the left and right, respectively. We chose to group by exclusive Brethren (green), open Brethren (red) and thus congregations not affiliated to a church (cyan). Average values are plotted in blue

Table 2 An overview of the top five and bottom five nodes in terms of importance and change value is provided for the network on the assemblies of Brethren in Germany S_2 considering degree centrality

Node	Importance
Dautphetal-Hommertshausen	0.159584
Steffenberg-Oberhoerlen	0.146634
Angelburg-Goennern	0.129008
Zwickau-Oberhohndorf (BF)	0.117686
Zwickau-Oberhohndorf (FB)	0.117686
...	
Memmingen	- 0.088595
Bietigheim-Bissingen	- 0.101346
Niederrossbach/Westerwald	- 0.117263
Frankfurt am Main	- 0.117263
Hof/Saale	- 0.117263
Node	Change
Duisburg-Homberg	0.045455
Hagen	0.042208
Altena	0.042208
Dillenburg-Donsbach	0.036611
Muenchen-Schwabing	0.035874
...	
Wesel-Lackhausen	0.000000
Hattingen	0.000000
Muelheim/Ruhr-Duempten	0.000000
Muelheim/Ruhr	0.000000
Essen-Borbeck	0.000000

have only a small number of congregations. Complicating matters further, some congregations may have changed their affiliations over time.

In Fig. 6, we present the closeness and degree centrality evolution of the groups. Notably, only blockfree congregations exhibit below-average centrality while exclusive

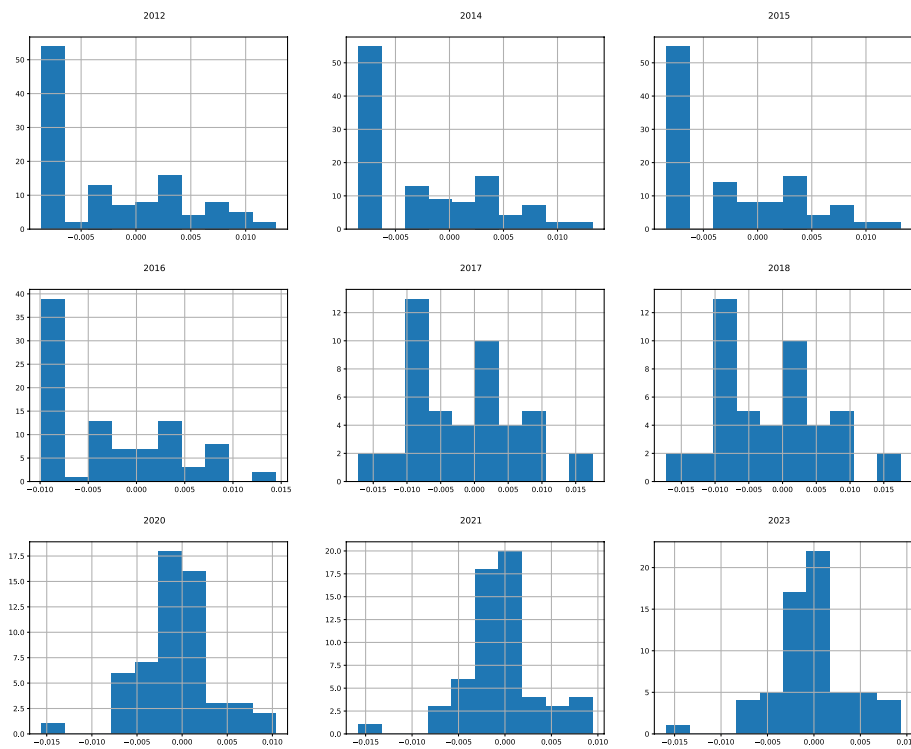


Fig. 7 Histograms showing the distribution of importance values over the time provided for the network on the assemblies of Brethren in Germany S_2 considering degree centrality

Brethren have the highest values. Another crucial insight is that from 2020 onwards, all groups share a common value, making them indistinguishable from each other.

Table 2 displays the top five and bottom five nodes ordered by importance and change value for individual congregations. The values, in contrast to S_1 , are relatively low. Grouping these values by congregation type reveals the following mean importance values: exclusive Brethren 0.012, open Brethren 0.12, and block-free congregations -0.01 ; the mean change values for these groups are as follows: exclusive Brethren 0.012, open Brethren 0.011, and block-free congregations 0.009. These findings are consistent with those in Fig. 6(right): blockfree congregations have lower values, but the overall change appears to be similar across all groups.

However, an examination of the distribution of importance (Fig. 7) and change (Fig. 8) over time reveals some intriguing findings. While both values begin with a considerable number of low values, they appear to approach a normal distribution over time. However, it is uncertain if this phenomenon is universal to social networks. For both measures, it seems that over time, a greater number of nodes exhibit above-average values, while initially, a greater number of nodes have below-average values.

Therefore, for larger networks that possess inherent group data, it is not only reasonable but also insightful to group the data. Although analyzing the total change and importance values can prove difficult, grouping the data simplifies the understanding and summarization of the longitudinal development. Moving forward, let us now examine extremely large networks that lack inherent group data.

