



# A systematic mapping study of developer social network research

Steffen Herbold\*, Aynur Amirfallah, Fabian Trautsch, Jens Grabowski

Institute of Computer Science, University of Goettingen, Germany

## ARTICLE INFO

### Article history:

Received 8 August 2019

Received in revised form 9 March 2020

Accepted 19 August 2020

Available online 2 September 2020

### Keywords:

Developer social networks

Mapping study

Literature survey

## ABSTRACT

Developer social networks (DSNs) are a tool for the analysis of community structures and collaborations between developers in software projects and software ecosystems. Within this paper, we present the results of a systematic mapping study on the use of DSNs in software engineering research. We identified 255 primary studies on DSNs. We mapped the primary studies to research directions, collected information about the data sources and the size of the studies, and conducted a bibliometric assessment. We found that nearly half of the research investigates the structure of developer communities. Other frequent topics are prediction systems build using DSNs, collaboration behavior between developers, and the roles of developers. Moreover, we determined that many publications use a small sample size regarding the number of projects, which could be problematic for the external validity of the research. Our study uncovered several open issues in the state of the art, e.g., studying inter-company collaborations, using multiple information sources for DSN research, as well as general lack of reporting guidelines or replication studies.

© 2020 Elsevier Inc. All rights reserved.

## 1. Introduction

Social structures within software development projects are a topic that received a lot of attention in different research communities, e.g., by researchers interested in open source development, global software engineering, and mining software repositories. DSNs are often inferred automatically from information that can be found in forges like GitHub, Mailing Lists (MLs), Issue Tracking Systems (ITs), and Version Control Systems (VCSs) of software development projects. The DSNs give valuable insights into the projects, e.g., regarding the importance of individuals (Joblin et al., 2015), patterns in communication behavior (Damian et al., 2007b), for the identification of single points of failure (Tamburri et al., 2019), gender-aspects (Catolino et al., 2019), and even bugs (Pinzger et al., 2008). Due to the magnitude of publications on DSNs, the diversity of topics addressed by DSNs, and the lack of a contemporary literature review, a novel literature study is required to ensure that researchers and practitioners can get a complete overview on the state of the art of DSNs. This article describes a mapping study performed based on the rigorous guidelines by Kitchenham and Charters (2007) for literature reviews with the goal to identify and map research on DSNs. We map the publications on DSNs to research topics and

analyze the scope of the publications in terms of data sources, number of projects, and number of people.

With our mapping study, we provide the following contributions.

- A contemporary overview of the state of the art of the literature on DSNs.
- A summary of the already investigated research directions, including the relevant literature.
- A summary of the data sources, as well as the size of the DSNs in terms of number of projects and people involved.
- A bibliometric assessment to identify influential publications, authors, venues, and interest in the topic over time.
- The identification of open issues within the current state of the art.

We found that 49% of all publications on DSNs analyze the structure of the community, either in general, or with respect to other aspects of software development, e.g., the evolution, or the impact on code quality. Other frequent topics in research are prediction systems based on DSNs, e.g., for defect prediction or bug triage, the collaboration behavior between developers, and the roles of developers. Regarding the way that studies are conducted, we found that 79% of the studies are based on a single data source and 70% of the studies use less than 11 projects to draw conclusions. These are concerning findings regarding the generalizability of results. Regardless, 80% of publications use social networks with at least 100 people modeled by the network, i.e., large networks are usually the foundation for analysis, which is good for the generalizability. Thus, we believe there is a need

\* Corresponding author.

E-mail addresses: [herbold@cs.uni-goettingen.de](mailto:herbold@cs.uni-goettingen.de) (S. Herbold), [aynur.amirfallah@stud.uni-goettingen.de](mailto:aynur.amirfallah@stud.uni-goettingen.de) (A. Amirfallah), [trautsch@cs.uni-goettingen.de](mailto:trautsch@cs.uni-goettingen.de) (F. Trautsch), [grabowski@cs.uni-goettingen.de](mailto:grabowski@cs.uni-goettingen.de) (J. Grabowski).

for studies with high external validity on DSNs, especially more studies that consider a large amount of different projects in order to derive generalizable conclusions for diverse populations. Other open issues in the state of the art are, e.g., inter-company collaborations and the use of data from multiple information sources for the analysis of DSNs. Finally, the extraction of data from the publications for this mapping study revealed a lack of reporting guidelines for DSNs, i.e., some publications fail to report basic meta data about the studies conducted, e.g., the number of projects considered, the number of developers involved, or how data was processed, e.g., to deal with duplicate identities.

The remainder of this paper is organized as follows. We give a definition of DSNs in Section 2. In Section 3, we present our methodology for the mapping study, including our research questions, inclusion and exclusion criteria for the literature, how we identified publications, and the data we collected for each included publication. In Section 4, we give the results of our review, by listing the primary studies we found and map them to DSN concepts according to our research questions. In Section 5, we discuss open issues regarding DSN research based on the results of our mapping study. Then, we discuss related prior literature studies in Section 6, and conclude the article in Section 7.

## 2. Definition of Developer Social Networks (DSNs)

A definition is difficult, because different data sources, research goals, and modeling approaches are used to represent DSNs in the literature. Due to this, publications on DSNs contain the specific definition of their DSN structure, but this varies between publications. For our purpose, we require a definition, that can be applied to validate if a construct is an instance of a DSN. We identified three necessary and sufficient conditions for DSNs.

1. A DSN is described by a graph  $G = (V, E)$  where  $V$  denotes a set of vertices and  $E$  a set of edges such that  $E \subseteq V \times V$ . The graph can be directed or undirected, depending on the intent of the researchers and the data that is used for modeling the DSN.
2. The vertices or a subset of the vertices must represent actors of a software development process, e.g., developers, users, or project managers.
3. The edges represent connections between vertices that are based on communication behavior (e.g., email communication) or collaboration behavior (e.g., contributions to the same software artifact).

An example of a DSN is given in Fig. 1. This figure depicts an anonymized excerpt of the DSN created by Bird et al. (2006a). The vertices in this graph represent different developers, which were active on Apache email lists. A directed edge between two vertices exists, if the developer has sent or replied to at least 150 emails of another developer.

## 3. Methodology

Our review follows the guidelines for systematic literature reviews proposed by Kitchenham and Charters (2007). Additionally, we used backward and forward snowballing, which was suggested for systematic literature studies by Wohlin (2014). In the following, we define our underlying research questions, inclusion and exclusion criteria, how we identified papers, and which data was collected for our study. We do not define our study as systematic literature review but as a systematic mapping study, because we did not perform any synthesis of the results, but only provide an overview of the literature.

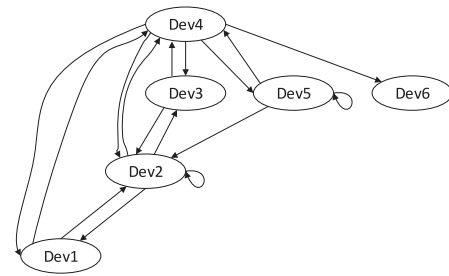


Fig. 1. Example of a DSN. It shows an anonymized excerpt of the DSN created by Bird et al. (2006a).

### 3.1. Research questions

In order to study the state of the art in DSNs, we defined the following five research questions to guide our mapping study.

- **RQ1.** What software engineering topics have been addressed by DSNs?
- **RQ2.** Which data sources are used for modeling of DSNs?
- **RQ3.** What is the scope of the analysis...
  - (a) with respect to number of projects considered
  - (b) and people modeled by the DSNs?
- **RQ4.** What are the most influential...
  - (a) publications?
  - (b) authors?
  - (c) venues?
- **RQ5.** How did the interest in DSN research evolve over time?

The first three research questions guide our analysis of the state of the art of DSNs. We want to get insights into both the topics that are under investigation within the research community, as well as the amount of studies on different topics through our analysis for RQ1. The research questions RQ2 and RQ3 guide our investigation of the scope of studies. Through the answer to RQ2, we want to get valuable information about the data sources that researchers use to define social relationships. Through RQ3, we want to gain insights into how large the studies are, e.g., if they are case studies of specific cases with few projects or if they are broad studies over hundreds of projects. The fourth and fifth question give us insights into the community of DSN research itself. RQ4 will tell us which work had the most impact, i.e., early foundational work and later work that presented new ideas for the use of DSNs that influenced many other publications. Moreover, we assess if there are authors who are clearly distinguished in the field of DSN research through their publications. We also look at the venues where DSN research is most often published to gain insights into which communities frequently use DSNs in their research. Through RQ5 we want to understand how the interest in DSN research evolves over time, e.g., if the interest is still growing or if the topics of interest change over time.

### 3.2. Inclusion and exclusion criteria

To identify which papers should be part of our review, we defined the following criteria for inclusion:

- publications that describe DSNs;
- publications that describe how DSNs may be created; and
- publications that describe theoretical aspects of DSNs.

Additionally, we used the following exclusion criteria:

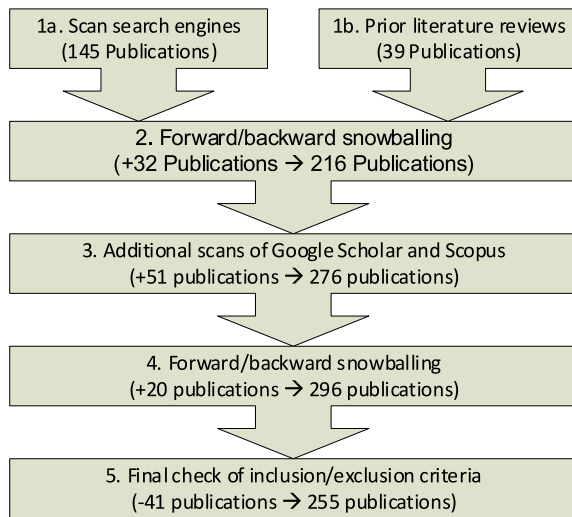


Fig. 2. Overview of the mapping study's workflow.

- publications that only summarize existing work without new contributions;
- publications that only consider social networks or graph structures in general, without a direct and specific relation to software development;
- publications that were not peer-reviewed; and
- publications that are not published in English.

### 3.3. Identification of primary studies

Fig. 2 summarizes our workflow for the identification of primary studies. We used a five step procedure.

1. Initial scan of the literature using search engines and prior literature studies to identify a seed of publications.
2. Backward and forward snowballing of publications found in the initial scan.
3. Second scan of the literature using search engines to capture the remainder of 2017 and to account for delayed indexing of publications.
4. Backward and forward snowballing of publications found in the second scan.
5. Final check of inclusion and exclusion criteria on all identified publications.

In the first step, we searched for publications by using five search engines: Google Scholar, IEEE Xplore, ACM Digital Library, Springer Link, Elsevier Search, and Scopus.<sup>1</sup> We used three queries for each search engine: “developer social networks”, “developer network”, and “collaborative networks OSS”. Table 1 gives an overview on the number of hits we had with our search terms in each of the search engines. This initial search was conducted between May 2017 and September 2017. Due to the extremely high number of hits, we considered only 750 hits per search engine and search term to get the literature seed for our mapping study. Next, we selected candidates for inclusion by reading the titles, abstracts, and, if it was necessary, the introduction and conclusion sections of the publications. We identified 145 publications through this procedure from the search engines. Additionally, we scanned the primary studies from prior related literature studies by Zhang et al. (2014a), Tamburri et al. (2013),

Manteli et al. (2012), and Abufouda and Abukwaik (2017) (see Section 6). We identified 39 additional publications from the prior studies. This difference is mainly due to the scope of the other literature studies, especially with respect to search terms. For example, Manteli et al. (2012) focus on global software engineering and, therefore, also use search terms that do not mention DSNs. Thus, we identified 184 publications in this first step.

In the second step, we checked the related work cited in each of the publications we found using the search engines. This step is also known as backward snowballing (Wohlin, 2014). Moreover, we used the “cited by” function of Google Scholar, to identify publications that cited the publications we identified with the search engines. This step is also known as forward snowballing (Wohlin, 2014). We also applied the snowballing to each additional publication we found. We identified 32 additional publications, i.e., 216 publications in total. The snowballing also served to mitigate potential negative effects because we did not consider every hit for the search terms with the search engines. Our assumption is that we find the literature we may have missed through the snowballing. Moreover, same as the use of the prior literature reviews as seed for the snowballing, the snowballing allowed us to identify literature that did not mention the DSN in the paper title or abstract and was, therefore, missed by our search.

