



Regular Article

From social networks to knowledge graphs: A plea for interdisciplinary approaches

Jens Dörpinghaus^{a,*}, Sonja Klante^b, Martin Christian^b, Christof Meigen^c, Carsten Düing^d^a Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany^b German Institute for Adult Education (DIE), Leibniz Centre for Lifelong Learning, Bonn, Germany^c Leipzig Research Center for Civilization Diseases – LIFE Child, Leipzig, Germany^d Mathematical Institute, University Koblenz-Landau, Koblenz, Germany

ARTICLE INFO

Keywords:

Social Network Analysis

Knowledge Graphs

Digital Humanities

Linked Data

Knowledge networks

Graph database

ABSTRACT

Contextual information is a crucial factor in interdisciplinary research on social networks. At the same time, missing linked data and missing methods for large-scale knowledge networks are a serious limitation. Only few researches have addressed the questions between data science, graph theory and social network analysis. This paper seeks to discuss how this gap can be closed. For this, we use an interdisciplinary approach trying to improve the interdisciplinary exchange and method exchange. We investigate how new methods from computer science, in particular knowledge graphs, can be used within the field of SNA. The contributions of this paper are (1) a knowledge graph model representation of a multi-layer social network, (2) the showcase of examples from different domains and (3) a perspective for more interdisciplinary discussion. We present an innovative solution to discuss coming challenges in SNA with methods from the digital humanities, in particular computer science.

Social network analysis (SNA) is an emerging topic and is increasingly becoming a vital factor in other scientific domains. They play an important role in the social sciences and have been widely used for several decades, both in theory and in application. It is an important issue to understand social interactions and networks and how they influence society. In the last few years, there has been a growing interest in using social networks in historical sciences. Quite recently, considerable attention has been paid to social networks in religious studies and especially in narrative studies and theology.

Scientists have always seen SNA as a part of the humanities and for the past years there has been a rapid rise in the use of methods from the digital humanities which comprises humanities and computer science. Most interestingly, most concepts within SNA which could have been developed within the field of graph theory, were introduced within the humanities and social sciences. For example, the concept of centrality, introduced by Freeman 1977 ? goes back to the works of Bavelas, 1948, see ? Thus, these concepts are not borrowed from graph theory, but refer to a parallel development within the network paradigm. In the next few years, the interdisciplinary research with computer science is likely to become an important component within SNA. Few researchers have addressed the question of how SNA can be embedded in the field of

graph theory and how SNA can benefit from the research done in the field of data science, in particular algorithmic, knowledge graphs and graph databases.

Thus, this paper seeks to address how this gap can be closed. We investigate how new methods from computer science, in particular knowledge graphs, can be used within the field of SNA. See Fig. 1 for an illustration. The aims of our research are (1) to understand and underline the benefits of this new approach, (2) to showcase examples and (3) to emphasize the need for a more interdisciplinary discussion. We present an innovative solution to address coming challenges in SNA with methods from the digital humanities, in particular with methods from computer science.

However, we need to highlight the difference between the analysis of (online) social networks using web mining technologies and the social network analysis as method within (computational) social sciences and the humanities. It is not possible to define a precise border between both. Both share a common theoretical environment, providing an actor-centric data network. However, while the first covers internet applications and mainly emerged within the field of computer science, the latter is clearly located in the field of the (digital) humanities, and computational social sciences. This explains why a varying amount of methods

* Corresponding author.

E-mail addresses: jens.doerpinghaus@bibb.de (J. Dörpinghaus), klante@die-bonn.de (S. Klante), christian@die-bonn.de (M. Christian), cmeigen@life.uni-leipzig.de (C. Meigen), dueing@uni-koblenz.de (C. Düing).<https://doi.org/10.1016/j.ssaho.2022.100337>

Received 21 March 2022; Received in revised form 29 August 2022; Accepted 16 September 2022

Available online 15 October 2022

2590-2911/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

from computer science are used in the diverse field of SNA. However, this work is mainly interested on the gap between the humanities, social sciences, and computer science. While some methods which we cover in this paper have already been applied, as our examples show, we will provide scientific arguments why and how these interdisciplinary approaches work in all domains.

In addition, the applications of knowledge graph on the analysis of (online) social network are usually limited to applications from data science and artificial intelligence. For example, linking entities ?, sentiment analysis ? or stance detection ? However, we will try to provide an approach that helps to apply these approaches within, and embed them within the origin scientific domain, e.g. the humanities or social sciences.

This paper is divided into five sections. First, we give an overview about related work and in the second section about the background in all areas of research. We will also focus on historical network analysis (HNA), because it helps to underline two challenges: Data linking and data quality. A new methodology is described in the third section, where we discuss the way from Social Networks to Knowledge Graphs. The fourth section is dedicated to example use cases. We will present several examples from different domains and outline, how they benefit from this new approach. Our conclusions are drawn in the final section.

1. Related work

Since only little research has been done in this field, we list all material available, even if their tasks are only tangentially related. In the literature, there are few examples of Knowledge Graphs and SNA. Knowledge Graphs have been widely addressed in combination with the analysis of social network platforms like Twitter or Facebook. Collarana et al. (2018) for example, extract Knowledge Graphs from Web Sources to answer questions related to the social networks. See also the works of Gabrilovich and Usunier (2016).

knowledge discovery, whereas Hyvönen et al. (2019) highlight the importance of new relation between cultural heritage. In the last few years, more research has been focusing on ontologies and interoperable data, see Cristofaro et al. (2021).

Li et al. (2021) present a review of recent progress on social network structures. The single research dedicated to graph theory was presented by Ponten et al. (2007). And indeed, although in the early phase of SNA there was an ongoing discussion with graph theory, this is rather a recently new explored area: “The use of Graph Theory on social media data is a promising approach to identify emergent properties of the complex physical and cognitive interactions that occur between humans and nature.” (de Juan et al., 2021, p. 1) Graph theory and social networks were also presented during research on COVID-19, see Gaeta et al. (2021).

But implicitly, a lot of research already used Knowledge Graphs. By combining actors and persons with additional information like locations, linked data is used. While it is still very common, to use external data linked to a network¹ it is also possible to define different node types.² Zheng and Skillicorn (2017) for example, store additional data on edges and relations in the network. Hence, their main approach is to use different layers in a network, cf. Zheng and Skillicorn (2017, p. 31). But adding additional data to the network is mainly used in other research areas than the social sciences. Collar (2013), for example, as an archaeologist, was among the first combining religious studies and archaeology using SNA. In her work “Religious Networks in the Roman Empire” she examines why some cults and religions within the Roman Empire either vanished or became meaningless while maintaining the same popularity. She combines several network approaches, which lead to multiple data types. Collar describes several methods which are widely used in archaeology: primarily the *proximal point analysis* (PPA), the *Small-World* or *Complexity* theory, *Relational Space*, *Closeness* and *Betweenness Centrality* measures and *Social Network Analyses* (SNA), cf. Collar (2013, p. 29). It is not yet known whether the SNA can be

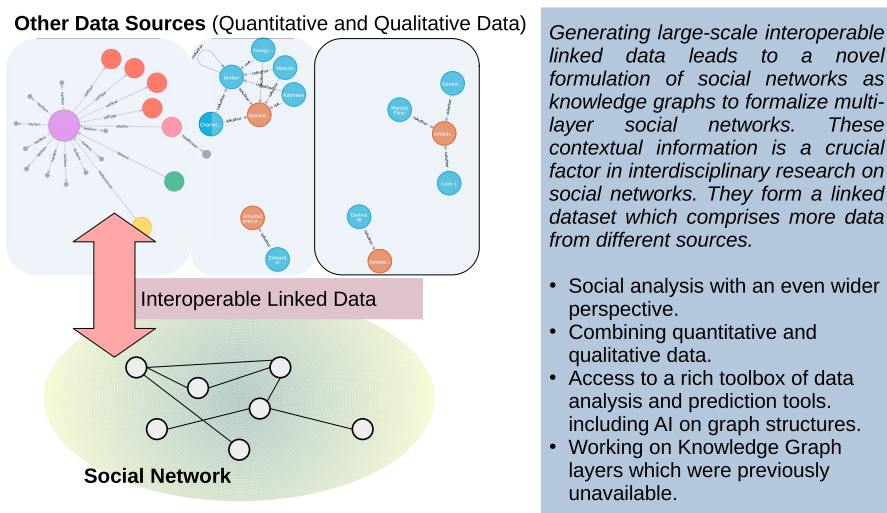


Fig. 1. The proposed method: Combining Social Network Analysis with interoperable data from other data sources and fields and building a Knowledge Graph. This allows the application of novel methods and heuristics on a wider range of data.¹

Several scholars like Keshavarzi (2021); Choudhary et al. (2021) have mentioned the overlap between social networks and Knowledge Graphs. But research has tended to focus on other aspects rather than investigating this overlap. An additional problem is that only few papers work on both aspects: SNA and Knowledge Graphs.

On the other hand, some preliminary work was carried out on Knowledge Graphs within Digital Humanities. Haslhofer et al. (2018) for example, point at the importance of connected knowledge and

¹ See for example the “risks” labeled to a social network in Suárez et al. (2021): As an external data, they are a property within the network but do not change the graph structure.

² Berhan et al. (2019) for example defined people, organizations, offices and students. Jackson (2010, p. 20) states: “It is important to emphasize that nodes can be individual people, firms, countries, or other organizations; a node can even be something like a web page belonging to a person or organization.”

generalized in all cases, since the lack of data is a challenge. Regarding New Testament research, the studies of [Duling \(2013\)](#), [Czachesz \(2011\)](#) and [White \(1992\)](#) should be mentioned: In particular, the works of Duling use different entities in the network. One of the main issues in what we do *not* know about a social network is in particular what we can not reproduce. Thus, “network analysis needs to be embedded within traditional research both to produce results and not to be ignored. For this achievement, Collar should be praised.” ([Van Nuffelen, 2015](#), p. 226).