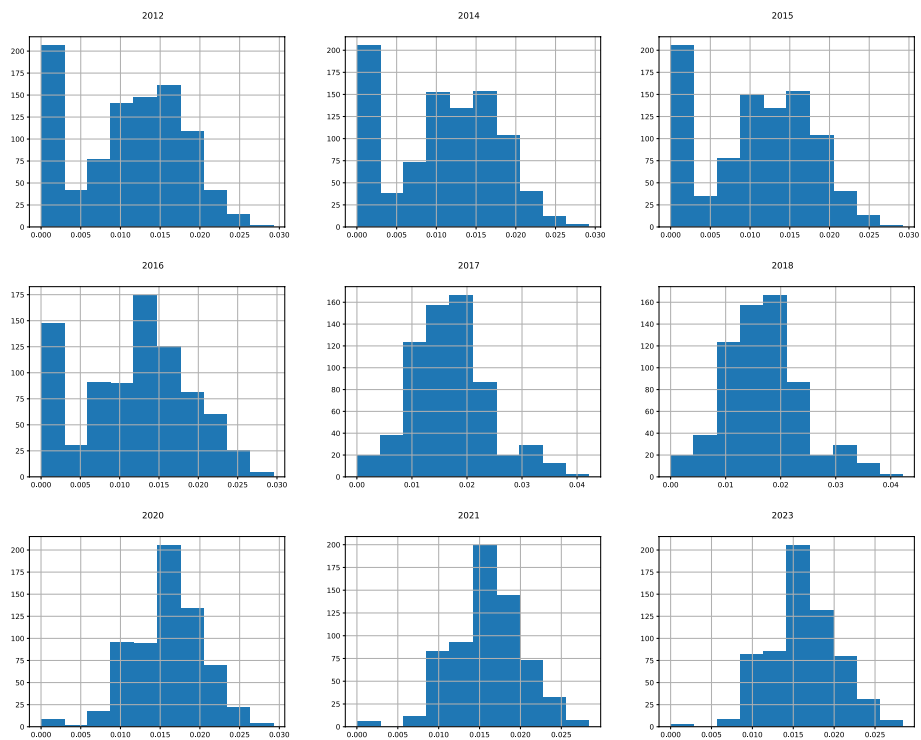


Fig. 8 Histograms showing the distribution of importance values over the time provided for the network on the assemblies of Brethren in Germany S_2 considering degree centrality

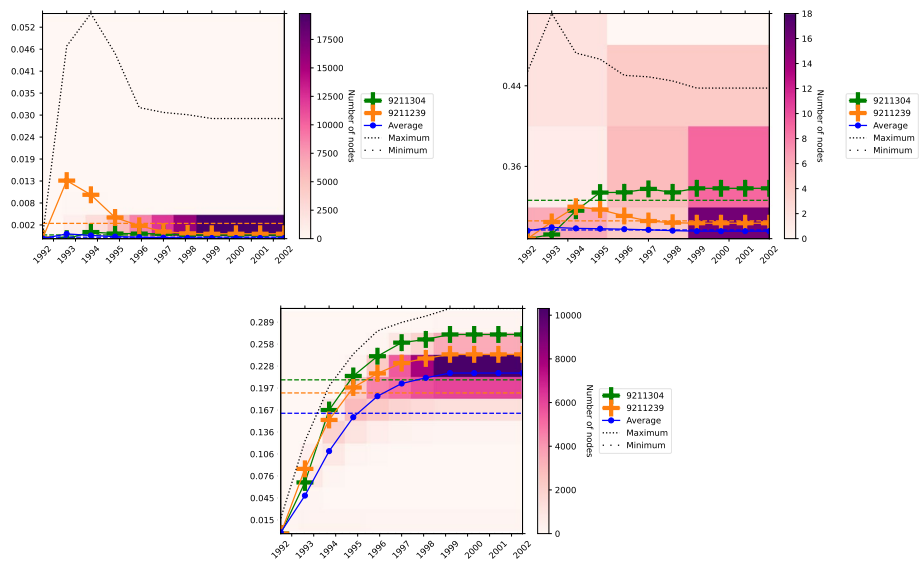


Fig. 9 Distribution of betweenness, degree and closeness centrality for S_3 and randomly chosen nodes over time is illustrated in a heatmap. The data is grouped into 10 bins between 0 and 1. Left: Randomly chosen nodes; Middle: Two nodes with highest importance; Right: Two nodes with highest change

Table 3 An overview of the top five and bottom five nodes in terms of importance and change value is provided for the High Energy Physics Phenomenology citation network S_3 considering betweenness centrality

Node	Importance
9302210	0.286474
9306320	0.222511
9303255	0.199420
9603208	0.188684
9410404	0.186773
...	
9209206	− 0.003858
9205218	− 0.003858
9210207	− 0.003858
9211330	− 0.003858
9208204	− 0.003858
Node	Change
9209262	0.092262
9303255	0.089568
9302210	0.087727
9203203	0.080529
9310366	0.078702
...	
9601270	0.000000
9307346	0.000000
9203208	0.000000
9311330	0.000000
9309254	0.000000

High energy physics phenomenology citation network

The high energy physics phenomenology citation network has no inherent groups, but contains 34,546 nodes, making manual data inspection impractical. Comparing randomly selected nodes, such as for betweenness centrality, proves unfruitful in identifying noteworthy actors, see Fig. 9. Nevertheless, these figures facilitate comprehending the overall shift in centrality measures over time. While betweenness only has a few nodes with a high degree, the number of nodes with a low degree increases significantly over time. However, for closeness centrality, the opposite is true. Since the network has a low density, it is expected to find a lot of nodes with a low varying degree.