In the third step, we repeated our search for literature from the first step. This was required, because the initial search already started in May 2017, i.e., we could not be confident that all papers from 2016 were indexed by the search engines and part of the data for 2017 was not available yet. Moreover, we wanted to include recent publications, that would be missing otherwise. Thus, we repeated the search engines Google Scholar in July 2018 and July 2019 and with SCOPUS in February 2020. This way, we identified 31 new publications using Google Scholar and 29 publications using SCOPUS, bringing our total number of publications to 276. Afterwards, in the fourth step, we performed an additional round of snowballing on these publications and identified 20 additional publications, i.e., a total of 296 publications.

Before we started with the data collection, we validated whether all identified candidates met the inclusion criteria or violate the exclusion criteria in our last step. This way, we excluded 41 of the identified publications, mainly because they were not peer reviewed (e.g., book chapters, preprints on arXiv), summarized only existing work (e.g., surveys, dissertation summaries), or because they did not contain anything specific to developer social networks, regardless of our initial assessment. This left us with 255 primary studies.

### 3.4. Data collection

Once all literature was identified, we proceeded with the collection of the data required to answer our research questions. For RQ1, we first extracted the research questions and/or hypothesis that were formulated to guide the research, as well as the contributions as listed in the introduction or summarized in the abstract from the publications. We used inductive coding (Thomas, 2006) performed by two researchers to identify the research topics of the papers from the hypothesis and contributions in order to obtain the necessary information to answer RQ1. For this, we printed the title, research questions/hypotheses, and contributions of each publication on a separate sheet of paper and sorted them incrementally by their topic, starting with a coarse-grained separation until we were satisfied that our categories provided a sufficient amount of detail for our mapping study. For RQ2 and RQ3, we extracted the data source, the number of projects, and the number of participants in the DSN used within the publications. For RQ4 and RQ5, we collected meta data about

<sup>1</sup> Scopus was only used for the additional search in the third step and not for the initial search.

**Table 1**

Search terms and number of hits for each search engine.

Search terms	Google Scholar	IEEE Xplore	ACM Digital Library	Springer Link	Elsevier Search	Scopus
Developers network	969,000	4,339	204,258	108,157	60,735	102
Developer social networks	235,000	513	249,607	48,021	26,642	131
Collaborative networks OSS	25,400	22	119,424	1,090	692	0

the publications themselves, i.e., the title, authors, publication venue, year, and number of citations. We organized the collected data in a spreadsheet which is made available as supplementary material (Herbold et al., 2020).

#### 4. Literature review

In this section, we provide the review of the state of the art of DSN research based on the data collection we described in Section 3. We systematically address different topics. We use the data from this review to answer our research questions in Section 5.

##### 4.1. Research directions

Based on the description of the contributions, the research questions, and the research hypotheses of publications, we identified seven general research directions regarding DSNs. For four of the general research directions we identified subtopics, i.e., specific aspects that were considered within the general direction. Table 2 shows our mapping of publications to the research directions including subtopics.

Nearly half of the publications we identified analyze the **community structures** in software development projects. Most of these publications analyzed the general structure of the DSN. However, we also identified seven more specific subtopics of the analysis of community structures: the evolution of the communities by considering DSNs over time; community structures in the context of global software engineering; the formation of teams within development projects; the correlation between the community structure and code quality; the analysis of socio-technical congruence; the simulation of community structures; and the identification of community smells.

DSNs are frequently used for the creation or improvement of **prediction** models for various aspects in software development projects. We identified seven subtopics of prediction approaches using DSNs: bug triage, i.e., support for assigning appropriate developers to work on bug reports; defect prediction, i.e., using the social structure of a project to enhance models that estimate the defect-proneness of different parts of software; recommendation of suitable developers for project work in general; predictions of the outcome of a project, i.e., if projects are likely successful; predictions of suitable Web services; predictions of build failures; and prediction of appropriate developers for code review.

The **collaboration behavior** was also scrutinized using DSNs. While DSNs are modeling some direct or indirect collaboration behavior in software development projects, the analysis of the collaboration behavior itself is in general not the focus. The publications we identified for this research direction focus directly on the collaboration behavior, e.g., which tools were used or how collaboration behavior was impacted by the structure of projects. In addition to research on collaboration behavior in general, we identified three more specific subtopics: collaboration behavior in global software engineering; problems in collaboration behavior and how they are reflected in DSNs; and collaboration between developers from different companies, including competitors in open source projects.

DSNs are also frequently used to assess the **roles of developers** within a development project, e.g., whether a developer is a

core developer or a peripheral developer. While the identification of roles for developers in general is the main topic of this research direction, we also identified two other subtopics; the analysis of how onboarding of peripheral developers within projects works; and how developers specialize within a project.

We also identified research regarding **tools** for DSN analysis, mostly for the visualization of DSNs based on different information sources.

The **validity of DSN research** was also considered by five publications. These publications do not question the validity of DSN research in general, but rather analyze how properties of DSN research may depend on the specific context of research projects, e.g., the scope of the analysis or the repository that was used as source for the DSNs.

Finally, we found one publication on a **data set** that directly contains the graph structure of a DSN. The lack of publications on data sets shows that researchers either generate DSNs from data they collect, or from more general data sets that do not model DSNs directly. Such data sets contain general information mined from software repositories from which a DSN is then built.

**Answer to RQ1:** Community structures are the dominant research direction. Other frequently studied directions are DSNs for predictions, collaboration behavior and developer roles. Tools, studies on validity, and data sets play only a minor role.

##### 4.2. Data sources

There are five major data sources which are used by 241 of the 255 publications:

- Forges like GitHub or SourceForge that are used by millions of developers for hosting and developing open source software. These forges offer an integration of VCSs and ITSs within a single environment, often coupled with other services like Web pages, hosting of releases, or Wikis. Thus, they are a rich source for collaborations between developers, both within a project, as well as across multiple projects.
- ITSs like Jira or Bugzilla are used for the collection, tracking, and management of issues and work items within projects, e.g., change requests, bug reports, or questions by users. ITSs allow the discussion about issues, the definition of work flows for issues, and different types of resolutions.
- VCSs like Git or SVN are systems that track and archive changes of files and folders over time. Typically, VCSs allow different development branches and support working collaboratively on the same resources (Sommerville et al., 2011).
- MLs are collections of email addresses that can be used for communication within software projects. MLs may be restricted, e.g., not everybody may be allowed to post or subscribe to a ML. Participants of MLs may be natural persons (e.g., developers, users), but also systems (e.g., continuous integration systems, ITSs).



**Table 2**

Overview of the literature on DSNs by research directions.

Category	#Pubs.	Publications
<b>Community structure</b>		
General	75	Allaho and Lee (2013), Amrit et al. (2004), Van Antwerp and Madey (2010), Avelino et al. (2019), Bana and Arora (2018), Batista et al. (2017), Behfar and Behfar (2016), Behfar et al. (2018), Hajiakhoond Bidoki and Sukthankar (2018), Bird et al. (2006a,b, 2008), Canfora et al. (2011), Cherry and Robillard (2008), Conaldi and Rullani (2010), Crowston and Howison (2005, 2006), dos Santos et al. (2011), Gao and Madey (2007a,b), Geipel et al. (2014), Gloor et al. (2003), González-Barahona et al. (2004), Avelino et al. (2017), He et al. (2012), Howison et al. (2006), Hu and Zhao (2008), Hu et al. (2012, 2016), Huang and Choi (2011), Ichimura and Uemoto (2015), Iyer and Lyytinen (2019), Jermakovics et al. (2011, 2013), Jiang et al. (2013), Joblin et al. (2015), Kamei et al. (2008), Kidane and Gloor (2007), Leibzon (2016), Lim and Bentley (2011), Lima et al. (2014), Linåker et al. (2019), Long and Siau (2007), Lopez-Fernandez et al. (2004), López-Fernández et al. (2008), Madey et al. (2002), Meneely and Williams (2009, 2010b,a), Mergel (2015), Nzeko'o et al. (2015), Qiu et al. (2019), Robertsa et al. (2006), Schwind et al. (2008), Singh (2010), Sowe et al. (2014), Sureka et al. (2011), Surian et al. (2010), Tamburri et al. (2019), Tan et al. (2007), Thung et al. (2013), Toral et al. (2010), Van Antwerp and Madey (2010a), Wagstrom et al. (2005), Wang et al. (2019), Wiggins et al. (2008), Wolf et al. (2009b), Xu and Madey (2004), Xu et al. (2005a,b), Yu and Ramaswamy (2013), Yu et al. (2014), Zanetti et al. (2012), Zhang et al. (2014) and Zhang et al. (2015)
DSN evolution	18	Aljemabi and Wang (2018), Datta et al. (2011), Hannemann and Klamma (2013), Hong et al. (2011), Joblin et al. (2017b), Kakimoto et al. (2006), Kavaler and Filkov (2017), Kumar and Gupta (2013), Kumar et al. (2019), Nakakoji et al. (2005), Ngamkajornwiwat et al. (2008), Rossi and Neville (2010), Sharma and Kaulgud (2011), Van Antwerp and Madey (2010b), Weiss et al. (2006), Yu et al. (2014), Zanetti et al. (2013) and Zhang et al. (2011)
Global SWE	10	Ahuja et al. (2003), Avritzer et al. (2010), Cataldo and Herbsleb (2008b), De Souza et al. (2007), Ehrlich and Chang (2006), Ehrlich and Cataldo (2012), Hinds and McGrath (2006), Hossain and Zhu (2009), Sarker et al. (0000) and Spinellis (2006)
Team formation	6	Çaglayan et al. (2013), Crowston et al. (2007), Hahn et al. (2006, 2008), Panichella et al. (2014b) and Singh and Tan (2010)
Impact on code quality	6	Bettenburg and Hassan (2010, 2013), Çaglayan and Bener (2016), Datta (2018), Hossain and Zhou (2008) and Mockus (2010)
Socio-technical congruence	5	Cataldo et al. (2008), De Souza et al. (2005), Kwan et al. (2011), Syeed and Hammouda (2013) and Valetto et al. (2007)
Simulation	4	Honsel et al. (2014), Honsel et al. (2015), Honsel et al. (2015) and Yu et al. (2008)
Community smells	2	Catolino et al. (2019) and Palomba et al. (2018)
<b>Prediction</b>		
Bug triage	16	Banitaan and Alenezi (2013), Bhattacharya and Neamtiu (2010), Bhattacharya et al. (2012b), Chen et al. (2010), Jeong et al. (2009), Sun et al. (2017), Wang et al. (2013), Wu et al. (2011), Wu et al. (2018), Xuan et al. (2012b), Yang et al. (2014), Zanetti et al. (2013), Zhang and Lee (2012), Zhang et al. (2013), Zhang et al. (2014a) and Zhang et al. (2014b)
Defect prediction	12	Abreu and Premraj (2009), Bhattacharya et al. (2012a), Biçer et al. (2011), Bird et al. (2009), Hu and Wong (2013), Meneely et al. (2008), Miranskyy et al. (2014), Pinzger et al. (2008), Simpson (2013), Wang and Wang (2016), Zhang et al. (2014b) and Zhang et al. (2019)
Project outcomes	9	Cataldo and Ehrlich (2012), Jarczyk et al. (2018), Liu and Iyer (2007), Peng et al. (2018), Peng (2019), Singh et al. (2011), Surian et al. (2013), Wang et al. (2012) and Wang (2012)
Developers for tasks in general	7	Dráždilová et al. (2012), Hossain et al. (2006), Hu et al. (2018), Li et al. (2016), McDonald (2003), Surian et al. (2011) and Wan et al. (2018)
Suitable web services	3	Bianchini et al. (2015, 2016a) and Bianchini et al. (2016b)
Build failures	2	Schröter (2010) and Wolf et al. (2009a)
Developers for code review	1	Kerzazi and El Asri (2016)
<b>Collaboration behavior</b>		
General	13	Cohen and Consens (2018), Damian et al. (2013), Duc et al. (2011), Feczak and Hossain (2009, 2011), Gharehyazie and Filkov (2017), Kerzazi and El Asri (2017), Licorish and MacDonell (2017), Omoronyia et al. (2009), Ortu et al. (2015), Wu et al. (2016), Xuan et al. (2012a) and Yang (2014)
Global SWE	10	Cataldo and Herbsleb (2008a), Chang and Ehrlich (2007), Damian et al. (2007b), Fonseca et al. (2006), Herbsleb and Mockus (2003), Mikawa et al. (2009), Nguyen et al. (2008), Sarker et al. (2011), Urdangarin et al. (2008) and Wolf et al. (2008)
Problems	10	Begel et al. (2010), Bernardi et al. (2012), Bhowmik et al. (2016), Cataldo et al. (2006), Damian et al. (2007a), Eli (2013), Orsila et al. (2009), Sapkota et al. (2020), Wang and Redmiles (2016) and Xuan et al. (2016)
Inter-company collaboration behavior	1	Teixeira et al. (2015)
<b>Developer roles</b>		
Identification	18	Zhang et al. (2012), Meneely et al. (2010), Bhattacharya et al. (2014b), Crowston et al. (2006), Datta et al. (2010), Dittrich et al. (2013), Huang and Liu (2005), Joblin et al. (2017a), Lee et al. (2013), Licorish and MacDonell (2014, 2015), Lim et al. (2010), Lim and Finkelstein (2012), Marczak et al. (2008), Pohl and Diehl (2008), Sharma et al. (2017), Sowe et al. (2006) and Yu and Ramaswamy (2007)
Onboarding	9	Bird et al. (2007), Canfora et al. (2012), Casalnuovo et al. (2015), Cheng et al. (2017), Ducheneaut (2005), El Asri et al. (2017), Gharehyazie et al. (2013, 2015) and Zhou and Mockus (2011)
Specialization	1	MacLean et al. (2011)
Tools	11	Borici et al. (2012), De Souza et al. (2004), de Souza et al. (2007), Gilbert and Karahalios (2007), Gote et al. (2019), Ogawa et al. (2007), Ohira et al. (2005b,a), Sarma et al. (2009), Schwind and Wegmann (2008) and Schwind et al. (2010)
DSN validity	5	Meneely and Williams (2011), Aljemabi and Wang (2017), Nia et al. (2010), Panichella et al. (2014a) and Tymchuk et al. (2014)
Datasets	1	MacLean and Knutson (2013)