There is an increasing sensitivity in other domains on the fact, that usually only some methodological parts of SNA are used. [Rollinger \(2020\)](#) for example, as a historian distinguished between informal SNA as ‘soft SNA’ and the “formal social network analysis on the one hand, with its involvement of statistics, mathematical computations undertaken by computer algorithms” ([Rollinger, 2020](#), p. 1). And although his work mainly focuses on the interface between historical science and social science, he also raises very important questions on the topic of this paper:

“What software was used in drawing them up, and what algorithms and software functions were employed to take quantitative measures? If SNA is to be taken seriously, these questions have to be addressed and answers provided, ideally by making the corresponding data sets publicly available.” ([Rollinger, 2020](#), 26)

Therefore, [Rollinger \(2020\)](#) already points at the interdisciplinary challenges between SNA and computer science, although the discussion is currently mainly between SNA and humanities. [Dörpinghaus \(2021\)](#) proposed another approach, a fourfold characterization based on the research of [Stegbauer and Häußling \(2010, p. 21\)](#) depending on the usage of methods from Computer Science and their feedback or dialogue to the original research domain. Here, the last category include the mathematical and/or computational models which abstracts the data to create novel models or method within the field of Digital Humanities.

Thus, not only the interdisciplinary research within Digital Humanities on the border between Humanities and Computer Science forms a new field, but we can also only build upon very little work within the field of mathematical formalization and algorithmic research.

2. Background

2.1. Social network analysis

SNA aims at an analytic representation of human interactions. In the social sciences, this was a paradigm shift. A comprehensive historical analysis of network research in the field of social sciences can be found in [Scott \(2012, 7ff\)](#), [Stegbauer and Häußling \(2010\)](#) or in [Rollinger \(2014, p. 345ff\)](#). The development began in the 1940s as so-called “sociometry” and had its breakthrough in the US in the 1970s. There is not a single definition which may be used, but according to [Stegbauer and Häußling \(2010, p. 21\)](#) we see several additions to the description and analysis of social relations between different actors as part of the social order: SNA can describe the systematic evaluation of empiric data, the graphical presentation of these data and, it can describe mathematical models describing these data. Thus, it is a highly interdisciplinary field which is a part of the so-called digital humanities.

In general, the definition of social networks seems to be so trivial that apparently most scholars begin with examples of friendship or exchange networks, see for example [Jackson \(2010, pp. 4–19\)](#). We propose a very generic graph-theoretical definition.

Definition 1. (Social Network). A Social Network is a Graph $G = (V, E)$ with vertices (nodes) $v \in V$ and edges (relations) $e \in E$. Both edges and vertices are part of previously well-defined categories, $V \subseteq C_1 \cup C_2 \cup \dots \cup C_n$ and $E \subseteq R_1 \cup R_2 \cup \dots \cup R_m$. The network G may either be directed or undirected.

For example, we may have 3 categories of nodes C_i containing actors,

organizations and location: $V \subseteq C_1 \cup C_2 \cup C_3$.

Network approaches have also been used in historical studies for some decades. Here, they are often called historical network analysis (HNA). It is important to mention this research here for two reasons: It is highly diverse, combining data from different sources and the heavy need to evaluate the quality of data. [Rollinger \(2020, p. 25\)](#) points at the “the problem of incomplete sources bedevils any historical network research, regardless of which genre of sources s/he bases his research on.” Thus, every node v and edge e in G has a source or origin which we may denote with $src(v)$ or $src(e)$ with

$$src : V \cup E \rightarrow \mathbb{S}$$

Here, $\mathbb{S}$ denotes the discrete space of sources.

[Reitmayer and Marx \(2010\)](#) notes that a lot of different methods are used and there is no common formal structure. They show, that Anglo-American historians started to work with network analysis much earlier whereas in Europe this trend started much later, see [Reitmayer and Marx \(2010, p. 869\)](#). Since only selected methods of network analysis are used, a network analysis as it is conducted within the social science would not be carried out. In particular, a subset of literature uses the term “network” without using methods from SNA or HNA. For example, [van de Kamp \(2020\)](#) analysis of the network of translator of devotional literature from Dutch and English into German. Although combining biographic research with networks, no methods or computational analysis are used.

Although SNA and HNA are emerging topics, most works show that it is the described data and source problem that is the greatest hurdle. Especially with focus on the ancient Mediterranean region, many data are still not systematically recorded, as [Leidwanger et al. \(2014\)](#) argues. He claims that different approaches need to be made interoperable, because a “more explicit dialogue concerning long-term histories and the changing social motivations for, and hence scales of, connectivity from the Neolithic to the Middle Ages, would surely promote a greater degree of commensurability in the questions different specialists ask” ([Leidwanger et al., 2014, p. 7](#)). [Rollinger \(2014, p. 368\)](#) also deals extensively with method criticism, the variable range of results and examines the source problem as reason for quality problems, cf. [Ganter \(2015, p. 183\)](#). Thus, we also need to add a measure for uncertainty. For every node v and edge e in G we may denote this with $q(v)$ or $q(e)$ with

$$q : V \cup E \rightarrow (0, 1)$$

Here, 0 labels a very uncertain, 1 a very certain node or edge. This may be used to label hypothetical findings, or it may be used for experimental analysis.

To sum up, in any case in SNA or HNA sources need to be interpreted, because they do not allow formal statements about the quality of the relationships or their temporal duration. While network analysis can show additional aspects, working with sources is crucial, see [Ganter \(2015, p. 183\)](#). Thus, when trying to overcome this problem in the spirit of [Rollinger \(2014, p. 384f\)](#), to create a network with as many data as available, there is the need to understand the network topology and structure.

Our little excursus through the definitions and problems of social networks have already pointed out, that SNA uses highly complex graph structures. We will now transfer these structures to the system of Knowledge Graphs.

2.2. Knowledge graphs

The term knowledge graph (sometimes also called a *semantic network*) was – just as social networks before we presented [Definition 1](#) – not uniquely defined, see [Fensel et al. \(2020, pp. 1–10\)](#). [Ehrlinger and Wöß \(2016\)](#) compared several definitions, but the only formal definition was related to RDF graphs (see below) which does not cover labeled property graphs. [Paulheim \(2017\)](#) as another example, gives a definition

of Knowledge Graphs limited to the definition of important features. Knowledge Graphs were introduced by Google in 2012, when the Google Knowledge Graph was published on the use of semantic knowledge in web search.³ This is a representation of general knowledge in graph format. Knowledge graphs also play an important role in the Semantic Web and are also called semantic networks in this context.

Thus, a *knowledge graph* is a systematic way to connect information and data to knowledge. It is thus a crucial concept on the way to generate knowledge and wisdom, to search within data, information and knowledge. Context is the most important topic to generate knowledge or even wisdom. Thus, connecting knowledge graphs with context is a crucial feature.

Definition 2. (Knowledge Graph). We define a knowledge graph as a graph $G = (E, R)$ with entities $e \in E = E_1 \cup E_2 \cup \dots \cup E_n$ coming from formal structures E_i like ontologies and relations R .

The relations $r \in R$ can be ontology relations, thus in general we can say every ontology E_i which is part of the data model is a subgraph G_i of G indicating $O \subseteq G$. In addition, we allow inter-ontology relations between two nodes e_1, e_2 with $e_1 \in E_1, e_2 \in E_2$ and $O_1 \neq E_2$. In more general terms, we define $R = \{R_1, \dots, R_n\}$ as a list of either inter-ontology or inner-ontology relations. Both E and R are finite discrete spaces. Every entity $e \in E$ may have some additional metainformation which needs to be defined with respect to the application of the knowledge graph. For instance, there may be several node sets (some ontologies, some actors (employees, stakeholders, ...), locations, ...) $E_1, \dots, E_n$ so that $E_i \subseteq E$ and $E = \bigcup_{i=1, \dots, n} E_i$. The same holds for R when several context relations come together such as “is relative of”, “has business affiliation”, “has visited”, etc. By using formal structures within the graph, we are implicitly using the model of a labeled property graph, see [Rodriguez and Neubauer \(2012\)](#) and [Rodriguez and Neubauer \(2010\)](#). Here, nodes and edges form a heterogeneous set. Nodes and edges can be identified by using a single or multiple labels, for example using $\lambda: E \rightarrow \Sigma$. We need to mention that both concepts are equivalent, since graph databases use the concept of labeled property graphs. But it is also obvious that by identifying the categories from [definition 1](#) with the formal structures in [definition 2](#) both are equivalent.

But we need to go one step beyond that. The main issues while using classical data representations like XML or JSON is that they cannot describe relations between entities. This led to the development of a W3C standard called RDF (Resource Description Framework). It is widely used withing web resources, but also within the context of scientific data representation. The benefits for using RDF are a far more generic possibility for representing data and the capability to link them by describing relations. It was developed in the 1990s for the Semantic Web and inspired a lot of approaches like *Linked Open Data*, see <https://www.w3.org/DesignIssues/LinkedData.html> or Tim Berners-Lee's *5 Stars Open Data*. Talking about the Semantic Web is trying to describe a concept which is flexible and easy but also so complex that it was not pushed in the last decades.

The Semantic Web is intended to describe resources and their metadata and make them interoperable. The goal of Semantic Web is to make information on the World Wide Web automatically and automatically interpretable. In addition, semantic meta-information is stored for all content meta-information is stored for all content, providing information about its meaning, context, or their relationships to other content. Tim Berners-Lee summarizes the main idea:

The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. [...] The Semantic Web is not a separate Web but an extension of the

current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. [Berners-Lee et al. \(2001\)](#).

This vision is not fulfilled till now. Although there are great hurdles for this holistic view, RDF leads to a graph structure which forms a *Knowledge Graph*. Several Ontology technologies are driven by and build upon RDF, for example RDFS and OWL.

Here we see another overlap between the informal wish to connect data in social networks presented in the last section and the idea of Knowledge Graphs: While in the humanities connected data is mainly hand-crafted, the usage of RDF would lead to an automatized linking of data points.

Closely related to this is the concept of Linked Open Data. This involves data that is linked to one another by its semantic meta-information and published in a freely accessible form. This is how different data collections like PubMed have been linked together. Thus, by combining the idea of social networks and Knowledge Graphs we should also improve the linking of (open) data. But before describing our method, we need to briefly address graph databases and the complexity of storing and retrieving data in graph format.