However, Table 3 allows us to identify nodes that have high importance or have undergone significant changes. We present the results for the top two ranked nodes in Fig. 12. It is apparent that publications 9,302,210 and 9,306,320 are of great importance, their betweenness centrality values have significantly improved and remain close to the maximum value. On the other hand, publications 9,209,262 and 9,303,255 start with high values, but experience a considerable drop, although they still exceed the average betweenness centrality value. Thus, in identifying nodes with unique

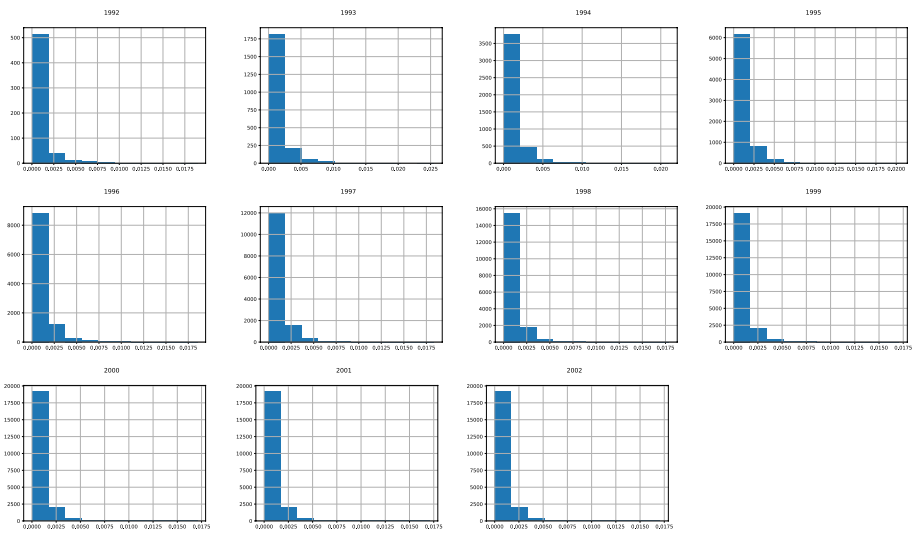


Fig. 10 Histograms showing the distribution of importance values over the time provided for the High Energy Physics Phenomenology citation network S_3 considering degree centrality

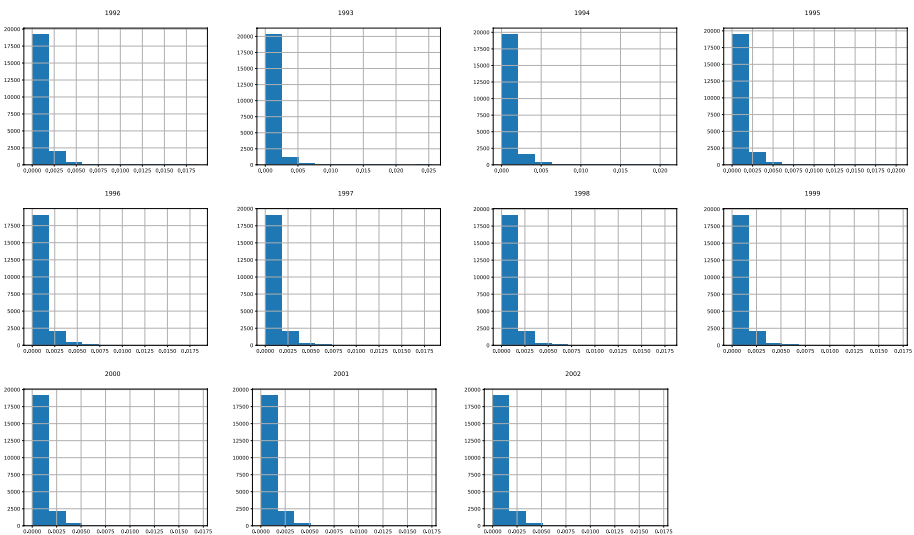


Fig. 11 Histograms showing the distribution of importance values over the time provided for the High Energy Physics Phenomenology citation network S_3 considering degree centrality

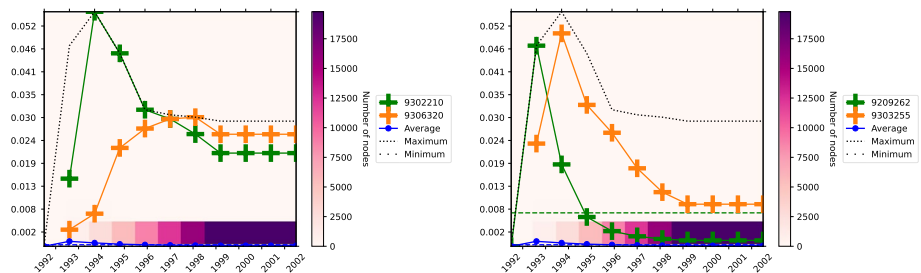


Fig. 12 Distribution of betweenness centrality for S_3 over time is illustrated in a heatmap. The data is grouped into 10 bins between 0 and 1. Left: Two nodes with highest importance; Right: Two nodes with highest change