**Table 3**

Data sources that were used for the modeling of the DSNs.

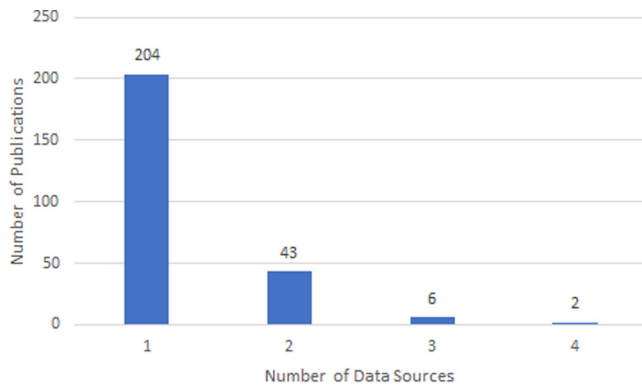
Data source	#Pubs.	Publications
Forge	64	Aljemabi and Wang (2018), Allaho and Lee (2013), Bana and Arora (2018), Batista et al. (2017), Behfar and Behfar (2016), Behfar et al. (2018), Hajiakhoond Bidoki and Sukthankar (2018), Caglayan et al. (2013), Catolino et al. (2019), Cohen and Consens (2018), Conaldi and Rullani (2010), dos Santos et al. (2011), Dráždilová et al. (2012), El Asri et al. (2017), Gao and Madey (2007a,b), Hahn et al. (2006, 2008), He et al. (2012), Hu et al. (2016, 2018), Huang and Choi (2011), Ichimura and Uemoto (2015), Iyer and Lyytinen (2019), Jarczyk et al. (2018), Jiang et al. (2013), Kerzazi and El Asri (2016, 2017), Lee et al. (2013), Leibzon (2016), Li et al. (2016), Lima et al. (2014), Liu and Iyer (2007), Madey et al. (2002), Mergel (2015), Ohira et al. (2005b,a), Peng et al. (2018), Peng (2019), Qiu et al. (2019), Sapkota et al. (2020), Singh (2010), Singh et al. (2011), Surian et al. (2010, 2011, 2013), Tamburri et al. (2019), Tan et al. (2007), Thung et al. (2013), Tymchuk et al. (2014), Van Antwerp and Madey (2010a), Wan et al. (2018), Wang et al. (2012), Wang (2012), Wang et al. (2019), Wu et al. (2016), Xu and Madey (2004), Xu et al. (2005a,b), Yu and Ramaswamy (2013), Yu et al. (2014), Zhang et al. (2014) and Zhang et al. (2015)
ITS	49	Abreu and Premraj (2009), Banitaan and Alenezi (2013), Bettenburg and Hassan (2010), Bhattacharya and Neamtii (2010), Bhattacharya et al. (2012b), Bhowmik et al. (2016), Cataldo et al. (2006, 2008), Cataldo and Ehrlich (2012), Chen et al. (2010), Crowston and Howison (2005), Crowston et al. (2006), Crowston and Howison (2006), Datta et al. (2010), Duc et al. (2011), Ehrlich and Cataldo (2012), Feczak and Hossain (2009, 2011), Hong et al. (2011), Hossain and Zhou (2008), Hossain and Zhu (2009), Howison et al. (2006), Jeong et al. (2009), Kumar and Gupta (2013), Kumar et al. (2019), Licorish and MacDonell (2014, 2015, 2017), Long and Siau (2007), Nguyen et al. (2008), Ortu et al. (2015), Sharma and Kaulgud (2011), Sureka et al. (2011), Wang et al. (2013), Wolf et al. (2008, 2009b,a), Wu et al. (2011), Wu et al. (2018), Xuan et al. (2012b), Yang et al. (2014), Zanetti et al. (2012, 2013), Zanetti et al. (2013), Zhang and Lee (2012), Zhang et al. (2013), Zhang et al. (2014a,b) and Zhou and Mockus (2011)
VCS	41	Van Antwerp and Madey (2010), Avelino et al. (2019), Bird et al. (2009), Casalnuovo et al. (2015), Çaglayan and Bener (2016), Cheng et al. (2017), De Souza et al. (2004, 2005), Dittrich et al. (2013), Ell (2013), González-Barahona et al. (2004), Gote et al. (2019), Avelino et al. (2017), Huang and Liu (2005), Jermakovics et al. (2011, 2013), Joblin et al. (2015, 2017a,b), Kakimoto et al. (2006), Lopez-Fernandez et al. (2004), López-Fernández et al. (2008), MacLean and Knutson (2013), Meneely et al. (2008), Meneely and Williams (2009, 2010b, 2011), Miranskyy et al. (2014), Mockus (2010), Orsila et al. (2009), Palomba et al. (2018), Pinzger et al. (2008), Pohl and Diehl (2008), Schwind and Wegmann (2008), Schwind et al. (2008, 2010), Sun et al. (2017), Teixeira et al. (2015), Valetto et al. (2007), Van Antwerp and Madey (2010b) and Yu and Ramaswamy (2007)
ML	23	Ahuja et al. (2003), Bird et al. (2006a,b, 2008), Gharehyazie et al. (2015), Gloor et al. (2003), Hossain et al. (2006), Kamei et al. (2008), Kavalier and Filkov (2017), Kidane and Gloor (2007), Nakakoji et al. (2005), Ngamkajornwiwat et al. (2008), Nia et al. (2010), Nzeko'o et al. (2015), Robertsa et al. (2006), Sharma et al. (2017), Sowe et al. (2006), Toral et al. (2010), Weiss et al. (2006), Xuan et al. (2016), Yu et al. (2008), Zhang et al. (2012) and Zhang et al. (2014b)
Other	14	Amrit et al. (2004), Bianchini et al. (2015, 2016a,b), Borici et al. (2012), Cataldo and Herbsleb (2008a), Damian et al. (2007a), de Souza et al. (2007), Hu and Zhao (2008), Hu et al. (2012), Hu and Wong (2013), Omoronyia et al. (2009), Wang and Wang (2016) and Yang (2014)
Survey	13	Avritzer et al. (2010), Chang and Ehrlich (2007), Cherry and Robillard (2008), Ehrlich and Chang (2006), Hinds and McGrath (2006), Lim et al. (2010), Lim and Bentley (2011), Lim and Finkelstein (2012), McDonald (2003), Mikawa et al. (2009), Sarker et al. (0000, 2011) and Urdangarin et al. (2008)
ITS & VCS	19	Aljemabi and Wang (2017), Bernardi et al. (2012), Bettenburg and Hassan (2013), Bhattacharya et al. (2012a, 2014b), Biçer et al. (2011), Datta et al. (2011), Datta (2018), Honsel et al. (2014), Honsel et al. (2015), Honsel et al. (2015), Kwan et al. (2011), Linåker et al. (2019), Meneely et al. (2010), Meneely and Williams (2010a), Schröter (2010), Simpson (2013), Spinellis (2006) and Zhang et al. (2019)
ML & VCS	14	Bird et al. (2007), Canfora et al. (2011, 2012), Ducheneaut (2005), Fonseca et al. (2006), Gharehyazie and Filkov (2017), Gilbert and Karahalios (2007), Hannemann and Klamma (2013), Ogawa et al. (2007), Singh and Tan (2010), Sowe et al. (2014), Syeed and Hammouda (2013), Xuan et al. (2012a) and Zhang et al. (2011)
Survey & Other	3	Damian et al. (2007b, 2013) and Marczak et al. (2008)
ML & ITS	2	MacLean et al. (2011) and Rossi and Neville (2010)
ML & Other	2	Crowston et al. (2007) and Wagstrom et al. (2005)
Forge & Survey	1	De Souza et al. (2007)
ITS & Survey	1	Herbsleb and Mockus (2003)
VCS & Forge	1	Geipel et al. (2014)
ITS, ML & VCS	3	Gharehyazie et al. (2013), Panichella et al. (2014b) and Sarma et al. (2009)
ML, ITS & Other	1	Wiggins et al. (2008)
ITS, Survey & Other	1	Cataldo and Herbsleb (2008b)
ITS, VCS & Other	1	Begel et al. (2010)
ITS, ML, CVS & Other	2	Panichella et al. (2014a) and Wang and Redmiles (2016)

- Surveys, i.e., interviews or questionnaires that were used to directly ask developers about their communication behavior within a development project.

In addition to the five major sources, there are other ways that researchers used to collect information about collaboration behavior which we summarized as “Other” in Table 3. These are IRC chats (Cataldo and Herbsleb, 2008b,a; Panichella et al., 2014a; Wang and Redmiles, 2016), plug-ins that monitor development environments (Omoronyia et al., 2009; Borici et al., 2012;

de Souza et al., 2007), manual inspection of project documents, e.g., requirements (Damian et al., 2013, 2007b; Marczak et al., 2008), owners of web service mash-ups (Bianchini et al., 2016a,b, 2015), the web site Ohloh that provides statistics about open source development<sup>2</sup> (Hu and Zhao, 2008; Hu et al., 2012), online discussion forums (Wiggins et al., 2008; Crowston et al., 2007), JAR files (Hu and Wong, 2013), the BlogLinks and Advogato social

<sup>2</sup> The name has changed to <https://www.openhub.net/>.



**Fig. 3.** Number of data sources that were used for the modeling of the DSNs within the publications.

networks<sup>3</sup> of software developers (Wagstrom et al., 2005), on site researchers that observe communication behavior (Damian et al., 2007a), employee directories (Begel et al., 2010), and the code review portal Gerrit (Yang, 2014). Additionally, one publication discusses DSNs from an abstract perspective and proposes the use of tracking for every communication including phone calls, emails, etc. (Amrit et al., 2004).

Fig. 3 depicts the number of data sources that were used for modeling DSNs. It highlights that 204 of the 255 publications build a DSN that is based on a single source, 43 publications used a combination of two data sources, six publication three data sources and two publications four data sources.

**Answer to RQ2:** Software repositories like forges, ITSs, VCSs and MLs are the main sources for DSNs, however, surveys are also sometimes used. Publications commonly use a single source for DSN modeling. The knowledge about DSNs built with multiple sources is limited.