2.3. Graph databases and complexity

Although graph databases are a relatively new field with constantly emerging technologies often missing common standards (like query languages) a lot of research has been done on the field of query optimization, query transformation and of course in storing and retrieving large scale knowledge graphs. While current state-of-the-art systems often use RDF data models which are a collection of nested graphs and SPARQL queries, the field is now driven by labeled property graphs to overcome their serious limitations. For example, nodes and edges have no internal structure which does not allow complex queries like subgraph matchings or traversals, and it is not possible to uniquely identify instances of relationships which have the same type, see [Desai et al. \(2018\)](#).

It is obvious, that graph databases show different query times on different situations and there is a considerable amount of literature on that topic. We refer to [Huang and Dong \(2013\)](#); [Desai et al. \(2018\)](#). A lot of research has also been done with respect to analyses and optimization of graph queries, especially with focus on Cypher and Neo4j. [Hölsch and Grossniklaus \(2016\)](#) for example focus on an algebraic query transformation without the usage of a relational database system to process graph data, whereas [Thakkar et al. \(2017\)](#) conclude, that there is a very confusing situation, it is “an unforeseen race of developing new task specific graph systems, query languages and data models, such as property graphs, key-value, wide column, resource description framework (RDF)”. See also the works of [Angles et al. \(2019\)](#); [Zhao and Han \(2010\)](#); [Eymer et al., 2019](#).

There seems to be no generally established procedure for categorizing graph-based queries. What we know about graph queries is largely based on different sources that categorize graph queries or describe them according to different criteria, see [Needham and Hodler \(2019\)](#); [Pokorný \(2018\)](#); [Wood \(2012\)](#). [Dörpinghaus and Stefan \(2020\)](#) provided a joint schema for this based on accessing attributes, node and edge types to be considered and entry points. We present an illustration in [Fig. 2](#).

The schema divides the categories for graph queries into *local structures* and *local & global structures*. The second structure category is called *local & global*, because some graph algorithms can act locally by specifying a start node or a subgraph. Furthermore, some categories, such as *CRPQ* or *ECRPQ*, were identified as subcategory of other categories. This is illustrated by the hierarchical structure of the categorization scheme. The category *Aggregation* belongs to graph queries that search for both local and global structures. For example, the category *Aggregation* can include questions such as “What is the degree of node

³ See <https://blog.google/products/search/introducing-knowledge-graph-things-not/>.

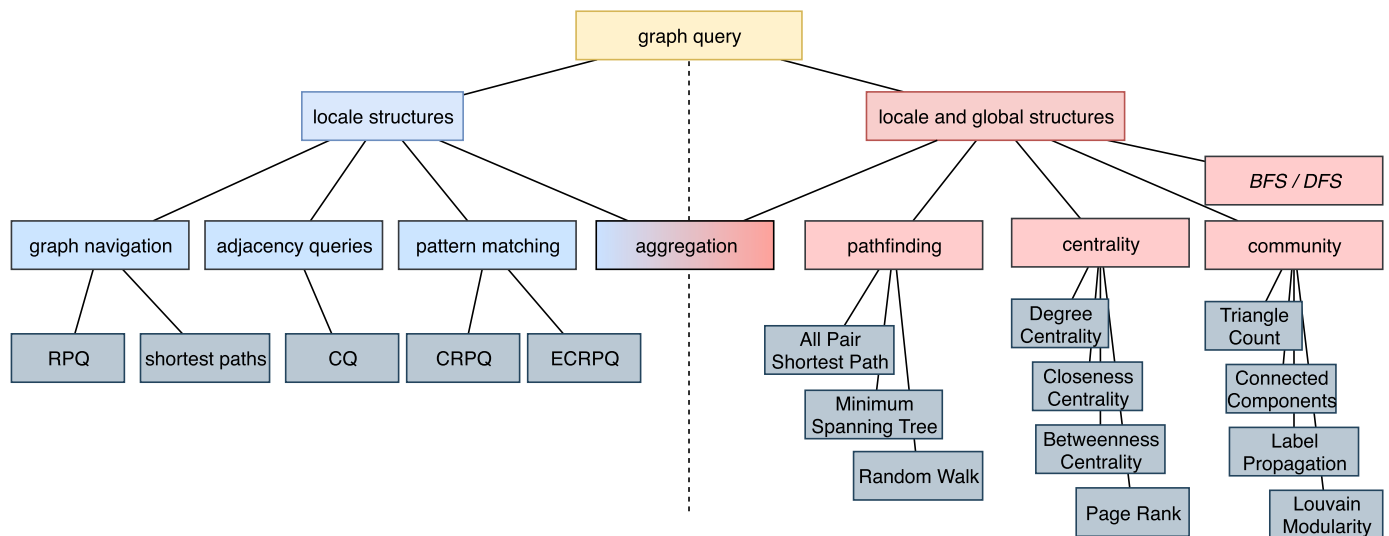


Fig. 2. An overview of the categories for graph queries unified from literature sources. These categories give a first overview and a categorization scheme for graph queries and their complexity. Picture adapted from Dörpinghaus and Stefan (2020).

A?” or “What is the average of the graph?”, the former referring to local and the latter to global structures.

While some problems are known to have solutions running in polynomial time (like pathfinding), others are well known to be quite hard. For example, graph navigation and pattern matching are more complex. RPQ is known to be PSPACE-complete, CRPQ and ECRPQ are EXPSpace-complete, see Bonifati and Dumbrava (2019).

Centrality measures for knowledge graphs are also quite complex. While several efficient algorithms have been proposed, see Grando et al. (2016), and some more specific problems are known to be NP-hard. For example, Group Closeness Maximization (GCM), see Chen et al. (2016) or Maximum Betweenness Centrality, see Fink and Spoerhase (2011).

3. Method: from social networks to knowledge graphs

Data and knowledge management, sometimes also called information management, is a core topic in computer science. Since we are using methods from this field to store and retrieve data, i.e. social networks, it is necessary to discuss some preliminary concepts of this field. In general, it is a quite interdisciplinary field touching economics (how efficient and expensive is the solution?), psychology (do people use and understand this solution in a way that was intended?), any other data-driven field, and of course informatics. This underlines the universality of this approach and after introducing the DIKW concept, we will discuss how these concepts can be utilised to generalize the concept of SNA towards knowledge graphs.

One of the core concept in data management is *DIKW*: Data, Information, Knowledge, Wisdom, see Hey (2004). First introduced by Zeleny (1987) it was extended by Ackoff (1989) who introduced the perspective of wisdom. Sometimes this hierarchy is depicted as a *knowledge pyramid*, sometimes it is a linear chain. Fig. 3 combines both perspectives: The linear perspective of understanding and context with past and future and the pyramid's perspective describing the amount of data leading to a smaller amount of information etc. In general, knowledge can be seen as *explicit* or *implicit*. Data is always given explicit. Implicit knowledge is not available for data mining, since it is only available as personal knowledge or experience. Thus, one of the key problems for digital humanities is the amount of implicit knowledge, for example historic events or not written down parts of history. In information theory, knowledge is obtained from data and information. Data are recorded, context-free facts like measured values from devices (mass spectroscopy) or basic notes (weight of patients), but also images (e.g. computer tomography). The latter in turn may be regarded as organized pixels of

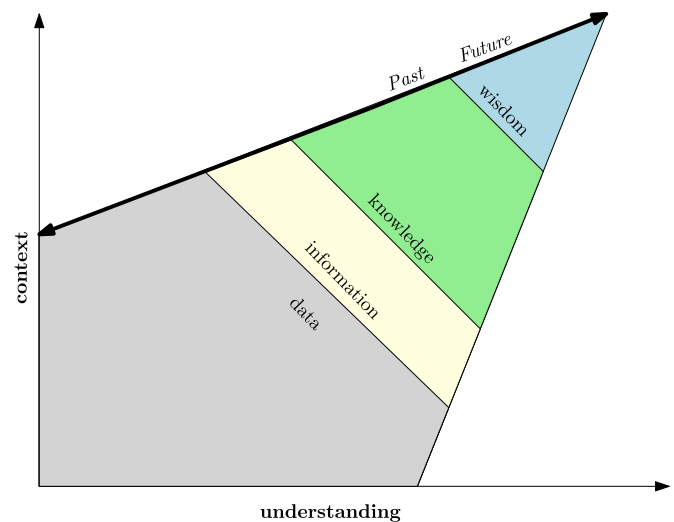


Fig. 3. DIKW hierarchy or ‘Knowledge Pyramid’ in both a linear and the pyramid perspective.

certain intensity values. In the beginning, these data are measured. If this data is enriched by context, which implies meaning and purpose, we get information. This information leads to knowledge and wisdom if – once again – enriched by context.⁴

In summary, SNA records data from social interactions and social relations into networks which display the sociological context of each person. Extending this network with more contextual information, like locations, organizations or other data points will result in more data. This will help to apply more methods from computer science. We can thus summarize the method in the following steps, which will be illustrated afterwards with several examples:

1. This step comprises a *Social Network* $G = (V, E)$ with actors and their (social) relation. This implicitly forms the first layer of a knowledge graph.

⁴ More information about this topic can be found in the work of Hey (2004) or Rowley (2007).

2. We may iteratively add several layers of other social categories, frames [?] and organizations. Each forms yet another data layer in the knowledge graph. However, till now this concept is widely used and compatible to the network paradigms in social sciences.
3. We may iteratively add several layers of other data, for example:
 - Entities from ontologies and other formal structures, for example to annotate data with concepts and attributes.
 - Empirical data evaluating actors and other concepts.
 - Sources and provenance of data points.

In this third step, we explicitly build a knowledge graph and rely on the concept of linked data. However, even the first step may be considered as a knowledge graph with a single data layer of actors.

But we may also consider the other direction: Any knowledge graph $G = (E, R)$, with entities $e \in E = E_1 \cup E_2 \cup \dots \cup E_n$, coming from formal structures E_i , where $E_1 = \{a_1, \dots, a_m\}$ is a set of m actors $a_1, \dots, a_m$, forms an *extended social network*.

All other structures $E_2, \dots, E_n$ describe data from other sources or other domains. These *data layers* play a crucial role for data integration within knowledge graphs, see Dörpinghaus and Jacobs (2020).