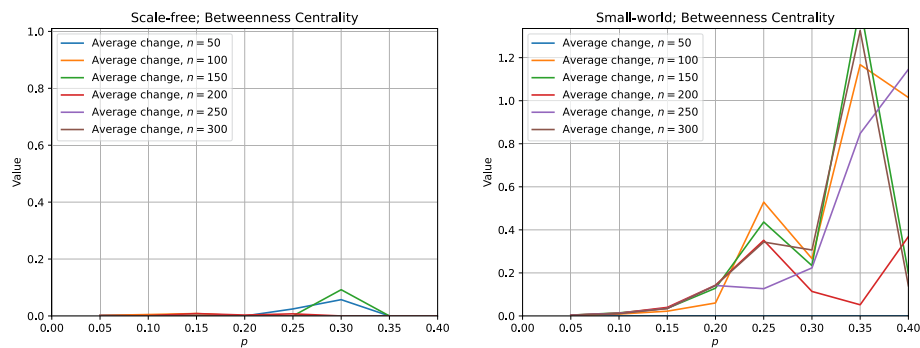


Fig. 13 Average change ratio for betweenness centrality on different runs of random graphs (scale-free, left, and small-world, right)

properties, importance and change are critical considerations for networks that comprise numerous nodes.

Again we will provide an examination of the distribution of importance (Fig. 10) and change (Fig. 11) over time. The findings on this large network differ from those provided for network S_2 . Over all years, there are numerous nodes with a low value for both measures, and this does not appear to be changing. The values appear to follow a power-law distribution. This clearly demonstrates that there is no universal solution for analyzing network change and importance. Instead, it is essential to apply visualizations that align with the network topology.

However, it is evident that this method overestimates nodes with a long lifespan, which we will discuss as an open research question. Here, normalizing by the lifespan of a node might help. Its bias or reasonable understanding heavily depends on the initial research question. This issue was problematic for networks with only a few time points, as seen in S_1 , but not a concern for S_3 where citations develop over time. However, these networks offer real-world data, and a crucial next step is to assess these methods on random graphs with specific, albeit artificial, properties.

Random graphs

We evaluate the degree centrality and betweenness centrality on random graphs, see Appendix C for details. First, we consider scale-free networks with n nodes, see (Jackson 2010). With this, we create a series of random Graphs $\mathcal{G}^s(n, i, p)$ which creates one initial scale-free network with n nodes and $i - 1$ more random graphs with a probability of $p/2$ for each node and edge to be deleted and $p/2$ for each node and edge to be deleted and a new one created. In addition, we consider scale-free networks and create a series of random Graphs $\mathcal{G}^w(n, i, p)$ which starts with one initial small world network with n nodes and $i - 1$ more random graphs with a probability of $p/2$ for each node and edge to be deleted and $p/2$ for each node and edge to be deleted and a new one created.

We will evaluate importance and change for degree, closeness and betweenness centrality on the following random graph series:

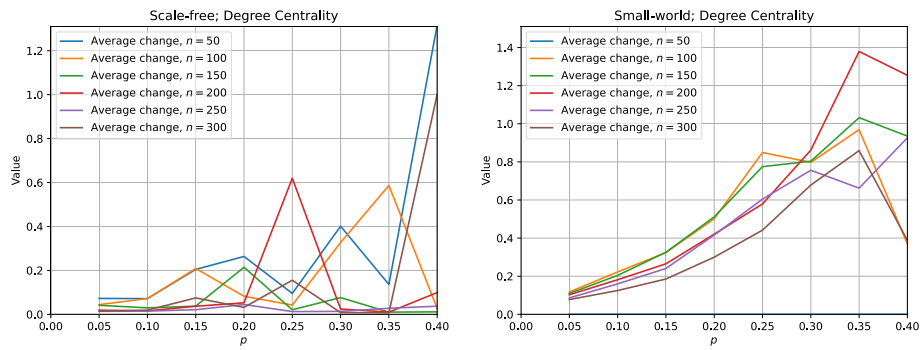


Fig. 14 Average change ratio for degree centrality on different runs of random graphs (scale-free, left, and small-world, right)

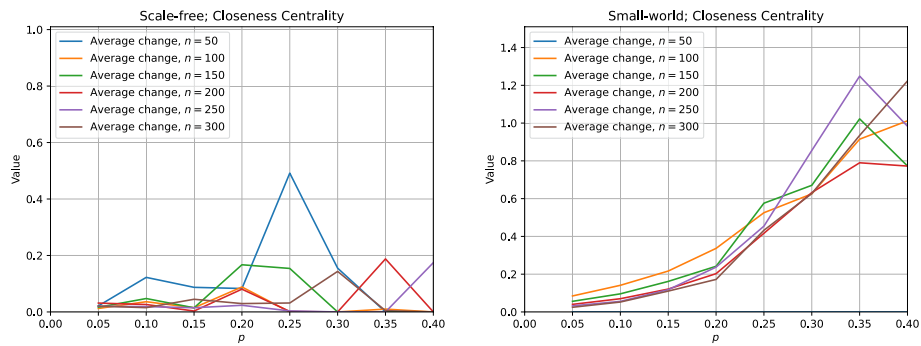


Fig. 15 Average change ratio for closeness centrality on different runs of random graphs (scale-free, left, and small-world, right)

- $\mathcal{G}^s(n, 15, p)$, $p \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4\}$,
 $n \in \{50, 100, 150, 200, 250, 300\}$
- $\mathcal{G}^w(n, 15, p)$, $p \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.45\}$,
 $n \in \{50, 100, 150, 200, 250, 300\}$

However, the overall importance average for all these graphs is close to zero, indicating that both random graphs display an artificial scenario in which the values for all three centrality measures are distributed almost equally without many outliers.