#### 4.3. Number of projects analyzed

A major factor regarding the external validity of results is the number of projects for which data is collected. If only data about very few projects is used for an empirical study about a phenomenon that can be studied using DSNs, the results may not generalize to other projects. The likelihood that the results generalize to software engineering in general increases with the number of projects that are analyzed. Table 4 shows the number of projects per publication. The data we collected shows that most papers on DSNs perform some sort of empirical study to demonstrate their approach or research a phenomenon. Only 12 of the 255 publications we identified did not perform any empirical study. Moreover, we identified 16 publications for which we could not identify the number of projects from the publication. There were two reasons for this: either the authors did not report how they selected a smaller subset from a larger database or the authors did not specify which projects were used at all. This is not only problematic for evaluating the external validity of a study, but also hinders replications of the results. Of the 227 publications for which we could identify the number of projects, 76 used only a single project for their empirical study, 69 used only 2–5 projects for the empirical study. In other words, about 33% of the publications on DSNs used a single project, another 30% used 2–5 projects. Both numbers are extremely low and do not allow for a

generalization of the findings due to the limited context covered by the projects. Another 12 publications only considered 6–10 projects, which is still a small number. On the bright side, 50 publications used more than 100 projects, i.e., larger sample sizes that usually allow to generalize findings. 38 of these publications use a forge as data source. Regardless, our analysis of the sample sizes with respect to the number of projects indicates a severe threat to the external validity of many empirical studies on DSNs.

**Answer to RQ3a:** Over 69% of all publications use less than 11 projects to evaluate their findings. Most publications with at least 100 projects use a forge as data source (38 of 50).

#### 4.4. Number of developers in the DSNs

The second major factor regarding the validity of results is the number of people that are part of the DSNs. Table 5 shows the data we collected regarding the number of people in the DSNs. In case a publication created multiple DSNs, e.g., one per project considered, we report the mean value of the people in the DSNs. The number of people modeled by the DSNs is relatively high. 77 publications have more than 1000 people as part of their DSNs, 15 publications actually model more than 100,000 people. Only four publications have very small networks with less than or equal to 10 people, another 32 publications consider less than or equal to 100 people. Thus, for the publications for which the data about the number of people is available, the networks that are considered are in general relatively large. When we looked closely at the data, we observed two reasons for this: first, while many publications consider only few projects, these projects tend to be very large, e.g., Mozilla Firefox and the Eclipse IDE. Moreover, our data also shows that MLs and forges are the most common data sources for DSNs. Both capture not only developers, but also users of the respective projects. We also found a very concerning general trend in the literature: 66 of the 240 publications that performed an empirical study did not report the number of participants in the DSN. This is a vital piece of information for the estimation of both the internal and external validity of empirical studies that should always be reported.

**Answer to RQ3b:** Most publications report networks that have more than 100 vertices. The number of developers is often much larger than the number of projects, because large-scale projects with big communities are analyzed.

#### 4.5. Influential publications

We collected data regarding the citation counts from Google Scholar. We take the pattern from the ACM Distinguished Paper awards to define our criterion for influential publications, and consider the top 10% with the most citations as influential. Since we have 255 publications, this means we consider the 25 publications with the most citations (Table 6). We note that the citations for the third most cited paper (Bird et al., 2006a) also include the citations for the paper (Bird et al., 2006b), because the two publications are considered as the same paper by Google Scholar. The 25 most influential publications address

- software development with globally distributed project members (Herbsleb and Mockus, 2003; Ahuja et al., 2003; Hinds and McGrath, 2006);

<sup>3</sup> Both are not available online anymore.

**Table 4**

Number of projects that were analyzed as part of an empirical study of DSNs. Missing means that the number of projects is not or not accurately reported in the publication, NA means that the publication did not conduct an empirical study.

#Projects	#Pubs.	Publications
1	76	Abreu and Premraj (2009), Ahuja et al. (2003), Avritzer et al. (2010), Bettenburg and Hassan (2010), Bird et al. (2006a,b), Caglayan et al. (2013), Cataldo et al. (2006), Cataldo and Herbsleb (2008a), Cataldo et al. (2008), Cataldo and Ehrlich (2012), Cherry and Robillard (2008), Damian et al. (2007a), Datta et al. (2010, 2011), Datta (2018), Ducheneaut (2005), Ehrlich and Cataldo (2012), Eli (2013), González-Barahona et al. (2004), He et al. (2012), Hong et al. (2011), Honsel et al. (2014), Honsel et al. (2015), Hossain et al. (2006), Hu and Zhao (2008), Kamei et al. (2008), Kumar and Gupta (2013), Kwan et al. (2011), Li et al. (2016), Licorish and MacDonell (2014, 2015, 2017), Lim et al. (2010), Lim and Bentley (2011), Lim and Finkelstein (2012), Linäker et al. (2019), MacLean et al. (2011), Marczak et al. (2008), McDonald (2003), Meneely et al. (2008), Meneely and Williams (2009), Meneely et al. (2010), Meneely and Williams (2010a), Mikawa et al. (2009), Miranskyy et al. (2014), Mockus (2010), Nakakoji et al. (2005), Ngamkajornwiwat et al. (2008), Nguyen et al. (2008), Omoronyia et al. (2009), Orsila et al. (2009), Pinzger et al. (2008), Pohl and Diehl (2008), Roberts et al. (2006), Rossi and Neville (2010), Sarma et al. (2009), Schröter (2010), Sharma and Kaulgud (2011), Sharma et al. (2017), Simpson (2013), Spinellis (2006), Sureka et al. (2011), Syeed and Hammouda (2013), Toral et al. (2010), Urdangarin et al. (2008), Wolf et al. (2008, 2009a), Wu et al. (2011), Yang et al. (2014), Zanetti et al. (2013), Zhang et al. (2011), Zhang and Lee (2012), Zhang et al. (2012) and Zhang et al. (2014a)
2–5	69	Aljemabi and Wang (2018), Banitaan and Alenezi (2013), Bernardi et al. (2012), Bettenburg and Hassan (2013), Bhattacharya and Neamtiu (2010), Bhattacharya et al. (2012b, 2014b), Bhowmik et al. (2016), Biçer et al. (2011), Bird et al. (2007, 2008, 2009), Canfora et al. (2011, 2012), Cataldo and Herbsleb (2008b), Çaglayan and Bener (2016), Chang and Ehrlich (2007), Chen et al. (2010), Crowston et al. (2007), Damian et al. (2013), de Souza et al. (2007), Dittrich et al. (2013), Duc et al. (2011), Ehrlich and Chang (2006), El Asri et al. (2017), Gilbert and Karahalios (2007), Gloor et al. (2003), Hannemann and Klamma (2013), Honsel et al. (2015), Hu and Wong (2013), Jeong et al. (2009), Jermakovics et al. (2011, 2013), Kakimoto et al. (2006), Kavalier and Filkov (2017), Kerzazi and El Asri (2016, 2017), Kidane and Gloor (2007), Kumar et al. (2019), Leibzon (2016), Lopez-Fernandez et al. (2004), López-Fernández et al. (2008), Meneely and Williams (2010b, 2011), Nia et al. (2010), Nzeko'o et al. (2015), Ogawa et al. (2007), Panichella et al. (2014b), Sarker et al. (2011), Schwind and Wegmann (2008), Schwind et al. (2008), Singh and Tan (2010), Sowe et al. (2006), Sun et al. (2017), Van Antwerp and Madey (2010b), Wang et al. (2013), Wang and Wang (2016), Wang and Redmiles (2016), Wiggins et al. (2008), Wu et al. (2018), Xuan et al. (2012b), Yang (2014), Yu and Ramaswamy (2007), Yu et al. (2008), Zanetti et al. (2013), Zhang et al. (2013, 2014b), Zhang et al. (2014b) and Zhang et al. (2019)
6–10	12	Bhattacharya et al. (2012a), Gharehyazie et al. (2013, 2015), Avelino et al. (2017), Huang and Liu (2005), Joblin et al. (2015, 2017a), Ortu et al. (2015), Palomba et al. (2018), Panichella et al. (2014a), Teixeira et al. (2015) and Zhou and Mockus (2011)
11–100	20	Aljemabi and Wang (2017), Catolino et al. (2019), Crowston and Howison (2005), De Souza et al. (2007), dos Santos et al. (2011), Geipel et al. (2014), Gharehyazie and Filkov (2017), Hinds and McGrath (2006), Hossain and Zhou (2008), Hossain and Zhu (2009), Joblin et al. (2017b), Sarker et al. (2000), Sowe et al. (2014), Tamburri et al. (2019), Weiss et al. (2006), Xuan et al. (2012a, 2016), Zanetti et al. (2012), Zhang et al. (2014) and Zhang et al. (2015)
> 100	50	Allaho and Lee (2013), Van Antwerp and Madey (2010), Avelino et al. (2019), Batista et al. (2017), Bianchini et al. (2015, 2016a), Hajiakhoond Bidoki and Sukthakar (2018), Casalnuovo et al. (2015), Cheng et al. (2017), Cohen and Consens (2018), Conaldi and Rullani (2010), Crowston et al. (2006), Crowston and Howison (2006), Dráždilová et al. (2012), Feczak and Hossain (2009, 2011), Hahn et al. (2006, 2008), Howison et al. (2006), Hu et al. (2012, 2018), Huang and Choi (2011), Iyer and Lytinen (2019), Jarczyk et al. (2018), Jiang et al. (2013), Lee et al. (2013), Liu and Iyer (2007), Long and Siau (2007), Madey et al. (2002), Mergel (2015), Ohira et al. (2005b), Peng et al. (2018), Peng (2019), Sapkota et al. (2020), Singh (2010), Singh et al. (2011), Surian et al. (2010, 2011, 2013), Tan et al. (2007), Thung et al. (2013), Tymchuk et al. (2014), Wan et al. (2018), Wang et al. (2012), Wang (2012), Wang et al. (2019), Wu et al. (2016), Xu et al. (2005a,b) and Yu and Ramaswamy (2013)
Missing	16	Bana and Arora (2018), Behfar and Behfar (2016), Behfar et al. (2018), Bianchini et al. (2016b), Damian et al. (2007b), Gao and Madey (2007a), Herbsleb and Mockus (2003), Hu et al. (2016), Lima et al. (2014), MacLean and Knutson (2013), Qiu et al. (2019), Van Antwerp and Madey (2010a), Wagstrom et al. (2005), Xu and Madey (2004), Yu et al. (2014) and Yu et al. (2014)
NA	12	Amrit et al. (2004), Begel et al. (2010), Borici et al. (2012), De Souza et al. (2004, 2005), Fonseca et al. (2006), Gao and Madey (2007b), Gote et al. (2019), Ichimura and Uemoto (2015), Ohira et al. (2005a), Schwind et al. (2010) and Valetto et al. (2007)

- community structures in software development projects (Bird et al., 2006a; Crowston and Howison, 2005; Ducheneaut, 2005; Madey et al., 2002; Bird et al., 2008; Crowston and Howison, 2006; Lopez-Fernandez et al., 2004; Xu et al., 2005b);
- the formation of teams in projects through collaboration (Hahn et al., 2008; Crowston et al., 2007);
- the identification of relationships between developers (Begel et al., 2010);
- the impact of coordination requirements between developers on tool design (Cataldo et al., 2006) and modularization (Cataldo et al., 2008);
- communication issues (Damian et al., 2007a) and trust (Sarker et al., 2011);
- the identification of core developers (Crowston et al., 2006);
- predictions to support software engineering processes, i.e., bug triage (Jeong et al., 2009), defect prediction (Pinzger et al., 2008; Meneely et al., 2008; Bird et al., 2009), build failure prediction (Wolf et al., 2009a), and collaborations (McDonald, 2003).

**Answer to RQ4a:** There are many publications on DSNs with a high citation count. The most influential publications address a very diverse number of topics, which highlights that there are many use cases for DSNs in software engineering research.

#### 4.6. Influential authors

We identified 481 different authors who contributed to the literature on DSNs. We use a bibliometric approach to identify the most influential of these authors, based on three different indicators: (1) the number of citations of all publications on DSNs; (2) the number of publications on DSNs; and (3) the number of publications on DSNs we identified as influential (Section 4.5). We consider the top-5 authors in each category to be the most influential. For the bibliometric data we collected, this means that an author has to have at least 1397 citations, 8 publications, or 3 influential publications to be considered as one of the most influential authors.



**Table 5**

Number of people that are inside the DSNs. Missing means that the number of people is not or not accurately reported in the publication, NA means that the publication did not conduct an empirical study.