Knowledge graphs play in general an important role in recent knowledge mining and discovery. A *knowledge graph* is – as discussed in the previous section – a systematic way to connect information and data to knowledge on a more abstract level than language graphs. It is thus a crucial concept on the way to store and retrieve linked data, to generate knowledge and wisdom,⁵ to search within data, information and knowledge and a prerequisite for a detailed data analysis. The “context” of data is a significant topic to generate the knowledge necessary for further analysis. Simple datasets store simple data relations.

Example 3.1. We will start with a very generic example, see Fig. 4. On the left, this illustration shows several social networks. They may contain actors, organizations, institutes, clubs and other actor-based entities. They share both internal and external relations (e.g. a person might belong to a club). These actors are linked with entities and annotations from other formal structures, for example locations and attributes. They form another context layer, whereas the actors form a context for these entities. In case of literary studies it is possible to add entities for characterizations, point of view, story plots, etc. These formal structures are the link to quantitative data, for example measures (age) or even survey or clinical data.

Example 3.2. Another example was presented by Dörpinghaus and Jacobs (2020), see Fig. 5. Here, a social network of scientists was implicitly generated with data from other sources. The study used bibliographic data from DBLP and PubMed which lead to a document layer, an author layer, a journal or venue layer etc. Limited to this data, the analysis would be limited to the study of co-authorship. By adding a project layer, we get additional information about institutes and companies, affiliated actors and other publications.

It is easy to see that we can now answer more questions on different layers. For example: Which people have collaborated in projects but have not published together or have published different findings? Which people might be a good choice for collaborating on a particular topic? Here, we may also ask novel questions. For example: Can we extend the theory of forbidden triads introduced by Granovetter (1973) to other entities as well? Or in other words: If two actors a and b belong to another entity c , how likely do they know each other?

Example 3.3. A third example are web based learning platforms. It gets more and more common to use online-learning-platforms, and the amount of reliable data for its analysis is growing. For platforms using

different types of metadata and user-log data, different methods of learning analytics and knowledge graphs can be used to prove educational assumptions on online learning and learning in digital environments.

Example 3.4. We will discuss another short example of a knowledge graph extension of a social network within the context of historical or religious networks. In Fig. 6 we can see a social network representation of Lk 1:1–2:22. There are two different sets of entities. E_1 describes persons, E_2 locations. The edges describe different relations between both entities from the same set and between entities from different sets. A more generalized network or graph can be found in Fig. 7. Here, every node and edge has some sort of source or provenance. The brown node “Lk” refers to the Gospel of Luke. If we find references or sources in scientific literature, they form orange nodes. There are also some other concepts or entities that can be seen as context, for example being “barren” or being a “priest”. These concepts will help to detect previously hidden knowledge with the help of computer science.

But we should also mention, that this concept needs a careful consideration of the context that will be used to build this knowledge graph. For example, we have added nodes describing the authors of scientific literature. This can be used to detect the coverage of certain authors, or, in other words: Did we cover a sufficient number of different authors and options to construct this network? But we need to discuss, if we want to tackle this question, before building the network.

Since we are now entering the interdisciplinary field of data science, we should also mention that knowledge graphs are generated out of original data, both structured and unstructured. Being secondary data, they address central ethical standards of science: reproducibility, transparency and a fair and – if possible – open, handling of data. These standards are also key for reusable AI and can be summarized with the “FAIR Data” principle, which was introduced in Wilkinson et al. (2016). FAIR as an acronym refers to Findable, Accessible, Interoperable and Re-useable. It evolved out of the Semantic Web, representing data as Linked Open Data. A central component of FAIR Data is the semantic preparation of knowledge in a format that allows not only the search and retrieval of (meta-)data, but also interoperability and reusability.

4. Example case studies

4.1. Applications from computational social sciences: the complexity of survey structures

Lee et al. (2019) presented a study on the perception about the frequency of attributes and measures in the ego-network of actors. They claim, that the perception biases are related solely to the structure of social networks, in particular “on the level of homophily and its asymmetric nature, as well as on the size of minority group” (Lee et al., 2019, p. 1). In Fig. 8 we illustrate their approach: First, a social network model was generated where the homophily parameter (regulating the probability of connections between actors from minority and majority groups) and minority-group size were adjusted according to their theoretical model. When balancing the aggregation of perception within an ego-network, this study claims that perception biases can be reduced in heterophilic, but not in homophilic networks.

In this case, the Knowledge Graph $G = (V, E)$ comprises a social network layer $S = (V_S, E_S)$ (Step 1) and survey data $\Sigma = (V_\Sigma, E_\Sigma)$ (Step 2). Beside of that we may also add more data layers, for example explaining survey data or adding more quantitative or qualitative data to S , see Fig. 8. We may now assume that one or more layers are partly incomplete, for example that S is missing social relations. There are several questions: Can we predict relations in S ? Is it possible to apply methods on S without the missing data, for example community detection or computing centrality measures?

It is quite obvious that the model introduced by Lee et al. (2019) could be used to re-predict relations in S . But there is a serious

⁵ At least while talking about the computer science task of “Knowledge Discovery”. This is based on a formal definition of knowledge in databases combined with data processing and transforming.

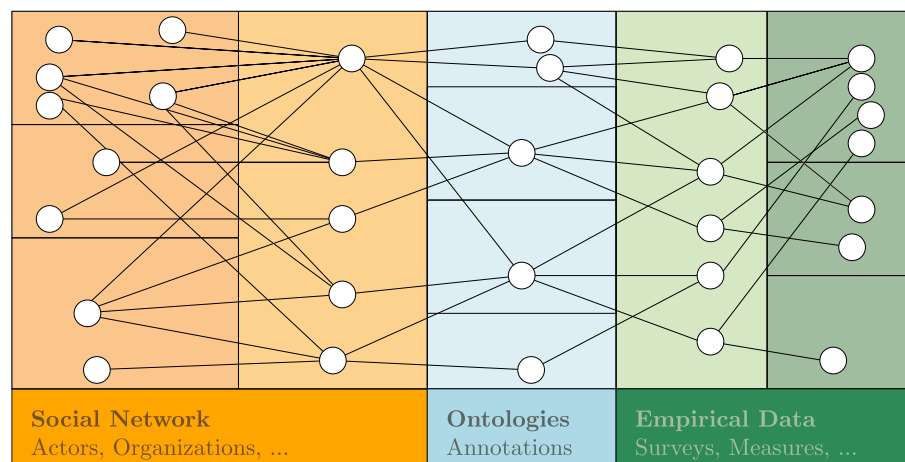


Fig. 4. This illustration shows several social networks (left). They may contain actors, organizations, institutes, clubs and other actor-based entities. These actors are linked with entities and annotations from other formal structures, for example locations, attributes or – in case of literary studies – entities of characterizations, point of view, story plots, etc. They are the link to quantitative data, for example measures (age) or even survey or clinical data.

shortcoming: The model was build on real-world scenarios to explain the perception biases within an ego-network. But unfortunately the survey data itself does only refer back to a single actor network S , not to a complete ego-network or a network. We may denote the hidden, complete social network with $\hat{S}$. Thus, this approach would only be feasible once we have survey data for such a complete social (sub-)network.

On the other hand, it is possible to apply methods like Community Detection on G , see Fig. 9. But given one particular method, the result on G , S and $\hat{S}$ may vary. Applying the method on S wouldn't be useful, since the relations (and actors not participating on the survey) are missing. But assuming the complete (sub-)network $\hat{S}$, we would expect that the results on $G|_S$ are the same or comparable to those on $\hat{S}|_S$.

In general, a *community* is a subset $C \subseteq V$ of nodes. A *community structure*⁶ is a set of communities $C_1, \dots, C_n$ so that $C_1 \cup \dots \cup C_n = V$. If these communities are pairwise disjoint as Jackson (2010, p. 444) suggests, we have hard borders between these sets. If nodes or actors may belong to multiple communities, we may speak of soft or overlapping community detection.

But beside of this definition, we lack a detailed understanding of *what* a community is. Several prominent definitions exist, and from a mathematical perspective they can't be merged. Thus, there is a little arbitrariness. Are communities "tightly knit groups with dense connections among their members" (Newman et al., 2006, p. 553) or solely unconnected margin actors? Jackson (2010, p. 449) summarizes: "As we vary what a community represents, the optimal method for identifying communities will correspondingly change. In particular, it is important to have an idea of how community structure affects network formation." Thus, we need to discuss how nodes may be "playing a similar role" (Fortunato, 2010, p. 1) and thus can be grouped together. This is still an ongoing discussion in research (see Newman et al., 2006, p. 553). Fortunato (2010) provides a detailed overview on methods and approaches, and especially on those called 'traditional methods'. Most research focuses on the performance of these algorithms on dense and large-scale networks. So in general, there is a lack of research for community research on small social networks and in particular on networks brought up from other scientific questions. Thus, the initial research questions remain open: Are results on $G|_S$ the same or comparable to those on $\hat{S}|_S$?

To sum up, combining different layers and a social network to a knowledge graph extends the possibilities to apply methods from social

network analysis. Since the survey data is usually mostly related to single actors or actor points, we can't apply all methods from SNA directly. Thus, it is not surprising, that only very little research has been done on the graph structures (cf. Cinelli et al., 2020; Rossetti et al., 2021). But some methods like community detection, clustering or grouping are still available and further research needs to be done since it remains unclear to which degree the results lead to the expected output.

4.2. Applications from computational pedagogy: learning in digital environments

Another application is the open web based learning space (OWL, <https://wb-web.de/lernen.html>) which aims to professionalize adult educators. It is developed by the German Institute for Adult Education, funded by the Federal Ministry of Education and Research, and is currently being enhanced in further projects. To support the heterogeneous target group, the platform offers learning content that addresses the needs to manage challenges in teaching and learning, but also to expand the relevant competences. Next to basic registration information, there is no data available at the moment that allows to create individual user profiles. As part of a new project, it is planned to collect data that can categorize users regarding their professional background and experience, as well as their area of expertise. This information can be used to display social connections of users.

Next to this data, the learning experience can be connected to the learners individually. Users can choose between more than 30 learning-pathways consisting of about 5–12 sequenced learning units. Each step addresses specific topics and problems, competence dimensions and includes tasks at specific competence levels (low, mid or high).