The results for the average change ratio are showcased in Figs. 13, 14, 15. It is worth noting that the two distinct random graphs reveal diverse behaviors, and various centrality measures exhibit different outcomes. Overall, scale-free networks typically have a low change ratio, even for high values of p . Small-world networks present a clearer view: the average change ratio increases for larger p , but then drops after a certain threshold. Additionally, the change in degree centrality appears to be the least influenced by graph topology.

Further research may be conducted in the direction of random graphs. The extent to which these simple random graph models reflect real-world networks remains unclear, and the differences between these two analytical approaches serve to highlight this question. Furthermore, our findings underscore the necessity for further investigation into

graph topology, particularly with regard to the viability of the methodologies proposed in this study.

Summarizing the findings on random graphs underscores the necessity of investigating the impact of graph structures on network robustness and the effects of network changes on centrality measures of other nodes. Our analysis of real-world networks demonstrated the benefits of conducting longitudinal studies across multiple time steps, as analyzing only a few timesteps can hinder identification of the proposed measures' significance. For larger networks and multiple time steps, nodes with distinct properties, significance, and variability can indeed be identified. Nonetheless, it is important to note that comparing values across various networks is not possible, and analyzing longitudinal social networks still requires manual effort.

Discussion and outlook

Various methods are available for handling longitudinal data in networks. However, all of them result in the algorithms analyzing the data being biased. In accordance with Dörpinghaus et al. (2023), we have proposed an expanded answer to the query of how to universally model longitudinal social networks in a single graph. We have enhanced the techniques through the provision of metrics for analyzing larger networks – specifically node importance and change – in addition to the evaluation and comparison of actor groups. The visual analysis was complemented by the use of histogram plots, which illustrate the evolution of the data over time, should any changes be observed.

The measures naturally extend centrality measures to longitudinal networks, but they require reinterpreting these results and adapting algorithms. It also remains unclear how well these measures work for different types of networks, for example since the proposed method overestimates nodes with a long lifespan and the network topology can vary. Three real-world examples demonstrate that small networks or those with only a few timesteps require significant manual effort and deeper understanding of the network, whereas the proposed measures can be reasonably utilized in larger networks. At this stage, it is not possible to determine whether an universal solution that incorporates additional normalization or other techniques can be established. However, our approach is effective for all centrality measures, but we have only focused on betweenness, closeness, and degree centrality. Further research should explore additional centrality measures and methods such as community detection, as addressing these questions is crucial for comprehending the algorithmic difficulties of temporal data in social network analysis. Nevertheless, we could address the third research question, namely how these analytical methods can be applied to real-world networks. It must be acknowledged that there is no universal solution; rather, the researcher must select the most appropriate methods for the network in question, taking into account its size and topology.

Our second question was whether we could approximate the change in centrality measures, change and importance over time. Our initial investigation (Dörpinghaus et al. 2023) yielded bounds based on degree centrality. Yet, prior knowledge of the change ratio p between different time points is necessary for these limits. As p increases, the sharpness of these boundaries diminishes. Further research should investigate various types of bounds, especially for other centrality measures. Additionally, analyzing

graph substructures that impact the temporal behavior of centrality measures could be beneficial, particularly when p is unknown.

In general, it does not appear that larger networks are more easily followed longitudinally. This claim would contradict the prevailing consensus in computer and social sciences, which posits that small, dense networks offer deeper analysis opportunities than large, “poor” networks. Additionally, the dynamic structure of a graph appears to play an important role in its analysis, although this was not explored in this study. For instance, a stable graph with marginal additions over time and an interaction network, where the graph varies considerably from one time step to the next and is typically close to empty at each time t , would yield disparate results. This makes it challenging to assess the generality of our claim and is a question for further research, although most networks appear to follow the first type.

Overall, we have presented evidence that the proposed methods offer a comprehensive understanding and practical measures for examining longitudinal networks. The primary benefit of our approach is that it takes into account the natural extension of centrality measures without relying on specific graphs, such as stream graphs. As demonstrated, this means that our approach can be readily applied to pre-existing networks.

Rewriting algorithms for analyzing longitudinal social networks and reinterpreting measures and algorithms requires interdisciplinary discussions between scientific domains. Thus, our paper advocates for more interdisciplinary exchange, especially among mathematics, computer science, social sciences, and the humanities.

Appendix A: Temporal graph structures

Similar to the approaches of Nicosia et al. (2013), Sizemore and Bassett (2018) we can study time-respecting structures in a graph as discussed in Dörpinghaus et al. (2023). However, Definition 1 of temporal social networks makes it easier to generalise graph structures as it keeps the generic definition of a graph.

A path p in a graph $H = (V, E)$ is a set of vertices v_1, \dots, v_t , $t \in \mathbb{N}$, for example written as

$$p = [v_1, \dots, v_t],$$

where $(v_i, v_{i+1}) \in E$ for $i \in \{1, \dots, t-1\}$. However, to track the meaning of time in a temporal social network $G = (V, E, T)$, we define p^t , which is a path p that exists at time t . In turn, we define $t(p)$ as the interval of time in which the path p exists in G .