#People	#Pubs.	Publications
1–10	4	Cherry and Robillard (2008), Damian et al. (2007b), Omoronyia et al. (2009) and Pohl and Diehl (2008)
11–100	32	Abreu and Premraj (2009), Ahuja et al. (2003), Avritzer et al. (2010), Banitaan and Alenezi (2013), Bird et al. (2008), Cataldo and Ehrlich (2012), Chang and Ehrlich (2007), Crowston et al. (2007), Damian et al. (2013), Datta et al. (2010), De Souza et al. (2007), Ehrlich and Chang (2006), Gharehyazie and Filkov (2017), Honsel et al. (2015), Huang and Liu (2005), Jarczyk et al. (2018), Jermakovics et al. (2011), Kakimoto et al. (2006), Kavalier and Filkov (2017), Lim and Bentley (2011), Marczak et al. (2008), Meneely et al. (2010), Meneely and Williams (2010a), Mikawa et al. (2009), Palomba et al. (2018), Panichella et al. (2014a), Sarker et al. (0000, 2011), Schwind et al. (2008), Urdangarin et al. (2008), Yu and Ramaswamy (2007) and Zhang et al. (2012)
101–1000	64	Aljemabi and Wang (2017), Bianchini et al. (2015, 2016a), Caglayan et al. (2013), Canfora et al. (2012), Cataldo et al. (2006), Cataldo and Herbsleb (2008a), Cataldo et al. (2008), Crowston and Howison (2006), Datta et al. (2011), Datta (2018), Dittrich et al. (2013), Duc et al. (2011), Ducheneaut (2005), Ehrlich and Cataldo (2012), Gharehyazie et al. (2013, 2015), Gloor et al. (2003), Avelino et al. (2017), Herbsleb and Mockus (2003), Honsel et al. (2015), Hossain et al. (2006), Huang and Choi (2011), Jermakovics et al. (2013), Joblin et al. (2015, 2017a,b), Kamei et al. (2008), Kerzazi and El Asri (2017), Kwan et al. (2011), Licorish and MacDonell (2014, 2015, 2017), Lim and Finkelstein (2012), Linäker et al. (2019), López-Fernández et al. (2008), MacLean et al. (2011), Meneely et al. (2008), Meneely and Williams (2009, 2010b, 2011), Mockus (2010), Ngamkajornwivat et al. (2008), Nguyen et al. (2008), Orsila et al. (2009), Ortu et al. (2015), Robertsa et al. (2006), Rossi and Neville (2010), Schwind and Wegmann (2008), Sowe et al. (2014), Spinellis (2006), Sun et al. (2017), Surian et al. (2011), Tamburri et al. (2019), Tymchuk et al. (2014), Weiss et al. (2006), Wolf et al. (2008, 2009a), Wu et al. (2011), Xuan et al. (2012a), Zanetti et al. (2012, 2013), Zhang et al. (2011) and Zhang and Lee (2012)
1001–10 000	35	Aljemabi and Wang (2018), Avelino et al. (2019), Behfar and Behfar (2016), Behfar et al. (2018), Bhowmik et al. (2016), Hajiakhoond Bidoki and Sukthar (2018), Bird et al. (2006a,b), Canfora et al. (2011), Casalnuovo et al. (2015), El Asri et al. (2017), Hannemann and Klammer (2013), He et al. (2012), Honsel et al. (2014), Hu and Zhao (2008), Jeong et al. (2009), Kerzazi and El Asri (2016), Kumar et al. (2019), Leibzon (2016), Li et al. (2016), Lim et al. (2010), Liu and Iyer (2007), Nakakoji et al. (2005), Nia et al. (2010), Nzeko'o et al. (2015), Ogawa et al. (2007), Singh (2010), Sowe et al. (2006), Sureka et al. (2011), Syeed and Hammouda (2013), Toral et al. (2010), Xuan et al. (2016), Yu et al. (2008), Zhang et al. (2014a) and Zhang et al. (2019)
10001–100 000	27	Van Antwerp and Madey (2010), Batista et al. (2017), Bhattacharya et al. (2014b), Bird et al. (2007), Cohen and Consens (2018), Dráždilová et al. (2012), Hu et al. (2012), Kumar and Gupta (2013), Long and Siau (2007), Madey et al. (2002), Panichella et al. (2014b), Qiu et al. (2019), Sapkota et al. (2020), Sarma et al. (2009), Sharma et al. (2017), Surian et al. (2010), Tan et al. (2007), Thung et al. (2013), Van Antwerp and Madey (2010a,b), Wan et al. (2018), Wang et al. (2013), Wu et al. (2018), Xuan et al. (2012b), Zanetti et al. (2013), Zhang et al. (2013) and Zhou and Mockus (2011)
> 100 000	15	Allaho and Lee (2013), Bernardi et al. (2012), Conaldi and Rullani (2010), Gao and Madey (2007a), Hahn et al. (2008), Hong et al. (2011), Hu et al. (2018), Jiang et al. (2013), Lima et al. (2014), Ohira et al. (2005b), Wang et al. (2019), Xu et al. (2005b), Yu and Ramaswamy (2013), Yu et al. (2014) and Yu et al. (2014)
Missing	66	Bana and Arora (2018), Bettenburg and Hassan (2010, 2013), Bhattacharya and Neamtui (2010), Bhattacharya et al. (2012b,a), Bianchini et al. (2016b), Biçer et al. (2011), Bird et al. (2009), Cataldo and Herbsleb (2008b), Catolino et al. (2019), Çaglayan and Bener (2016), Chen et al. (2010), Cheng et al. (2017), Crowston and Howison (2005), Crowston et al. (2006), Damian et al. (2007a), de Souza et al. (2007), dos Santos et al. (2011), Ell (2013), Feczak and Hossain (2009, 2011), Geipel et al. (2014), Gilbert and Karahalios (2007), González-Barahona et al. (2004), Hahn et al. (2006), Hinds and McGrath (2006), Hossain and Zhou (2008), Hossain and Zhu (2009), Howison et al. (2006), Hu and Wong (2013), Hu et al. (2016), Iyer and Lyytinen (2019), Kidane and Gloor (2007), Lee et al. (2013), Lopez-Fernandez et al. (2004), MacLean and Knutson (2013), McDonald (2003), Mergel (2015), Miranskyy et al. (2014), Peng et al. (2018), Peng (2019), Pinzger et al. (2008), Schröter (2010), Sharma and Kaulgud (2011), Simpson (2013), Singh and Tan (2010), Singh et al. (2011), Surian et al. (2013), Teixeira et al. (2015), Wagstrom et al. (2005), Wang et al. (2012), Wang (2012), Wang and Wang (2016), Wang and Redmiles (2016), Wiggins et al. (2008), Wolf et al. (2009b), Wu et al. (2016), Xu and Madey (2004), Xu et al. (2005a), Yang (2014), Yang et al. (2014), Zhang et al. (2014b), Zhang et al. (2014b) and Zhang et al. (2015)
NA	12	Amrit et al. (2004), Begel et al. (2010), Borici et al. (2012), De Souza et al. (2004, 2005), Fonseca et al. (2006), Gao and Madey (2007b), Gote et al. (2019), Ichimura and Uemoto (2015), Ohira et al. (2005a), Schwind et al. (2010) and Valetto et al. (2007)

Table 7 shows the nine most influential authors we identified according to these criteria. Below, we briefly summarize the research directions of the influential authors. We discuss authors that frequently collaborated with each other as a group.

- James D. Herbsleb and Kathleen M. Carley are co-authors of three influential publications as well as several other publications. Herbsleb and Carley are both professors at Carnegie Mellon University. Their work covers structures and collaboration in global software engineering as well as socio-technical congruence within projects.
- Premkumar Devanbu was the PhD advisor of Christian Bird, who wrote his dissertation on DSNs. Their work addressed social structures and openness of open source projects, as well as build failure prediction.
- Kevin Crowston was the PhD advisor of James Howison, who wrote his dissertation on DSNs. Their work addressed community structures for open source software development.
- Daniela Damian collaborated with different authors as part of her work on communication between developers from different perspectives.

- Gregory Madey was the lead author of the first paper on DSNs we identified. He enabled many early papers through the SourceForge Research Data Archive (Van Antwerp and Madey, 2008).
- Vladimir Filkov contributed to different aspects, including homophily, developer initiation into projects, communication behavior, as well as general structural aspects of DSNs.

**Answer to RQ4b:** We identified 9 out of 481 authors as highly influential. The most influential author is James D. Herbsleb with over 2200 citations and 7 publications within the field of DSNs, three of which are among the top 10% of all publications with respect to the number of citations.

#### 4.7. Important venues

The identified papers were published in 118 different venues, i.e., journals, conferences, and workshops. Table 8 lists the venues at which most papers on DSNs were published. Three

**Table 6**

Top 10% of publications ranked by number of citations. Data according to Google Scholar collected on 2020-02-24.

Title	Authors	Year	#Cit.
An empirical study of speed and communication in globally distributed software development	James D. Herbsleb, Audris Mockus	2003	1127
Individual Centrality and Performance in Virtual R&D Groups: An Empirical Study	Manju K. Ahuja, Dennis F. Galletta, Kathleen M. Carley	2003	665
Mining email social networks	Christian Bird, Alex Gourley, Premkumar Devanbu, Michael Gertz, Anand Swaminathan	2006	644
The social structure of free and open source software development	Kevin Crowston, James Howison	2005	602
Identification of Coordination Requirements: Implications for the Design of Collaboration and Awareness Tools	Marcelo Cataldo, Patrick A. Wagstrom, James D. Herbsleb, Kathleen M. Carley	2006	465
Socialization in an Open Source Software Community: A Socio-Technical Analysis	Nicolas Ducheneaut	2005	459
Improving Bug Triage with Bug Tossing Graphs	Gaeul Jeong, Sunghun Kim, Thomas Zimmermann	2009	434
The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory	Gregory Madey, Vincent Freeh, Renee Tynan	2002	342
The role of communication and trust in global virtual teams: A social network perspective	Saonee Sarker, Manju K. Ahuja, Suprateek Sarker, Sarah Kirkeby	2011	313
Latent social structure in open source projects	Christian Bird, David Pattison, Raissa D'Souza, Vladimir Filkov, Premkumar Devanbu	2008	301
Socio-Technical Congruence: A Framework for Assessing the Impact of Technical and Work Dependencies on Software Development Productivity	Marcelo Cataldo, James D. Herbsleb, Kathleen M. Carley	2008	300
Emergence of New Project Teams from Open Source Software Developer Networks: Impact of Prior Collaboration Ties	Jungpil Hahn, Jae Y. Moon, Chen Zhang	2008	296
Can developer-module networks predict failures?	Martin Pinzger, Nachiappan Nagappan, Brendan Murphy	2008	243
Predicting failures with developer networks and social network analysis	Andrew Meneely, Laurie Williams, Will Snipes, Jason Osborne	2008	241
Self-organization of teams for free/libre open source software development	Kevin Crowston, Qing Li, Kangning Wei, U. Y. Eseryel, James Howison	2007	240
Recommending collaboration with social networks: A comparative evaluation	David W. McDonald	2003	226
Codebook: discovering and exploiting relationships in software repositories	Andrew Begel, Yit P. Khoo, Thomas Zimmermann	2010	222
Predicting build failures using social network analysis on developer communication	Timo Wolf, Adrian Schröter, Daniela Damian, Thanh H.D. Nguyen	2009	217
Awareness in the Wild: Why Communication Breakdowns Occur	Daniela Damian, Luis Izquierdo, Janice Singer, Irwin Kwan	2007	214
Structures that work: social structure, work structure and coordination ease in geographically distributed teams	Pamela Hinds, Cathleen McGrath	2006	209
Applying social network analysis to the information in CVS repositories	Luis Lopez-Fernandez, Gregorio Robles, Jesus M. Gonzales-Barahona	2004	207
Core and Periphery in Free/Libre and Open Source Software Team Communications	Kevin Crowston, Kangning Wei, Qing Li, James Howison	2006	204
Hierarchy and centralization in free and open source software team communications	Kevin Crowston, James Howison	2006	200
A Topological Analysis of the Open Source Software Development Community	Jin Xu, Yongqin Gao, Scott Christley, Gregory Madey	2005	198
Putting It All Together: Using Socio-technical Networks to Predict Failures	Christian Bird, Nachiappan Nagappan, Harald Gall, Brendan Murphy, Premkumar Devanbu	2009	197

conferences stand out: the International Conference on Open Source Software (OSS), the International Conference on Software Engineering (ICSE), and the International Conference on Mining Software Repositories (MSR). 18% of all papers on DSNs were published at these three venues. This is not surprising, as most publications analyze open source projects or ecosystems and employ software repository mining techniques. The ICSE is the top conference in the software engineering field, which highlights that there are papers of outstanding quality on DSNs. We note that the venues with most publications are mostly conferences,

which is in line with the general conference-centered publication system of computer science research. The only three journals that made it into this list are Empirical Software Engineering, Information and Software Technology and the Journal of Systems and Software. However, there are also publications in other premier journals, e.g., the IEEE Transactions on Software Engineering (Kwan et al., 2011; Herbsleb and Mockus, 2003; Lim and Finkelstein, 2012; Catolino et al., 2019), ACM Transactions on Software Engineering Methodology (Singh, 2010), Management Information Systems Quarterly (Singh et al., 2011), and PLOS ONE (Xuan et al., 2016; Sapkota et al., 2020).