Each user learning on the OWL learning platform generates individual user data that is used for the correct display of learning units and for the system to know which content to show next. Based on this, it is possible to receive feedback depending on the result of tasks and assessments. This information is stored in a learning record store using xAPI statements. The statements consist of an individual (hashed) ID, a verb and an object. This means that every statement can show who did what on the platform. Example: User A answered Task B.

These xAPI statements shall be used to compare the learning experience of users and to identify patterns that might relate to social connections between users. A knowledge graph for instance could display a connection between the success in answering an assessment of a specific learning pathway and the area of professional expertise.

Each learning pathway and learning unit is equipped with a broad set of various metadata. This information is used for classification of

⁶ The corresponding graph-theoretical problem is called *graph partition*. Other network researches use a different nomenclature: Easley and Kleinberg (2010, p. 70) for example, name communities 'regions'.

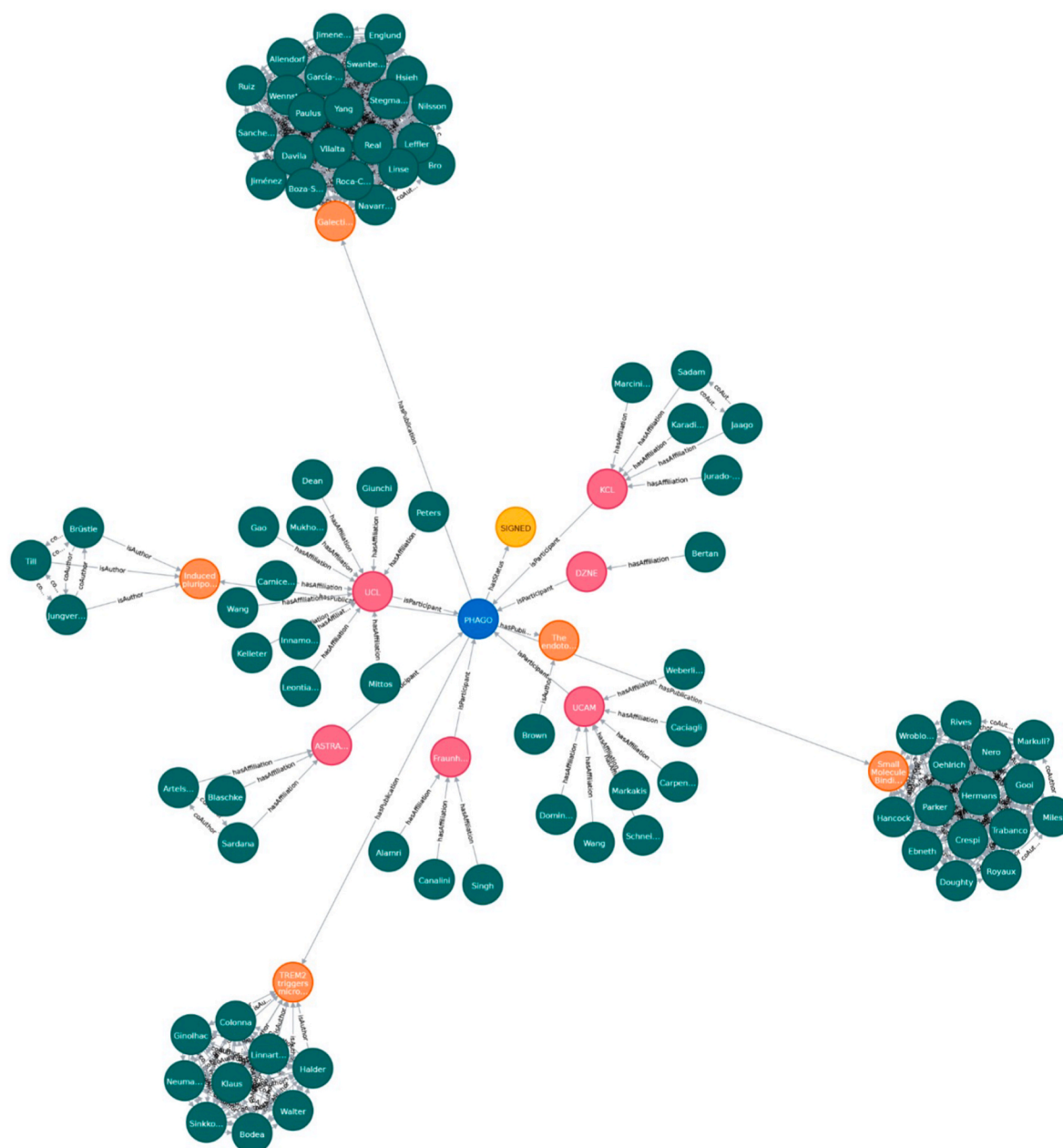


Fig. 5. A linked data knowledge graphs combining scientific projects data (H2020) with text mining on documents from other sources (DBLP or PubMed) presented in Dörpinghaus and Jacobs (2020). Here, a subnetwork around the PHAGO project is presented. Blue nodes refer to projects, red nodes to research institutes, green nodes to actors and orange nodes to documents. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

learning materials (e.g. competences), for filtering, recommendations, user information and so on. Furthermore, it is suitable for the interoperable use of material between different platforms. Applying this data to a knowledge graph would offer the opportunity to identify differentiated entry points to adjust and improve the learning material. Furthermore, it could be used to predict which user type might be interested in which topic and at which learning level.

During the development of the platform, the didactic concept and the implementation has been constantly reviewed. The aim of the formative evaluation accompanying the development was to continuously validate the acceptance and added value of the concept with the target group. The accompanying research showed, amongst other results (cf. Schöb et al., 2020), that both the interviewed teachers and the providers of train-the-trainer formats value the didactic implementation and the instructional support the platform provides for the self-directed learning

process. Difficulties arose with user guidance and usability, and users lacked just-in-time support for the design of courses (For further details we refer to Schöb et al., 2020).

It is necessary to take a deeper look at the latter points and to connect to the latest research. The evaluation options through web analytics tools offer initial insights, but many points remain open in order to explain how, where and why users interact or do not interact with the platform and the learning content.

The evaluation of the data by a knowledge graph seems more promising. Concepts such as “tailor-made content” or “successful learning” would have to be specified more precisely for such an evaluation and must be linked to a dataset (consisting of different combinations). The goal is to automate the interpretation of the data in order to close knowledge gaps. Ideally, the knowledge graph can be applied to put the content, the user and the behavior of the user in a relating

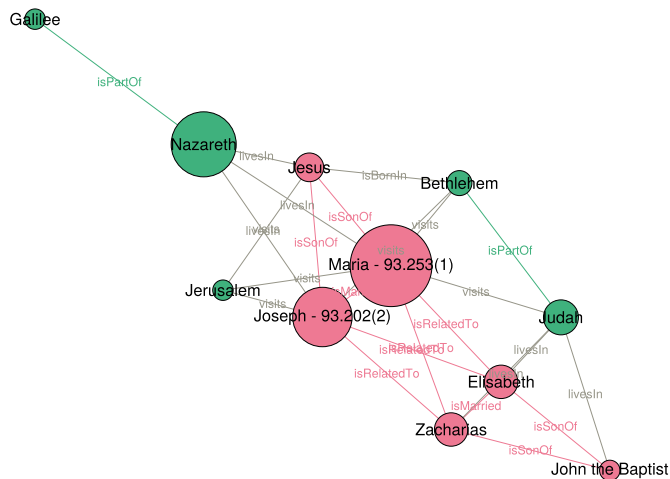


Fig. 6. The social network representation of Lk 1:1–2:22. Red nodes describe person, green nodes locations. The edges describe different relations. The node size reflects the betweenness centrality of each node. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

context. Semantic connections between different data-sets need to be trained by educational scientific knowledge. Nevertheless, the visualization of the graph could also already be used to get a better understanding of the connection of users, data and the progress of users.

For the following aspects, it is important to know the data structure of the OWL platform. There is metadata that is following the LOM (learning objects metadata) standard (cf. Kegel et al., 2002). This set of data contains information about duration, keywords, didactic aspects, authors, media and so on. Furthermore, the data from the learning record store contains information about the individual user progress. There is information available that tracks when which content has been selected, which tasks have been answered or been skipped, the score of answered tasks and the behavior for nearly all buttons (continue, repeat, skip, ...). Putting these datasets into context might serve in answering the following questions:

1. How do users interact with the learning platform? Is the user guidance adequate? Where do users bounce, or where needs the platform to be adapted? How do users find the suitable learning content?
2. Which circumstances let users bounce content? Can user data show where instruction is missing, tasks aren't appropriate or not precise enough, where users are missing content or where clarity and transparency of the learning goals are missing? Which additional data is needed for sound analytics?
3. Which content and tasks are processed by whom and in which way?

Following these questions it would be possible to create relations between chosen content, successful/non-successful answered tasks, the addressed competences, levels, kind of content presentation, task types, the information about users (e.g. experience, context of work, duration of learning experience, ...) and the recommendation system. Following this information, it might be possible to discover patterns that can affect further research, lead to better content creation and the optimization of content.

4.3. From graphs to knowledge graphs: enriching the lonsea graph with interpretations and external knowledge

In the project Lonsea at the University of Heidelberg (later continued

at the University of Basel) information about international organizations in the era of the League of Nations – roughly between the two world wars – was digitalized, analyzed and made publicly available.⁷

The main types of data collected were organizations and their members. One goal was to perform social network analysis to identify important individuals who, via their membership in various organizations, played an important role connecting the international network.

The naïve graph (Nodes = people, Edges = two people are in the same organization, Step 1) consisted of 12 468 people and 5 820 842 connections between these people. Many of the connections come in a few larger organizations. Instead of dealing directly with a graph of people and their connections, we introduced the organizations as nodes, thereby increasing the number of nodes by 1 161, but reducing the number of edges to, 21 047 (0.3% of the original graph). Of course, in the analyses concerning measures for centrality, betweenness etc., organizations were now “competing” with people and had to be treated separately.

By adding the organizations (Step 2), we not just reduced the size of the graph, we also gained the possibility to add more types of edges, namely between organizations. In the original documents, there are 22 types of relations organizations can have, for example “associated with” or “financed by” or “has institutional relationships with”. This allowed us to add 279 edges between our organization nodes. But it does not end here: Organizations have categories like “Feminism” or “Medicine and Hygiene” (20 nodes and 1 161 edges), and both people and organizations have residences in places (3 871 nodes and 1 676 + 17 023 edges), and organizations held congresses in various places (+2 598 nodes and 5 196 edges).