Unless otherwise noted, we use G for a temporal social network $G = (V, E, T)$ and H for any undirected graph.

We can add this generic notation for other structures as well. For example, we denote the time-respecting degree of a node v by $d^t(v)$. In this way, we get a series of *temporal degree centrality measures* (TDC) for a node $v \in V$ denoted by

$$dc^t(v) = \frac{d^t(v)}{n-1}.$$

In addition, we can analyse the *temporal degree distribution* which tells us about the network structure since we can distinguish between sparsely and densely connected networks.

Betweenness centrality (BC) was first introduced by Freeman (1977) and considers other indirect links, see (Schweizer 1996). Given a node v , $bc(v)$ is defined as

$$bc(v) = \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v(k, j)}{P(k, j)} \cdot \frac{2}{(n-1)(n-2)},$$

that is, we compute the number of all shortest paths $P_v(k, j)$ in a network for all starting and ending nodes $k, j \in V$ that pass through v . Let $P(k, j)$ denote the total number of shortest paths between k and j . Then the importance of v is given by the ratio of the two values of P_v and P . Again, for any time $t \in \mathcal{T}$ we may set $P_v^t(k, j)$ and $P^t(k, j)$ accordingly, such that

$$bc^t(v) = \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v^t(k, j)}{P^t(k, j)} \cdot \frac{2}{(n-1)(n-2)}$$

defines the series of *temporal betweenness centrality* (TBC). This definition is similar to that of Sizemore and Bassett (2018), who, however, used the concept of fastest paths.

We will proceed similarly with *closeness centrality* (CC). Given a node $i \in V$ we can compute the average distance between the first and other nodes $j \in V$ with $\sum_{j \neq i} d(i, j)$, where $d(i, j)$ denotes the length of a shortest path between i and j . Then, according to Jackson (2010), we can compute closeness-centrality as follows:

$$cl(v) = \frac{n-1}{\sum_{u \in V} d(u, v)}.$$

Again, with a definition of $d^t(i, j)$ for the length of a shortest path at time $t \in \mathcal{T}$ at hand, we can define *temporal closeness centrality* (TCC) as

$$cl^t(v) = \frac{n-1}{\sum_{u \in V} d^t(u, v)}.$$

However, these definitions are currently not more than a containment of well-known centrality measures on time snapshots of the temporal social network. They allow an interpretation of these snapshots, comparable to static social networks, and they provide a series of centrality measures that can be interpreted as the progression of these measures over time.

For social networks, perceiving the world with as few snapshots as possible is most feasible. Other approaches, e.g. defining paths closely so that they could split up from one time to another, if the interval is so small that an event lasts less, is often necessary to model traffic (Pan and Saramäki 2011). Social interaction, on the other hand, does usually change on the basis of longer lasting events. This is a crucial observation, because computing temporal paths with increasing timestamps from one node to the next is computationally hard, see (Santoro and Sarpe 2022).

While interdisciplinary approaches are available, applications from humanities and in particular historical networks research lead to a different perspective on data. For example, a closed organization may still have an influence on parts of the network or may be referred to later. However, with our novel approach, we will evaluate the behavior of analysis methods like centrality measures and community detection and discuss limitations and challenges for further research.

Appendix B: Random graphs

For our analysis, we rely on random graphs. The *degree distribution* provides us with information about the network structure since we can distinguish between sparsely and densely connected networks. In social network analysis (SNA), the following two graphs are widely considered:

Definition 4 (Scale-free network) A network is scale-free if the fraction of nodes with degree s follows a power law $s^{-\alpha}$, where $\alpha > 1$.

Definition 5 (Small world network (Watts 1999)) Let $G = (V, E)$ be a connected graph with n nodes and average node degree k . Then G is a small-world network if $k \ll n$ and $k \gg 1$.

Bollobás et al. (2003) introduced a widely used graph model with three random parameters $\alpha + \beta + \gamma = 1$. These values define probabilities and thus define attachment rules to add new vertices between either existing or new nodes. This model allows loops and multiple edges, where a loop denotes one edge where the endvertices are identical, and multiple edges denote a finite number of edges that share the same endvertices. Thus, we convert the random graphs to undirected graphs. For testing purposes, we scale the number of nodes n and use $\alpha = 0.41$, $\beta = 0.54$, and $\gamma = 0.05$. This random graph model is generic and feasible for computer simulations for measuring and evaluation purposes, see (Bollobás and Riordan 2003; Kivelä et al. 2014).

One of the core concepts important in social network research is the graph diameter $D(G)$. From the 1960s on, it was widely discussed whether the average path length of social networks is near six, see (Milgram 1967). However, there is an ongoing discussion on this issue, see for example (Watts and Strogatz 1998; Kleinfeld 2002). However, it was shown that in a scale-free network the diameter is always lower than $\log(n)$, and if the fixed number m of earlier vertices is larger than 1, in general the diameter is lower than $\frac{\log(n)}{\log \log(n)}$, see (Riordan 2004). Here, n describes not only the number of steps to create the random graph, but also the number of nodes in the graph. While the connection between a particular graph and a particular diameter is quite complex, see (Ma et al. 2020), we can rely on these bounds. For small-world random graphs we find (Gu et al. 2013) the almost surely upper bound $D(G) \leq \frac{72}{p} \log^2 n$ while (Martel and Nguyen 2004) proved the diameter is usually bound by $\log(n)$.