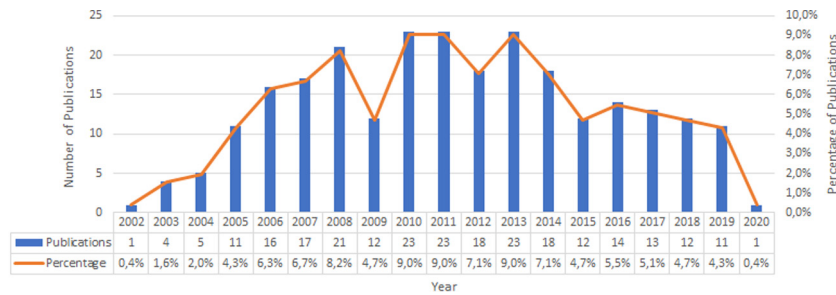


Fig. 4. Publications per year.

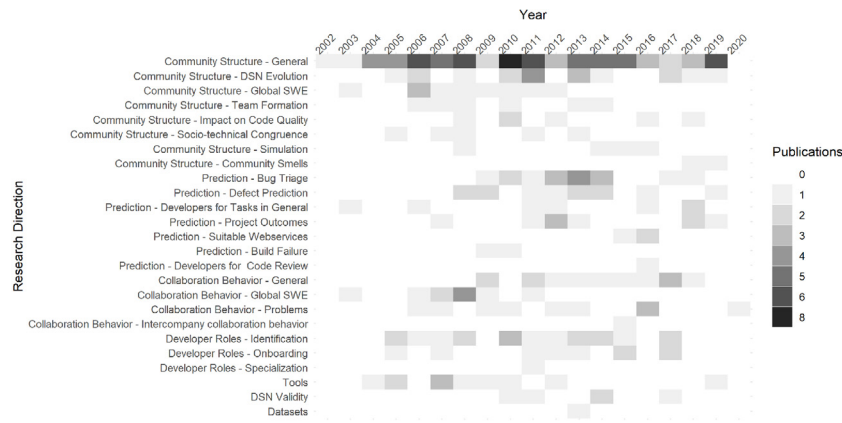


Fig. 5. Research directions over time.

Table 7

Most influential authors according to the number of citations, number of publications, and number of influential publications.

Author	#Cit.	#Pubs.	#Influential Pubs.
James D. Herbsleb	2271	7	3
Kathleen M. Carley	1488	4	3
Premkumar Devanbu	1477	10	3
Christian Bird	1472	8	3
James Howison	1397	6	4
Kevin Crowston	1397	6	4
Daniela Damian	1026	11	2
Gregory Madey	704	9	2
Vladimir Filkov	498	9	1

**Answer to RQ4c:** The papers on DSNs were published in 118 different venues, including journals, conferences, and workshops. The most prominent venues are the ICSE, the OSS, and the MSR. Only three journals are in the list of the most important venues, which highlights that most DSNs research is published at conferences.

#### 4.8. Interest over time

Another interesting aspect is the interest of researchers with respect to DSNs over time measured by the number of publications per year. Fig. 4 depicts the total number of publications per year since the initial publication by Madey et al. (2002) in 2002. The topic quickly gained traction in the research community with rising numbers of publications until the interest became steady with 11 to 21 publications per year between 2005 and 2018. However, there seems to be a slight decline in the interest in DSNs since 2014. We note that due to the time of our search, the data for 2020 (and possibly 2019) is incomplete.

Fig. 5 shows a heat map of the research directions over time. The interest in the general structure of DSNs has been steadily high over time. For all research directions, we see a fairly steady interest in the years 2005 to 2013. From 2013 onwards, the decline in overall interest in DSNs is also reflected in a more erratic interest in specific research directions. The most notable decline since 2013 is regarding global software engineering and tools, for which there are no publications anymore, even though the interest before was fairly high. Another important aspect is regarding the recent research, i.e., the years 2018 onwards. We note that the diversity of the research topics for which DSNs are used has declined, i.e., the ongoing research currently seems to be focused on community structures and predictive usages of DSNs. However, we see that there are still new topics emerging, i.e., the analysis of community smells (Palomba et al., 2018; Catolino et al., 2019).

**Answer to RQ5:** The interest in DSNs peaked from 2006 till 2013. Regardless, there is still a steady interest in the topic, albeit we note a decrease in the diversity of research topics since 2018.

## 5. Discussion of open issues

Our mapping study shows that DSNs are a versatile method for software engineering research. Mostly, they are used for the analysis of social structures and communication. However, the applications of DSNs range beyond that, e.g., for predictive purposes. Within this section, we discuss open problems in DSN research.

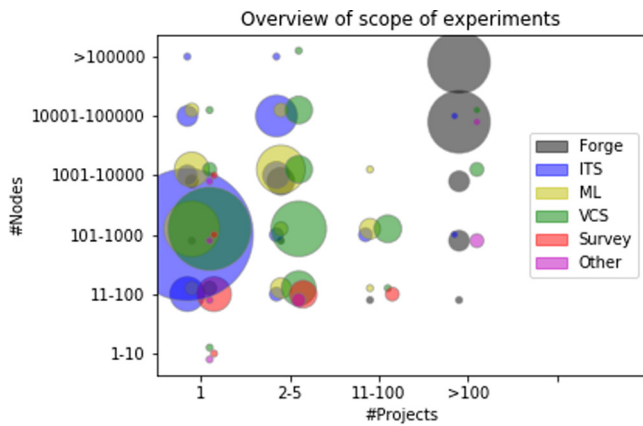
### 5.1. General issues

Here, we discuss general issues within the current body of work on DSNs, that should be addressed by future work.

**Table 8**

Most important publication venues determined by the number of papers published. We omitted labels like IEEE, ACM, or similar from the conference names, as they often changed slightly throughout the years.

Venue	#Pubs.
International Conference on Software Engineering (ICSE)	17
International Conference on Open Source Software (OSS)	16
International Conference on Mining Software Repositories (MSR)	15
(Workshop until 2007, Working Conference until 2015)	
Conference on Computer Supported Cooperative Work (CSCW)	10
International Conference on the Foundations of Software Engineering (FSE)	8
Hawaii International Conference on System Sciences (HICSS)	8
Asia-Pacific Software Engineering Conference (APSEC)	8
Empirical Software Engineering, Springer	7
International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)	7
Information and Software Technology, Elsevier	6
Journal of Systems and Software (JSS)	5
International Conference on Global Software Engineering (ICGSE)	5
International Conference on Software Maintenance and Evolution (ICSME)	5
(ICSM until 2013)	



**Fig. 6.** Relationship between study sizes and data sources.

#### 5.1.1. Lack of guidelines

There are no guidelines on how to conduct DSN research. Therefore, the studies on DSNs are performed and described very heterogeneously. This is not an issue in itself, as heterogeneity can also be positive if different aspects are analyzed. Moreover, many publications perform well-designed case studies and report all important data regardless of the lack of guidelines. However, we observed several issues that result from the inconsistent way studies with DSNs are performed:

- lack of reporting of the exact data sources and/or selection criteria for case study subjects;
- lack of reporting of important meta data about the study, e.g., number of projects, number of people; and
- lack of reporting of pre-processing steps performed with the data, e.g., to merge identities in case the same people used multiple aliases.

The development of guidelines for research on DSNs can, therefore, help to enhance the quality of DSN research in general.

#### 5.1.2. Studies with high external validity

Our data shows that many results regarding DSNs were obtained only on very few projects, i.e., over 69% of the publications used less than 11 projects to conduct their research. While this does not mean that the results are wrong or would not generalize to other contexts, this poses a threat to the generalizability of

results, as such small samples can only represent populations with a limited context. This problem is to some degree further aggravated, because there is an overlap in the data that is used, i.e., multiple studies using the same data, sometimes the same single project (e.g., IBM Jazz or the Global Studio Project). Moreover, we noted a strong relation between the data sources and the size of studies. Fig. 6 shows the size of the studies with relation to the data source. The larger circles mean more publications. Almost all publications with large numbers of people and projects were based on data from forges. Thus, an open issue considered for all future publications is to use larger sample sizes regarding the number of projects, to enable a better generalizability of results. This could either be done by harnessing data from forges or by collecting data for more projects from other data sources.

#### 5.1.3. Lack of replications

There is general lack of replications in DSN research. The publications are more or less independent of each other, the exception being multiple publications by the same authors building on each other. We did not find any study that explicitly tried to replicate prior results. The lack of replications is especially problematic due to the often very small numbers of projects considered (see above). Thus, we believe that replication studies on DSN research are required for all research directions so far.

#### 5.2. Open topics

Here, we discuss potential future directions of DSN research.

##### 5.2.1. Inter-company collaborations

Since more and more companies contribute to open source software and/or develop their own software products as open source, the collaboration between developers of competing companies becomes an important issue. If developers from competing organizations contribute to the same project, this could lead to issues within a project, that could be analyzed through DSNs, e.g., with respect to team formation, onboarding, collaboration problems, and even impacts on the socio-technical congruence of projects. Within our mapping study, we only discovered one publication in this direction (Teixeira et al., 2015).

##### 5.2.2. DSNs from multiple sources

The use of multiple sources for DSN studies allows a deeper analysis of developer communities. For example, how does the community on a ML differ from the community that can be



observed in pull request discussions or in an ITS? Can we infer something about onboarding of developers from their integration in different DSNs? Do projects that use an ITS and a ML exhibit different collaboration properties than projects that just use an ITS or a ML? What exactly is the temporal-spatial relationship between the DSN structures of different sources? Does research regarding the team formation of projects based on MLs yield the same results as research on team formation on ITSs? How does migration to a new ITS affect the community structure? All of these are currently open questions. Especially the comparison of DSNs that are based on different data sources has been neglected so far, with only a single publication that directly compares the DSN structure obtained from ITS data with that obtained from VCS data (Aljemabi and Wang, 2017).

### 5.2.3. Diversity in DSNs

The role of gender and other issues related to diversity is an important recent trend in current software engineering research. There is already one recent publication that touches on the relation between gender and DSNs (Catolino et al., 2019). We believe that insights into the question if and how gender or other diversity related features impact DSNs can give researchers and practitioners valuable insights that may help to make software engineering both more inclusive and more effective.

### 5.2.4. Applications using DSNs

The current literature on DSNs has a strong focus on understanding community structures and the implications of the community structure on issues like developer roles, team formation, and collaboration behavior. However, there are only relatively few actionable applications of DSNs. CodeBook (Begel et al., 2010) is a notable exception that demonstrates how DSNs can be used to improve the daily life of software developers. While other publications also study applications of DSNs, e.g., for defect prediction, failure prediction, or developer recommendations, they are mostly not accompanied by a tool that makes the research actionable for practitioners. The tool papers that we identified cover mostly the visualization of DSNs. While visualizations are a useful tool for the analysis of communities, they are not actionable applications of DSNs. We believe that research that produces actionable tools can have a big impact, e.g., on already considered issues like bug triage or developer recommendations.

### 5.2.5. Data sets

We only identified a single publication that published a DSN as data set. While there are other publications that are based on public data sets, e.g., the source forge dump (Van Antwerp and Madey, 2008), these data sets are not yet DSNs. They only contain the data necessary to create a DSN. While there are certainly use cases, in which new DSNs must be created, e.g., because different information is used to create links between developers, there are also cases for which dedicated data sets on DSNs would have advantages. For example, benchmark data sets could allow, e.g., to compare different approaches for developer recommendation or the identification of core developers. Moreover, the collection of data from a large amount of software repositories can be very time consuming. Data sets for a large amount of projects could help with this issue, and, e.g., enable larger studies with MLs as sources for projects.