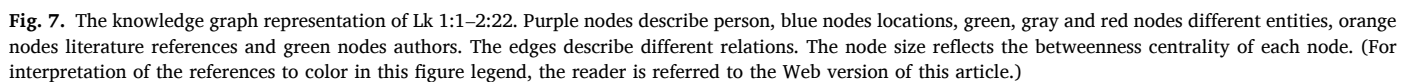
In terms of “external” knowledge, we used Google services for geocoding to get the location of all 3 086 places (which allows to identify places very close to each other, and potentially add an edge between them), and we added 3 202 links to external databases like Wikidata or the Library of Congress, which allows us to import links between people from outside sources. This iterative process is part of Step 3.

Our initial graph was very simple: the nodes were people like Sigmund Freud, C.G. Jung or Albert Einstein, with Freud and Jung being connected by both playing a role in the *International Institute of Psychagogy and Psychotherapy*, and no path from Sigmund Freud to Albert Einstein, see Fig. 10.

With our enriched graph, we have a path *Sigmund Freud* → *International Institute of Psychagogy and Psychotherapy* → (being a member of) *Federation of International Semi-Official and Private Institutions Established at Geneva* → (having as a member) *International Student Service* → *Léon Steinig* → *World Union of Jewish Students* → *Albert Einstein*. This path seems rather brittle, since people in organizations which were part of the *Federation of International Semi-Official and Private Institutions Established at Geneva* (having among its 24 members organizations like the *League of Jewish Women*, the *Bahai Cause* or the *International Commission for the teaching of Mathematics*) probably did not know each other – but when Einstein wanted to contact Freud to exchange ideas on the topic of war and violence, the communication went indeed through Léon Steinig (a fact that was unknown to the author when picking this example).

In our new graph, we not only have more categories of nodes, which should be taken into account in all kinds of social network analyses, but also more and less important edges, corresponding to the probability or intensity of personal contact, making our graph more like a flow network, with, for example, the question of reachability or path length (what is the shortest path through – however loose – connections from one node to another) being replaced by the max-flow problem (what are the weakest links you have to go through to reach one node from another). Assigning weights (or *capacities* in network terms) to the edges is however non-trivial and requires a lot of domain knowledge. Just by naming the subject, it is very much likely that people assembled in the

⁷ See Further lonsea.org.



To summarize, we started off with a simple social network graph, and introduced many new categories of nodes and edges, and queried external data sources (Google Maps and Wikidata in this case) to add even more edges, and ended up with a much richer graph, which delivers much more interesting results, but requires continuous fine-tuning and more care in the interpretation of the results.

Although SNA has developed within social sciences and is under constant discussion in other sciences like history and theology as well, see the work of Rollinger [Rollinger \(2014\)](#), or Collar [Collar \(2013\)](#). It is of some interest for theology, as Duling stated out: “interest in SNA by Biblical scholars has been sporadic, but steady, and is apparently growing.”[Duling \(2013\)](#)

10

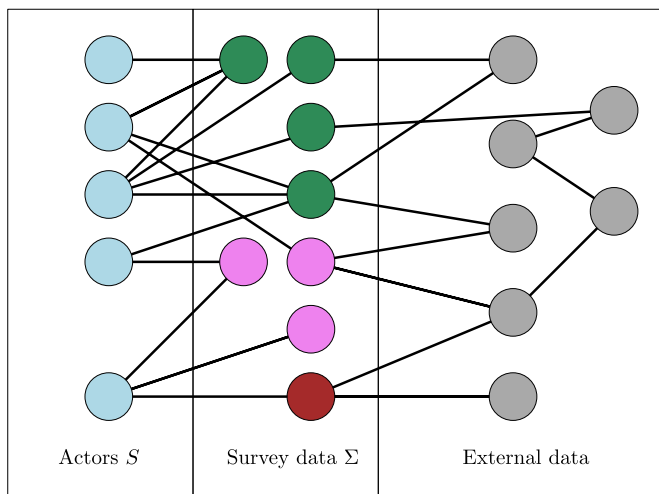


Fig. 8. The social network representation of the approach presented in Lee et al. (2019) extended with external data.

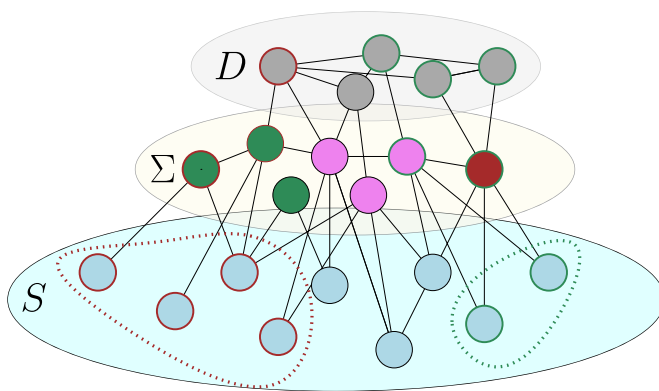


Fig. 9. The social network representation of the approach presented in Lee et al. (2019), a social network S with missing relations and survey data Σ extended with external data D . Here, we use data from other layers for community detection on S .

Social Networks – already a knowledge graph per definition, as we have shown above – can be easily extended to a “full” knowledge graph in combining literature sources and other knowledge entities. Here, we will present a short example from theology and consider yet unanswered research questions.

To visualize and analyze early Christianity as described in the New Testament, we use an example from Dörpinghaus (2020) to propose the usage of SNA in order to determine the concept of space – both social and topographical. These methods are interdisciplinary and requires an exegetical approach on Luke-Acts and a historical approach on all other historical sources available for early Christianity. For further details, we refer to Dörpinghaus (2020, 2021); McClure (2020).

In Example 3.4 we introduced a social network representation of Lk 1:1–2:22 with different sets of entities. The social relations between actors in the narrative forms a social networks as described in Step 1. Building a more complex network was proposed in Dörpinghaus (2020). Including locations and social groups is part of Step 2. Iteratively adding external literature references and evidences, and other entities was described in Step 3. Fig. 11 shows an illustration of the social network described in Acts 1–12. The node's size describes the betweenness centrality measure for each actor. Here, the algorithmic challenge is to compute distance measures. These problems are in $\mathcal{P}$. The instance graph is quite small, and thus a solution can be computed in very short time. But a question remains unclear: Are algorithms generalizable on

multi-layer knowledge graphs, and remain they robust for additional data stored? Dörpinghaus et al. (2022) investigated the questions how multiple layers within graphs impact methods and algorithms developed for single-purpose networks. Their research suggests, that although some graphs are robust for additional layers, in general more research within graph theory is necessary to generalize this assumption and more research within social sciences is needed to define the expected behavior.

However, de Valeriola (2021) asks another question important for research on historical data: Are centrality measures robust enough on networks that may have quality issues due to their sources? He sets up two scenarios dealing with incomplete data and inaccuracies, and four experiments which lead to a transfer of the initial research gap to the world of graph theory. This and his experimental results clearly underline the need of interdisciplinary research in this field. De Valeriola summarizes:

Our results show that, from a global point of view, centrality is a sufficiently stable quantity to be used in such a context. However, we have also shown that the hazards experienced by the sources have impacts that differ in magnitude depending on the network studied, the comparison statistic used, and the centrality metric considered. This finding calls for caution in the choice of network tools applied, and for nuance when interpreting the results of an analysis of historical networks using centrality. (de Valeriola, 2021, 120)

This shows another gap in research: Social networks are providing statistical results on a discrete data set, while the nature of Humanities are different. Can these ‘nuances’ be expressed in mathematical equations, and can we formulate ‘hazards experienced by the sources’ in probabilities, uncertainties or even logical expressions (if ... then ... else) in a network?

Another yet unanswered question is the complexity of these algorithms on knowledge graph, see Dörpinghaus and Stefan (2022). Here, the algorithmic challenge is to compute distance measures, e.g. betweenness centrality, for each actor. But querying the graph may also be complex. In general, we consider these problems to be in $\mathcal{P}$. The instance graph is quite small, and thus a solution may be computed in very short time. Nevertheless, while some problems are known to have solutions running in polynomial time, others are well known to be quite hard. See the section on graph databases and complexity, 2.3. In particular centrality measures for knowledge graphs are known to be $\mathcal{NP}$ -hard, for example Group Closeness Maximization (GCM), see Chen et al. (2016) or Maximum Betweenness Centrality, see Fink and Spoerhase (2011).

5. Discussion

In this paper, we introduced a novel formulation of social networks as knowledge graphs to formalize multi-layer social networks within the language of computer science. These contextual information is a crucial factor in interdisciplinary research on social networks. They form a linked dataset which comprises more data from different sources. The main conclusion that can be drawn is that interdisciplinary approaches will both help to model large scale social networks with additional information layers, and tackle the challenges to analyze them. To our knowledge, this is the first systematic approach combining these different perspectives.