The diameter of a scale-free graph is in general quite low, while in small-world graphs it is bound by $\log(n)$. However, we may expect random graphs to have a

different behavior from real-world social networks. Thus, for some of the following proofs we will assume that $D(G) \leq 5$.

Acknowledgements

Not applicable.

Author contributions

Jens Dörpinghaus: Conceptualization, Conceptualization, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review and Editing, Visualization, Project administration Vera Weil: Conceptualization, Conceptualization, Validation, Formal analysis, Investigation, Writing - Review and Editing Martin Sommer: Conceptualization, Validation, Formal analysis, Investigation, Writing - Review and Editing.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials

The datasets analysed during the current study are available in Dörpinghaus (XXXX, 2023); Gehrke et al. (2003); Leskovec et al. (2005). The random datasets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 9 May 2024 Accepted: 19 August 2024

Published online: 29 August 2024

References

- Barats C, Schafer V, Fickers A (2020) Fading away... the challenge of sustainability in digital studies. *DHQ: Digital Hum Quart* 14(3)
- Bellotti E (2014) *Qualitative networks: mixed methods in sociological research*. Routledge, London
- Bollobás B, Borgs C, Chayes JT, Riordan O (2003) Directed scale-free graphs. In: *SODA* 3:132–139
- Bollobás B, Riordan OM (2003) Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, 1–34
- Brownlee WH (1955) John the baptist in the new light of ancient scrolls. *Interpretation* 9(1):71–90
- Cencetti G, Battiston F, Lepri B, Karsai M (2021) Temporal properties of higher-order interactions in social networks. *Sci Rep* 11(1):7028
- Chauchot CM (2021) *John the Baptist as a rewritten figure in Luke-Acts*. Routledge, London
- Cinaglia P, Cannataro M (2022) Network alignment and motif discovery in dynamic networks. *Netw Model Anal Health Inform Bioinform* 11(1):38
- Dörpinghaus J (2023) Algorithmic challenges towards temporal data in social network analysis: a case study on the assemblies of Brethren in Germany. In: *Joint conference graphs and networks in the fourth dimension—time and temporality as categories of connectedness (GraphNR 2023)*
- Dörpinghaus, J. Social network analysis of Luke-Acts. <https://doi.org/10.5281/zenodo.7152121>
- Dörpinghaus J (2022) Social network analysis of luke-acts. Zenodo. <https://doi.org/10.5281/zenodo.7152121>
- Dörpinghaus J, Klante S, Christian M, Meigen C, Düing C (2022) From social networks to knowledge graphs: a plea for interdisciplinary approaches. *Soc Sci Hum Open* 6(1):100337
- Dörpinghaus J, Weil V, Sommer MW (2023) Towards modelling and analysis of longitudinal social networks. *Ann Comput Sci Inf Syst* 37:81–89
- Dörpinghaus J, Stenschke C (2021) Ein kollaborativer Workflow zur historischen Netzwerkanalyse mit Open Source Software. *Proceedings of the 13th free and open source conference*
- Espinosa-Rada A, Bellotti E, Everett MG, Stadtfeld C (2024) Co-evolution of a socio-cognitive scientific network: A case study of citation dynamics among astronomers. *Soc Netw* 78:92–108
- Everett MG, Borgatti SP (1999) The centrality of groups and classes. *J Math Sociol* 23(3):181–201
- Franzke M, Emrich T, Züfle A, Renz M (2018) Pattern search in temporal social networks. In: *Proceedings of the 21st international conference on extending database technology*
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry*, 35–41
- Gehrke J, Ginsparg P, Kleinberg J (2003) Overview of the 2003 kdd cup. *ACM SIGKDD Explorations Newsl* 5(2):149–151
- Geldbach E (2023) *Der bund evangelisch-freikirchlicher gemeinden (befg)*. In: *Handbuch der Religionen*, pp. 1–20. Westarp Science Fachverlag
- Grüninger M (2011) Verification of the owl-time ontology. In: *The semantic web—ISWC 2011: 10th international semantic web conference, Bonn, Germany, October 23–27, 2011, Proceedings, Part 1* 10, pp. 225–240. Springer
- Gu L, Huang HL, Zhang XD (2013) The clustering coefficient and the diameter of small-world networks. *Acta Math Sinica, Engl Ser* 29(1):199–208
- Hanneke S, Fu W, Xing EP (2010) Discrete temporal models of social networks. *Electron J Stat* 4:585–605
- Hobbs JR, Pan F (2004) An ontology of time for the semantic web. *ACM Trans Asian Lang Inf Process (TALIP)* 3(1):66–85
- Holme P, Saramäki J (2012) Temporal networks. *Phys Rep* 519(3):97–125