## 6. Related work

Our systematic mapping study is not the first literature study that covers DSNs. Within this section, we discuss related literature studies on DSNs, their differences to our work, and how we utilized them as sanity checks for our work.

Closest to our work is the survey by Zhang et al. (2014a). Similar to our work, the authors analyzed the data sources, as well as topics that were addressed with DSNs. However, there are several notable differences between the work by Zhanget al. and our work. First, the search strategy by Zhanget al. is different from ours. They used the search term “developer network” and identified 20 publications related to DSNs within the first 50 hits on Google Scholar. Using these publications as seed, the authors performed one round of forward/backward snowballing and identified a total of 86 primary studies this way. Due to this limited search, the authors only identified a limited body of the relevant literature. In comparison, we use more search terms and multiple search engines, consider 750 instead of 50 hits per search term/search engine, and performed exhaustive backward and forward snowballing until no further papers were identified, and performed an exhaustive search with Scopus. Moreover, the focus of the presentation and focus of the work from Zhanget al. differs from ours. We provide a systematic mapping of approaches to topics through inductive coding, i.e., are interested in the general topics and trends that are analyzed. In comparison, Zhanget al. provide a more detailed description of different approaches to address research topics, but do not systematically map publications into different categories. However, we used the description of research topics described by Zhanget al. as starting point for our inductive coding, but identified additional topics, e.g., global software engineering, team formation, inter-company collaboration behavior, and developer onboarding. Another difference to our work is that Zhanget al. also report on the metrics that were used for the analysis of the DSN, an aspect that is not covered by our mapping study.

The literature study by Tamburri et al. (2013) uses grounded theory to identify different types of social structures within open source software development. Thus, their focus is different from ours, which is on DSNs in general, not on social structures. However, DSNs play an important role in the study by Tamburri et al. and are part of the literature that they identify. Our inductive coding approach is in principle similar to the approach used by Tamburri et al. in the sense that we inductively infer the relevant concepts from the identified publications. However, the goal was completely different, i.e., we wanted to identify research topics instead of types of organizational structures. Thus, while we identify broad topics of research, Tamburri et al. identified detailed information for the topic of social and collaboration structures. Due to the different focus, the search strategies also differ. Most importantly, the search by Tamburri et al. also covers search terms like “organizational”, “knowledge community” and similar to account for the different focus. Moreover, the search engines used are different from ours. They used Web of Science, EBSCO, JSTOR knowledge storage, Wiley InterScience and ProQuest in addition to the search engines we used. On the other hand, we used Google Scholar, which was not used by Tamburri et al.. The authors identified 143 publications for their study.

Manteli et al. (2012) performed a literature study to analyze DSNs with respect to global software development. Their focus was on coordination, cooperation, and communication aspects of global software development. This scope of this survey is narrower than our mapping study of DSNs without further restrictions. Thus, while we identify broad topics of research, Manteli et al. identified detailed information for the DSNs in global software engineering. This shows in the difference in search terms and inclusion criteria. Moreover, there is a difference in search engines used. Manteli et al. used EBSCO and Wiley InterScience in addition to the engines we used, but did not use Google Scholar. The authors identified 23 primary studies on DSNs with a relation to global software development.

Abufouda and Abukwaik (2017) performed a systematic literature review on DSNs with the goal to identify how reliable constructed social networks are. This goal is different from our general focus, which shows, e.g., in the different exclusion criteria. The authors used the same search engines we also used, with the exception of Google Scholar and SCOPUS which were not considered. The authors identify 23 primary studies that meet the criteria for their survey. The data the authors collected is very detailed with respect to the required description of the model and covers aspects like vertex types, edge types, and validation criteria. Thus, the work by Abufouda and Abuwaik focuses on evaluating aspects related to the internal validity of studies. In comparison, we collect data related to the external validity of DSN studies in our work, i.e., the scope of the analysis that is conducted.

In addition to our comparison with related work above, there are several differences between our work and all the related literature. No other work performed a bibliometric assessment of influential authors, papers, and venues or the trends over time. Moreover, no work in the literature provides information about the scope of the networks, i.e., the number of projects and participants that are analyzed through DSNs in a publication. Therefore, none of the prior literature studies provides answers to our research questions RQ3–RQ5. Additionally, as discussed above, none of the works provides a complete or systematic mapping of research topics which could be used to answer RQ1 or empirical data to support the answer for RQ2. Overall, our work goes well beyond the currently available literature studies on DSNs, both in terms of identified publications as well as due to the research questions we address.

## 7. Conclusion

This article presents the results of our systematic mapping study on DSNs. We identified 255 primary studies published between 2002 and 2020. Our results show that DSNs were used for the analysis of many different software engineering research topics since their initial use in the year 2002 (Madey et al., 2002), e.g., the analysis of community structures, the creation or improvement of prediction models, the study of collaboration behavior, and the identification of developer roles within projects. Moreover, we found that the social networks are often modeled based on data collected from a single repository, e.g., a forge like GitHub or SourceForge, version control systems like Git or SVN, issue tracking systems like Jira or Bugzilla, or mailing lists. A few notable exceptions use on-site observation centric techniques instead of data from software repositories. We observed a tendency that many publications only use a small set of sample projects, which may inhibit the generalizability of findings. Related to this is a general lack of replications within the body of work, i.e., we did not find a single replication study. Through a bibliometric assessment we found that there are many highly cited papers on DSNs on a diverse set of topics, which highlights the many use cases for DSNs in software engineering research. Our data shows that the interest in DSNs in research is still high, though there is a slight declining trend in recent years after the interest peaked in 2014.

Based on our findings, we suggest that future research addresses aspects that were neglected so far, e.g., inter-company collaborations (Teixeira et al., 2015), practical applications of DSNs (Begel et al., 2010), and the relationship between social network structures and diversity (Catolino et al., 2019). Moreover, we believe that replication studies can help to address the question if current results from the state of the art generalize beyond the often relatively small set of projects that were used in many publications.

## CRediT authorship contribution statement

**Steffen Herbold:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. **Aynur Amirfallah:** Formal analysis, Investigation, Writing - review & editing. **Fabian Trautsch:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Jens Grabowski:** Writing - review & editing, Supervision, Project administration.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Thomas, D.R., 2006. A general inductive approach for analyzing qualitative evaluation data. *Amer. J. Eval.* 27 (2), 237–246. <http://dx.doi.org/10.1177/1098214005283748>, arXiv:<https://doi.org/10.1177/1098214005283748>.
- Kitchenham, B., Charters, S., 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering (Version 2.3). Technical Report EBSE-2007-01, Keele Univ., EBSE.
- Van Antwerp, M., Madey, G., 2008. Advances in the sourceforge research data archive. In: Workshop on Public Data About Software Development (WoPDaSD) At the 4th International Conference on Open Source Systems, Milan, Italy, pp. 1–6.
- Sommerville, I., et al., 2011. Software Engineering. Pearson, Boston.
- Manteli, C., Van Vliet, H., Van Den Hooff, B., 2012. Adopting a social network perspective in global software development. In: Global Software Engineering (ICGSE), 2012 IEEE Seventh International Conference on. IEEE, pp. 124–133.
- Tamburri, D.A., Lago, P., Vliet, H.v., 2013. Organizational social structures for software engineering. *ACM Comput. Surv.* 46 (1), 3.
- Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. In: EASE '14, ACM, New York, NY, USA, pp. 38:1–38:10. <http://dx.doi.org/10.1145/2601248.2601268>, URL <http://doi.acm.org/10.1145/2601248.2601268>.
- Zhang, W., Nie, L., Jiang, H., 2014a. Developer social networks in software engineering: construction, analysis, and applications. *Sci. China Inf. Sci.* 57 (12), 1–23.
- Abufouda, M., Abukwaik, H., 2017. On using network science in mining developers collaboration in software engineering: A systematic literature review. *Int. J. Data Min. Knowl. Manage. Process* 7 (5/6), 1–20. <http://dx.doi.org/10.5121/ijdkp.2017.7601>.
- Abreu, R., Premraj, R., 2009. How developer communication frequency relates to bug introducing changes. In: Proceedings of the Joint International and Annual ERCIM Workshops on Principles of Software Evolution (IWPSE) and Software Evolution (Evol) Workshops. ACM, pp. 153–158.
- Ahuja, M.K., Galletta, D.F., Carley, K.M., 2003. Individual centrality and performance in virtual r&d groups: An empirical study. *Manage. Sci.* 49 (1), 21–38.
- Aljemabi, M.A., Wang, Z., 2017. Empirical study on the similarity and difference between vcs-dsn and bts-dsn. In: Proceedings of the 2017 International Conference on Management Engineering, Software Engineering and Service Sciences. ACM, pp. 30–37.
- Aljemabi, M.A., Wang, Z., 2018. Empirical study on the evolution of developer social networks. *IEEE Access* 6, 51049–51060.
- Allaho, M.Y., Lee, W.-C., 2013. Analyzing the social ties and structure of contributors in open source software community. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM, pp. 56–60.
- Amrit, C., Hillegersberg, J., Kumar, K., 2004. A social network perspective of Conway's law. In: Proceedings of the CSCW Workshop on Social Networks, Chicago, IL, USA.
- Avelino, G., Passos, L., Hora, A., Valente, M.T., 2017. Assessing code authorship: The case of the linux kernel. In: Balaguer, F., Di Cosmo, R., Garrido, A., Kon, F., Robles, G., Zacchiroli, S. (Eds.), Open Source Systems: Towards Robust Practices. Springer International Publishing, Cham, pp. 151–163.