However, we already discussed that there is no precise border between different approaches towards SNA. For example, the difference between the analysis of (online) social networks using web mining technologies and the social network analysis as method within (computational) social sciences and the humanities is not totally clear. A computer scientist might be surprised about the need to formalize the interdisciplinary work on knowledge graphs because he already used these approaches. However, this does not hold for the humanities and social sciences. Although the concept of knowledge graphs was roughly

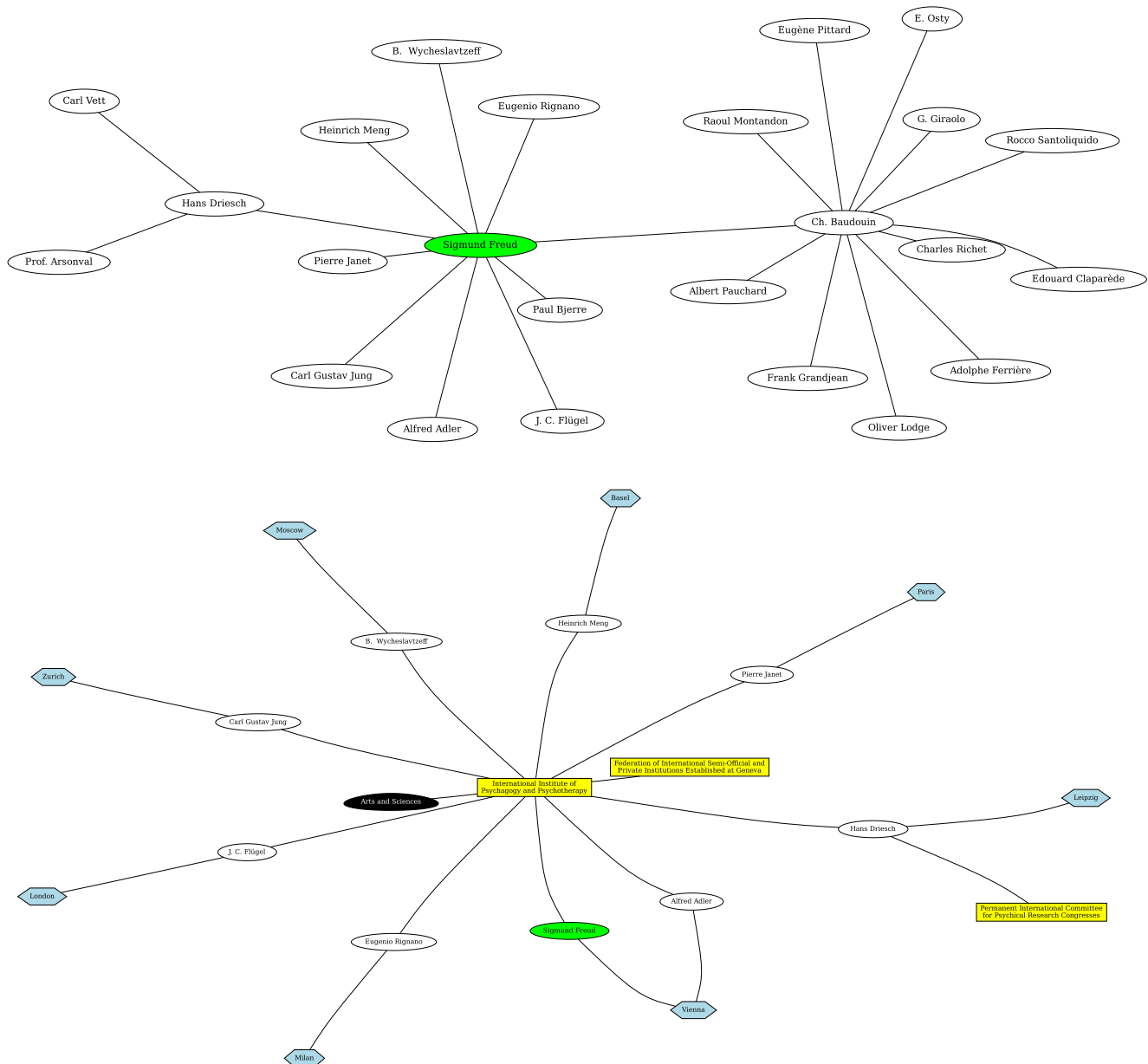


Fig. 10. Selected Nodes and Edges near Sigmund Freud (green node), (a) in the naïve social network graph, (b) with organizations, cities, categories. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

introduced? it is not yet embedded as a methodological element. Having this said, while some methods which we cover in this paper have already been applied, as our examples show, we did provide scientific arguments why and how these interdisciplinary approaches work in all domains.

Till now, this holistic perspective is not considered very often in research. Thus, missing linked data and missing methods for large-scale knowledge networks are a serious limitation. Only few researches have addressed the questions between data science, graph theory and social network analysis. This paper discussed how this gap can be closed. We investigate how new methods from computer science, in particular knowledge graphs, can be used within the field of SNA. The contributions of this paper are a knowledge graph model representation of a multi-layer social network, the showcase of examples from different domains and a perspective for more interdisciplinary discussion. Some of these examples were closer to the 'classical' approaches in SNA or HNA, some of them with missing data layers, others were more abstract

representations of social interactions, for example within digital environments. We presented knowledge graphs as an innovative solution for coming challenges in SNA with methods from the digital humanities, in particular computer science. As we have argued elsewhere, a collaboration between graph theory and social network analysis may be considered a promising aspect of next generation network science in the humanities.

Having this said, this novel approach did not only consider the application of knowledge graphs for social network analysis, which was widely used within the analysis of online social networks and platforms, but also the concept of knowledge graphs within the network paradigm within the humanities and social sciences. This provides a scientific foundation for the application of algorithms and AI approaches, not only on the data, but within the field of digital humanities and computational social sciences.

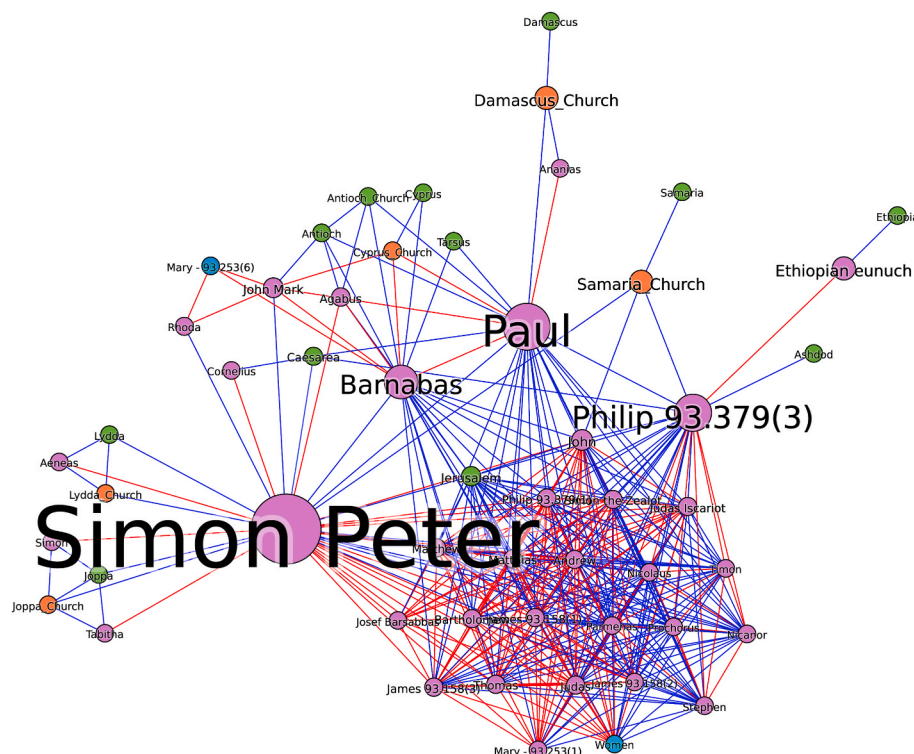


Fig. 11. The social network of Acts 1–12 (cf. Dörpinghaus, 2020). Weak ties are represented by blue edges, strong ones by red. The size of a node indicates its betweenness centrality. The node's color reflects their group affiliation. Purple denotes persons, green denotes places, Orange communities, Blue groups of people. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

6. Conclusions and outlook

We introduced a novel formulation of social networks as knowledge graphs to formalize multi-layer social networks within the language of computer science. The main conclusion that can be drawn is that interdisciplinary approaches will both help to model large scale social networks with additional information layers, and tackle the challenges to analyze them. To our knowledge, this is the first systematic approach combining these different perspectives.

Future research should consider the potential effects of multi-layer knowledge graphs more carefully, for example to understand their impact on centrality measures, community detection and other algorithms. This should also include a perspective from the humanities and social sciences: Are these effects casting a new light on the data, or are they artifacts? Besides, countless inquiries on real-world graph structures and their impacts on discrete structures exist. Thus, a close collaboration between graph theory and social sciences would help to improve our understanding of network structures and their interplay with other data and the flow of information. We believe that apart from looking for network structures, future research should also look for a deeper understanding of different, both quantitative and qualitative, data and their influence on social relations and interactions within the humanities. For example, the extension of survey data with other data sources would present a yet undiscovered network of data to explore the social structures hidden in the social layer.

CRedit authorship contribution statement

Jens Dörpinghaus: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sonja Klante:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources,

Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Martin Christian:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Christof Meigen:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Carsten Düing:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

References

- Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3–9.
- Angles, R., Thakkar, H., & Tomaszuk, D. (2019). Rdf and property graphs interoperability: Status and issues. In *Proceedings of the 13th Alberto mendelzon international workshop on foundations of data management*. Paraguay: Asunción.
- Berhan, L. M., Adams, A. L., McKether, W. L., & Kumar, R. (2019). Board 14: Social networks analysis of african american engineering students at a pwi and an hbcu—a comparative study. In *2019 ASEE annual conference & exposition*.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284, 34–43.
- Bonifati, A., & Dumbra, S. (2019). Graph queries: From theory to practice. *ACM SIGMOD Record*, 47, 5–16.
- Chen, C., Wang, W., & Wang, X. (2016). Efficient maximum closeness centrality group identification. In *Australasian database conference* (pp. 43–55). Springer.
- Choudhary, N., Rao, N., Katariya, S., Subbian, K., & Reddy, C. K. (2021). Self-supervised hyperboloid representations from logical queries over knowledge graphs. In *Proceedings of the web conference 2021* (pp. 1373–1384).
- Cinelli, M., Peel, L., Iovanella, A., & Delvenne, J. C. (2020). Network constraints on the mixing patterns of binary node metadata. *Physical Review E*, 102, Article 062310.
- Collar, A. (2013). *Religious networks in the roman Empire*. Cambridge University Press.
- Collarana, D., Galkin, M., Lange, C., Scerri, S., Auer, S., & Vidal, M. E. (2018). Synthesizing knowledge graphs from web sources with the minte framework. In *International semantic web conference* (pp. 359–375). Springer.