- Holme P, Saramäki J (2019) A map of approaches to temporal networks. *Temporal Netw Theory*, 1–24
- Holthaus S, Vanheiden K-H, Schmidt M, Jaeger H (2003) 150 Jahre Brüderbewegung in Deutschland
- Jackson MO (2010) Social and economic networks. University Press, Princeton. <https://doi.org/10.1515/9781400833993>
- Kessler, V.: 'women, forgive us': A german case study. *HTS Teologiese Studies/Theological Studies* 78(2) (2022)
- Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. *J Complex Netw* 2(3):203–271
- Kleinfeld JS (2002) The small world problem. *Society* 39(2):61–66
- Latapy M, Viard T, Magnien C (2018) Stream graphs and link streams for the modeling of interactions over time. *Soc Netw Anal Min* 8:1–29
- Latapy M, Magnien C, Viard T (2019) Weighted, bipartite, or directed stream graphs for the modeling of temporal networks. *Temporal Netw Theory*, 49–64
- Lazega E (2016) Synchronization costs in the organizational society: intermediary relational infrastructures in the dynamics of multilevel networks. *Multilevel network analysis for the social sciences: theory, methods and applications*, 47–77
- Lazega E (2017) Organized mobility and relational turnover as context for social mechanisms: a dynamic invariant at the heart of stability from movement. *Knowl Netw*, 119–142
- Lazega E, Snijders TA (2015) *Multilevel network analysis for the social sciences: theory, methods and applications*, vol 12. Springer, Berlin
- Lehmann S (2019) Fundamental structures in temporal communication networks. *Temporal Netw Theory*, 25–48
- Leidwanger J, Knappett C, Arnaud P, Arthur P, Blake E, Broodbank C, Brughmans T, Evans T, Graham S, Greene ES, et al (2014) A manifesto for the study of ancient mediterranean maritime networks. *Antiquity* 88(342)
- Lemercier C (2015) Taking time seriously. how do we deal with change in historical networks? In: *Knoten und Kanten III. Soziale Netzwerkanalyse in Geschichts- und Politikforschung*, pp. 183–211. Transcript, ???
- Leskovec J, Kleinberg J, Faloutsos C (2005) Graphs over time: densification laws, shrinking diameters and possible explanations. In: *Proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining*, pp 177–187
- Liese A (2007) Taufverständnisse in der brüderbewegung. *Zeitschrift für Theologie und Gemeinde* 12:272–286
- Ma F, Wang X, Wang P (2020) Scale-free networks with invariable diameter and density feature: Counterexamples. *Phys Rev E* 101(2):022315
- Martel C, Nguyen V (2004) Analyzing kleinberg's (and other) small-world models. In: *Proceedings of the twenty-third annual ACM symposium on principles of distributed computing*, pp 179–188
- Milgram S (1967) The small world problem. *Psychol Today* 2(1):60–67
- Naima M, Latapy M, Magnien C (2023) Temporal betweenness centrality on shortest paths variants. arXiv preprint [arXiv:2305.01080](https://arxiv.org/abs/2305.01080)
- Nicosia V, Tang J, Mascolo C, Musolesi M, Russo G, Latora V (2013) Graph metrics for temporal networks. *Temporal Netw* 15–40
- Pan RK, Saramäki J (2011) Path lengths, correlations, and centrality in temporal networks. *Phys Rev E* 84(1):016105
- Peixoto TP, Rosvall M (2019) Modelling temporal networks with markov chains, community structures and change points. *Temporal Netw Theory*, 65–81
- Rasti S, Vogiatzis C (2022) Novel centrality metrics for studying essentiality in protein-protein interaction networks based on group structures. *Networks* 80(1):3–50
- Riedel F, Runkel S (2015) Understanding churchscapes: theology, geography and music of the closed brethren in Germany. *The changing world religion map: sacred places, identities, practices and politics*, 2753–2782
- Riordan O (2004) The diameter of a scale-free random graph. *Combinatorica* 24(1):5–34
- Ryan L, D'Angelo A (2018) Changing times: migrants' social network analysis and the challenges of longitudinal research. *Soc Netw* 53:148–158
- Santoro D, Sarpe I (2022) Onbra: Rigorous estimation of the temporal betweenness centrality in temporal networks. In: *Proceedings of the ACM web conference 2022*, pp 1579–1588
- Schafer R (2004) *Der aufbau von leitungsstrukturen in gemeindegründungsarbeiten der brüdergemeinden in deutschland*. PhD thesis
- Scholtes I, Wider N, Pftzner R, Garas A, Tessone CJ, Schweitzer F (2014) Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nat Commun* 5(1):5024
- Schweizer T (1996) *Muster Sozialer Ordnung: Netzwerkanalyse Als Fundament der Sozialethnologie*. D. Reimer, Berlin
- Sizemore AE, Bassett DS (2018) Dynamic graph metrics: tutorial, toolbox, and tale. *Neuroimage* 180:417–427
- Taylor D, Myers SA, Clauset A, Porter MA, Mucha PJ (2017) Eigenvector-based centrality measures for temporal networks. *Multiscale Model Simul* 15(1):537–574
- Valeriola S (2021) Can historians trust centrality? *J Historical Netw Res* 6(1)
- Watts DJ (1999) Networks, dynamics, and the small-world phenomenon. *Am J Sociol* 105(2):493–527
- Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442
- Xu KS, Hero AO (2013) Dynamic stochastic blockmodels: statistical models for time-evolving networks. In: *Social computing, behavioral-cultural modeling and prediction: 6th international conference, SBP 2013, Washington, DC, USA, April 2–5, 2013. Proceedings* 6, pp. 201–210. Springer
- Yi JS, Elmqvist N, Lee S (2010) Timematrix: analyzing temporal social networks using interactive matrix-based visualizations. *Int J Hum Comput Interact* 26(11–12):1031–1051
- Yu E-Y, Fu Y, Chen X, Xie M, Chen D-B (2020) Identifying critical nodes in temporal networks by network embedding. *Sci Rep* 10(1):12494

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.