- Cristofaro, S., Sanfilippo, E. M., Sichera, P., & Spampinato, D. (2021). *Towards the representation of claims in ontologies for the digital humanities*.
- Czachesz, I. (2011). Women, charity and mobility in early christianity: Weak links and the historical transformation of religions. In I. Czachesz, & T. Biró (Eds.), *Changing minds. Religion and cognition through the ages* (pp. 129–154). Leuven: Peeters.
- Desai, M., G Mehta, R., & P Rana, D. (2018). Issues and challenges in big graph modelling for smart city: An extensive survey. *International Journal of Computational Intelligence & IoT*, 1.
- Dörpinghaus, J. (2020). Soziale Netzwerke im frühen Christentum nach der Darstellung in App 1-12. Available at: <http://uir.unisa.ac.za/handle/10500/26609>.
- Dörpinghaus, J. (2021). Die soziale Netzwerkanalyse: Neue Perspektiven für die Auslegung biblischer Texte? *Biblich erneuerte Theologie*, 5.
- Dörpinghaus, J., & Jacobs, M. (2020). Knowledge detection and discovery using semantic graph embeddings on large knowledge graphs generated on text mining results. In *15th conference on computer science and information systems* (pp. 169–178). FedCSIS. <https://doi.org/10.15439/2020F36>.
- Dörpinghaus, J., & Stefan, A. (2020). Optimization of retrieval algorithms on large scale knowledge graphs. In *2020 15th conference on computer science and information systems (FedCSIS)* (pp. 227–236). IEEE.
- Dörpinghaus, J., & Stefan, A. (2022). Semantic graph queries on linked data in knowledge graphs. In S. Fidanova (Ed.), *Recent advances in computational optimization: Results of the workshop on computational optimization WCO 2020*. Cham: Springer International Publishing.
- Dörpinghaus, J., Weil, V., Düing, C., & Sommer, M. W. (2022). *Centrality measures in multi-layer knowledge graphs*. arXiv preprint arXiv:2203.09219.
- Duling, D. C. (2013). Paul's aegean network: The strength of strong ties. *Biblical Theology Bulletin*, 43, 135–154.
- Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge: University Press.
- Ehrlinger, L., & Wöb, W. (2016). Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCES)*. <http://ceur-ws.org/Vol-1695/>.
- Eymer, J., Dexter, P., & Liu, Y. D. (2019). Toward lazy evaluation in a graph database. *SPLASH*.
- Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., & Wahler, A. (2020). *Introduction: What is a knowledge graph?* Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-37439-6_1. URL: doi:10.1007/978-3-030-37439-6_1.
- Fink, M., & Spoerhase, J. (2011). Maximum betweenness centrality: Approximability and tractable cases. In *International workshop on algorithms and computation* (pp. 9–20). Springer.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486, 75–174.
- Gabrilovich, E., & Usunier, N. (2016). Constructing and mining web-scale knowledge graphs. In *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval* (pp. 1195–1197).
- Gaeta, A., Loia, V., & Orciuoli, F. (2021). A method based on graph theory and three way decisions to evaluate critical regions in epidemic diffusion. *Applied Intelligence*, 51, 2939–2955.
- Ganter, A. (2015). Rezension: C. Rollinger: Amicitia sanctissima colenda. Freundschaft und soziale Netzwerke in der späten Republik. (Studien zur Alten Geschichte, Bd. 19). *Historische Zeitschrift*, 301, 182–184.
- Grando, F., Noble, D., & Lamb, L. C. (2016). An analysis of centrality measures for complex and social networks. In *2016 IEEE global communications conference (GLOBECOM)* (pp. 1–6). IEEE.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Haslhofer, B., Isaac, A., & Simon, R. (2018). *Knowledge graphs in the libraries and digital humanities domain*. arXiv preprint arXiv:1803.03198.
- Hey, J. (2004). The data, information, knowledge, wisdom chain: The metaphorical link. *Intergovernmental Oceanographic Commission*, 26, 1–18.
- Hölsch, J., & Grossniklaus, M. (2016). An algebra and equivalences to transform graph patterns in neo4j. In *EDBT/ICDT 2016 workshops: EDBT workshop on querying graph structured data (GraphQ)*.
- Huang, H., & Dong, Z. (2013). Research on architecture and query performance based on distributed graph database neo4j. In *2013 3rd international conference on consumer electronics, communications and networks* (pp. 533–536). IEEE.
- Hyvönen, E., Rantala, H., et al. (2019). Knowledge-based relation discovery in cultural heritage knowledge graphs. In *Digital humanities in nordic countries proceedings of the digital humanities in the nordic countries 4th conference*. CEUR-WS. org.
- Jackson, M. O. (2010). *Social and economic networks*. Princeton: Princeton University Press. <https://doi.org/10.1515/9781400833993>
- de Juan, S., Ospina-Álvarez, A., Villasanté, S., & Ruiz-Frau, A. (2021). A graph theory approach to assess nature's contribution to people at a global scale. *Scientific Reports*, 11, 1–13.
- van de Kamp, J. (2020). *Übersetzungen von Erbauungsliteratur und die Rolle von Netzwerken am Ende des 17. Jahrhunderts. Beiträge zur historischen Theologie*. Tübingen: Mohr Siebeck.
- Kegel, I., Lange, R. T., Manion, J., Mason, J., Matlack, E. R., Matlack, R., Mazow, C., McDonald, W. A., Melton, B., Muramatsu, B., Riley, K., Robson, R., Saunders, R., Schoening, J. R., Sinita, K. M., Stroup, G., Suthers, D. D., Tonkel, J. R., Towle, B., ... White, S. (2002). Draft standard for learning object metadata. URL: <https://www.ieeetsc.org/working-groups/wg12LDM/omDescription/>.
- Keshavarzi, A. (2021). *Towards meta-data discovery and knowledge discovery on knowledge graphs*. Ph.D. thesis. University of Georgia.
- Lee, E., Karimi, F., Wagner, C., Jo, H. H., Strohmaier, M., & Galesic, M. (2019). Homophily and minority-group size explain perception biases in social networks. *Nature Human Behaviour*, 3, 1078–1087.
- Leidwanger, J., Knappett, C., Arnaud, P., Arthur, P., Blake, E., Broodbank, C., Brughmans, T., Evans, T., Graham, S., Greene, E. S., et al. (2014). A manifesto for the study of ancient mediterranean maritime networks. *Antiquity*, 88.
- Li, N., Huang, Q., Ge, X., He, M., Cui, S., Huang, P., Li, S., & Fung, S. F. (2021). A review of the research progress of social network structure. *Complexity*, 2021.
- McClure, J. M. (2020). Jesus's social network and the four gospels: Exploring the relational dynamics of the gospels using social network analysis. *Biblical Theology Bulletin*, 50, 35–53.
- Needham, M., & Hodler, A. E. (2019). *Graph algorithms*. O'Reilly Media, Inc.
- Newman, M. E., Barabási, A. L. E., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton: University press.
- Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8, 489–508.
- Pokorný, J. (2018). Functional querying in graph databases. *Vietnam Journal of Computer Science*, 5, 95–105. <https://doi.org/10.1007/s40595-017-0104-6>. URL: doi:10.1007/s40595-017-0104-6.
- Ponten, S., Bartolomei, F., & Stam, C. (2007). Small-world networks and epilepsy: Graph theoretical analysis of intracerebrally recorded mesial temporal lobe seizures. *Clinical Neurophysiology*, 118, 918–927.
- Reitmayer, M., & Marx, C. (2010). Netzwerkansätze in der Geschichtswissenschaft. In C. Stegbauer, & R. Häußling (Eds.), *Handbuch netzwerkforschung* (pp. 869–880). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rodriguez, M. A., & Neubauer, P. (2010). Constructions from dots and lines. *Bulletin of the American Society for Information Science and Technology*, 36, 35–41.
- Rodriguez, M. A., & Neubauer, P. (2012). The graph traversal pattern. In *Graph data management: Techniques and applications* (pp. 29–46). IGI Global.
- Rollinger, C. (2014). *Amicitia sanctissima colenda. Freundschaft und soziale Netzwerke in der Späten Republik*.
- Rollinger, C. (2020). Prolegomena. problems and perspectives of historical network research and ancient history. *Journal of Historical Network Research*, 4, 1–35.
- Rossetti, G., Citraro, S., & Milli, L. (2021). Conformity: A path-aware homophily measure for node-attributed networks. *IEEE Intelligent Systems*, 36, 25–34.
- Rowley, J. (2007). The wisdom hierarchy: Representations of the dijkstra hierarchy. *Journal of Information Science*, 33, 163–180.
- Schöb, S., Klante, S., & Lehmann, A. (2020). *Kompetenzentwicklung lehrender digitalisieren: Chancen und herausforderungen für die erwachsenen-/weiterbildungspraxis* (Vol. 1993, p. 28). Der pädagogische Blick-Weinheim: Beltz Juventa.
- Scott, J. (2012). *Social network analysis*. SAGE Publications.
- Stegbauer, C., & Häußling, R. (2010). *Handbuch netzwerkforschung*. VS Verlag für Sozialwissenschaften.
- Suárez, D., Díaz-Puente, J. M., & Bettoni, M. (2021). Risks identification and management related to rural innovation projects through social networks analysis: A case study in Spain. *Land*, 10, 613.
- Thakkar, H., Punjani, D., Auer, S., & Vidal, M. E. (2017). Towards an integrated graph algebra for graph pattern matching with gremlin. In *International conference on database and expert systems applications* (pp. 81–91). Springer.
- de Valeriola, S. (2021). Can historians trust centrality? *Journal of Historical Network Research*, 6.
- Van Nuffelen, P. (2015). Religious networks. *Classical Review*, 65, 224–226.
- White, L. (1992). *Semeia 56: Social networks in the early christian environment*. Atlanta: Society of Biblical Literature.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3.
- Wood, P. T. (2012). Query languages for graph databases. *SIGMOD Rec*, 41, 50–60. <https://doi.org/10.1145/2206869.2206879>. URL: <http://doi.acm.org/10.1145/2206869.2206879>.
- Zeleny, M. (1987). Management support systems: Towards integrated knowledge management. *Human Systems Management*, 7, 59–70.
- Zhao, P., & Han, J. (2010). On graph query optimization in large networks. *Proceedings of the VLDB Endowment*, 3, 340–351.
- Zheng, Q., & Skillicorn, D. (2017). *Social networks with rich edge semantics*. Taylor & Francis.