- Avelino, G., Passos, L., Hora, A., Valente, M.T., 2019. Measuring and analyzing code authorship in 1+118 open source projects. *Sci. Comput. Program.* 176, 14–32. <http://dx.doi.org/10.1016/j.scico.2019.03.001>, URL <http://www.sciencedirect.com/science/article/pii/S0167642318300388>.
- Avritzer, A., Paulish, D., Cai, Y., Sethi, K., 2010. Coordination implications of software architecture in a global software development project. *J. Syst. Softw.* 83 (10), 1881–1895. <http://dx.doi.org/10.1016/j.jss.2010.05.070>, URL <http://www.sciencedirect.com/science/article/pii/S0164121210001330>.
- Bana, R., Arora, A., 2018. Influence indexing of developers, repositories, technologies and programming languages on social coding community github. In: 2018 Eleventh International Conference on Contemporary Computing (IC3). pp. 1–6. <http://dx.doi.org/10.1109/IC3.2018.8530644>.
- Banitaan, S., Alenezi, M., 2013. Decoba: Utilizing developers communities in bug assignment. In: Machine Learning and Applications (ICMLA), 2013 12th International Conference on, Vol. 2. IEEE, pp. 66–71.
- Batista, N.A., Brandão, M.A., Alves, G.B., da Silva, A.P.C., Moro, M.M., 2017. Collaboration strength metrics and analyses on github. In: Proceedings of the International Conference on Web Intelligence. In: WI '17, Association for Computing Machinery, New York, NY, USA, pp. 170–178. <http://dx.doi.org/10.1145/3106426.3106480>.
- Begel, A., Khoo, Y.P., Zimmermann, T., 2010. Codebook: discovering and exploiting relationships in software repositories. In: Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 1. ACM, pp. 125–134.
- Behfar, S.K., Behfar, Q., 2016. Intragroup density predicting intergroup tie strength within open-source-software collaboration network. In: Distributed Computing and Artificial Intelligence, 13th International Conference. Springer, pp. 165–173.
- Behfar, S.K., Turkina, E., Burger-Helmchen, T., 2018. Knowledge management in oss communities: Relationship between dense and sparse network structures. *Int. J. Inf. Manage.* 38 (1), 167–174.
- Bernardi, M.L., Canfora, G., Di Lucca, G.A., Di Penta, M., Distant, D., 2012. Do developers introduce bugs when they do not communicate? the case of eclipse and mozilla. In: Software Maintenance and Reengineering (CSMR), 2012 16th European Conference on. IEEE, pp. 139–148.
- Bettenburg, N., Hassan, A.E., 2010. Studying the impact of social structures on software quality. In: Program Comprehension (ICPC), 2010 IEEE 18th International Conference on. IEEE, pp. 124–133.
- Bettenburg, N., Hassan, A.E., 2013. Studying the impact of social interactions on software quality. *Empir. Softw. Eng.* 18 (2), 375–431.
- Bhattacharya, P., Iliofotou, M., Neamtii, I., Faloutsos, M., 2012a. Graph-based analysis and prediction for software evolution. In: Software Engineering (ICSE), 2012 34th International Conference on. IEEE, pp. 419–429.
- Bhattacharya, P., Neamtii, I., 2010. Fine-grained incremental learning and multi-feature tossing graphs to improve bug triaging. In: Software Maintenance (ICSM), 2010 IEEE International Conference on. IEEE, pp. 1–10.
- Bhattacharya, P., Neamtii, I., Faloutsos, M., 2014b. Determining developers. In: 2014 IEEE International Conference on Software Maintenance and Evolution (ICSME). IEEE, pp. 11–20.
- Bhattacharya, P., Neamtii, I., Shelton, C.R., 2012b. Automated, highly-accurate, bug assignment using machine learning and tossing graphs. *J. Syst. Softw.* 85 (10), 2275–2292. <http://dx.doi.org/10.1016/j.jss.2012.04.053>, automated software evolution. URL <http://www.sciencedirect.com/science/article/pii/S0164121212001240>.
- Bhowmik, T., Niu, N., Wang, W., Cheng, J.-R.C., Li, L., Cao, X., 2016. Optimal group size for software change tasks: a social information foraging perspective. *IEEE Trans. Cybern.* 46 (8), 1784–1795.
- Bianchini, D., De Antonellis, V., Melchiori, M., 2015. Developers' networks contribution to web application design. In: Proceedings of the 17th International Conference on Information Integration and Web-Based Applications & Services. In: iiWAS '15, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/2837185.2837241>.
- Bianchini, D., De Antonellis, V., Melchiori, M., 2016a. The role of developers' social relationships in improving service selection. *Int. J. Web Inf. Syst.*
- Bianchini, D., De Antonellis, V., Melchiori, M., 2016b. A social network-based framework for data services selection in modern web application design. In: CAISE Forum. pp. 73–80.
- Biçer, S., Bener, A.B., Çağlayan, B., 2011. Defect prediction using social network analysis on issue repositories. In: Proceedings of the 2011 International Conference on Software and Systems Process. ACM, pp. 63–71.
- Bird, C., Gourley, A., Devanbu, P., Gertz, M., Swaminathan, A., 2006a. Mining email social networks. In: Proceedings of the 2006 International Workshop on Mining Software Repositories. ACM, pp. 137–143.
- Bird, C., Gourley, A., Devanbu, P., Gertz, M., Swaminathan, A., 2006b. Mining email social networks in postgres. In: Proceedings of the 2006 International Workshop on Mining Software Repositories. ACM, pp. 185–186.
- Bird, C., Gourley, A., Devanbu, P., Swaminathan, A., Hsu, G., 2007. Open borders? immigration in open source projects. In: Mining Software Repositories, 2007. ICSE Workshops MSR'07. Fourth International Workshop on. IEEE, p. 6.
- Bird, C., Nagappan, N., Gall, H., Murphy, B., Devanbu, P., 2009. Putting it all together: Using socio-technical networks to predict failures. In: Software Reliability Engineering, 2009. ISSRE'09. 20th International Symposium on. IEEE, pp. 109–119.
- Bird, C., Pattison, D., D'Souza, R., Filkov, V., Devanbu, P., 2008. Latent social structure in open source projects. In: Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering. ACM, pp. 24–35.
- Borici, A., Blincoe, K., Schröter, A., Valetto, G., Damian, D., 2012. Proxiscientia: Toward real-time visualization of task and developer dependencies in collaborating software development teams. In: Proceedings of the 5th International Workshop on Co-Operative and Human Aspects of Software Engineering. IEEE Press, pp. 5–11.
- Çağlayan, B., Bener, A.B., 2016. Effect of developer collaboration activity on software quality in two large scale projects. *J. Syst. Softw.* 118, 288–296.
- Çağlayan, B., Bener, A.B., Miranskyy, A., 2013. Emergence of developer teams in the collaboration network. In: Cooperative and Human Aspects of Software Engineering (CHASE), 2013 6th International Workshop on. IEEE, pp. 33–40.
- Canfora, G., Cerulo, L., Cimitile, M., Di Penta, M., 2011. Social interactions around cross-system bug fixings: the case of freebsd and openbsd. In: Proceedings of the 8th Working Conference on Mining Software Repositories. ACM, pp. 143–152.
- Canfora, G., Di Penta, M., Oliveto, R., Panichella, S., 2012. Who is going to mentor newcomers in open source projects?. In: Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering. In: FSE '12, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/2393596.2393647>.
- Casaluovo, C., Vasilescu, B., Devanbu, P., Filkov, V., 2015. Developer onboarding in github: The role of prior social links and language experience. In: Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering. In: ESEC/FSE 2015, Association for Computing Machinery, New York, NY, USA, pp. 817–828. <http://dx.doi.org/10.1145/2786805.2786854>.
- Cataldo, M., Ehrlich, K., 2012. The impact of communication structure on new product development outcomes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 3081–3090.
- Cataldo, M., Herbsleb, J.D., 2008a. Communication networks in geographically distributed software development. In: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work. ACM, pp. 579–588.
- Cataldo, M., Herbsleb, J.D., 2008b. Communication patterns in geographically distributed software development and engineers' contributions to the development effort. In: Proceedings of the 2008 International Workshop on Cooperative and Human Aspects of Software Engineering. ACM, pp. 25–28.
- Cataldo, M., Herbsleb, J.D., Carley, K.M., 2008. Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity. In: Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. ACM, pp. 2–11.
- Cataldo, M., Wagstrom, P.A., Herbsleb, J.D., Carley, K.M., 2006. Identification of coordination requirements: implications for the design of collaboration and awareness tools. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work. ACM, pp. 353–362.
- Catolino, G., Palomba, F., Tamburri, D.A., Serebrenik, A., Ferrucci, F., 2019. Gender diversity and women in software teams: How do they affect community smells?. In: 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). pp. 11–20. <http://dx.doi.org/10.1109/ICSE-SEIS.2019.00010>.
- Chang, K.T., Ehrlich, K., 2007. Out of sight but not out of mind?: Informal networks, communication and media use in global software teams. In: Proceedings of the 2007 Conference of the Center for Advanced Studies on Collaborative Research. IBM Corp., pp. 86–97.
- Chen, L., Wang, X., Liu, C., 2010. Improving bug assignment with bug tossing graphs and bug similarities. In: 2010 International Conference on Biomedical Engineering and Computer Science. pp. 1–5. <http://dx.doi.org/10.1109/ICBECs.2010.5462287>.
- Cheng, C., Li, B., Li, Z.-Y., Zhao, Y.-Q., Liao, F.-L., 2017. Developer role evolution in open source software ecosystem: An explanatory study on gnome. *J. Comput. Sci. Tech.* 32 (2), 396–414.
- Cherry, S., Robillard, P.N., 2008. The social side of software engineering—A real ad hoc collaboration network. *Int. J. Hum.-Comput. Stud.* 66 (7), 495–505. <http://dx.doi.org/10.1016/j.ijhcs.2008.01.002>, collaborative and social aspects of software development. URL <http://www.sciencedirect.com/science/article/pii/S1071581908000141>.
- Cohen, E., Consens, M.P., 2018. Large-scale analysis of the co-commit patterns of the active developers in github's top repositories. In: 2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR). pp. 426–436.
- Conaldi, G., Rullani, F., 2010. The meso-level structure of f/oss collaboration network: local communities and their innovativeness. In: IFIP International Conference on Open Source Systems. Springer, pp. 42–52.
- Crowston, K., Howison, J., 2005. The social structure of free and open source software development. *First Monday* 10 (2).



- Crowston, K., Howison, J., 2006. Hierarchy and centralization in free and open source software team communications. *Knowl. Technol. Policy* 18 (4), 65–85.
- Crowston, K., Li, Q., Wei, K., Eseryel, U.Y., Howison, J., 2007. Self-organization of teams for free/libre open source software development. *Inf. Softw. Technol.* 49 (6), 564–575.
- Crowston, K., Wei, K., Li, Q., Howison, J., 2006. Core and periphery in free/libre and open source software team communications. In: *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, Vol. 6. IEEE, p. 118a.
- Damian, D., Helms, R., Kwan, I., Marczak, S., Koelewijn, B., 2013. The role of domain knowledge and cross-functional communication in socio-technical coordination. In: *Software Engineering (ICSE), 2013 35th International Conference on*. IEEE, pp. 442–451.
- Damian, D., Izquierdo, L., Singer, J., Kwan, I., 2007a. Awareness in the wild: Why communication breakdowns occur. In: *Global Software Engineering, 2007. ICGSE 2007. Second IEEE International Conference on*. IEEE, pp. 81–90.
- Damian, D., Marczak, S., Kwan, I., 2007b. Collaboration patterns and the impact of distance on awareness in requirements-centred social networks. In: *Requirements Engineering Conference, 2007. RE'07. 15th IEEE International. IEEE*, pp. 59–68.
- Datta, S., 2018. How does developer interaction relate to software quality? an examination of product development data. *Empir. Softw. Eng.* 23 (3), 1153–1187.
- Datta, S., Kaulgud, V., Sharma, V.S., Kumar, N., 2010. A social network based study of software team dynamics. In: *Proceedings of the 3rd India Software Engineering Conference. ACM*, pp. 33–42.
- Datta, S., Sindhgatta, R., Sengupta, B., 2011. Evolution of developer collaboration on the jazz platform: a study of a large scale agile project. In: *Proceedings of the 4th India Software Engineering Conference. ACM*, pp. 21–30.
- De Souza, C., Dourish, P., Redmiles, D., Quirk, S., Trainer, E., 2004. From technical dependencies to social dependencies. In: *Workshop on Social Networks for Design and Analysis: Using Network Information in CSCW*.
- De Souza, C., Froehlich, J., Dourish, P., 2005. Seeking the source: software source code as a social and technical artifact. In: *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work. ACM*, pp. 197–206.
- De Souza, C.R., Hildenbrand, T., Redmiles, D., 2007. Toward visualization and analysis of traceability relationships in distributed and offshore software development projects. In: *International Conference on Software Engineering Approaches for Offshore and Outsourced Development. Springer*, pp. 182–199.
- Dittrich, A., Gunes, M.H., Dascalu, S., 2013. Network analysis of software repositories: identifying subject matter experts. In: *Complex Networks. Springer*, pp. 187–198.
- Dráždilová, P., Babskova, A., Martinovič, J., Slaninová, K., Minks, Š., 2012. Method for identification of suitable persons in collaborators' networks. In: *IFIP International Conference on Computer Information Systems and Industrial Management. Springer*, pp. 101–110.
- Duc, A.N., Cruzes, D.S., Ayala, C., Conradi, R., 2011. Impact of stakeholder type and collaboration on issue resolution time in oss projects. In: *IFIP International Conference on Open Source Systems. Springer*, pp. 1–16.
- Ducheneaut, N., 2005. Socialization in an open source software community: A socio-technical analysis. *Comput. Support. Coop. Work (CSCW)* 14 (4), 323–368.
- Ehrlich, K., Cataldo, M., 2012. All-for-one and one-for-all?: a multi-level analysis of communication patterns and individual performance in geographically distributed software development. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. ACM*, pp. 945–954.
- Ehrlich, K., Chang, K., 2006. Leveraging expertise in global software teams: Going outside boundaries. In: *Global Software Engineering, 2006. ICGSE'06. International Conference on*. IEEE, pp. 149–158.
- El Asri, I., Kerzazi, N., Benhiba, L., Janati, M., 2017. From periphery to core: A temporal analysis of github contributors' collaboration network. In: *Working Conference on Virtual Enterprises. Springer*, pp. 217–229.
- Ell, J., 2013. Identifying failure inducing developer pairs within developer networks. In: *Proceedings of the 2013 International Conference on Software Engineering. IEEE Press*, pp. 1471–1473.
- Feczak, S., Hossain, L., 2009. Measuring coordination gaps of open source groups through social networks. In: *ICEIS (5)*. pp. 84–90.
- Feczak, S., Hossain, L., 2011. Exploring computer supported collaborative coordination through social networks. *J. High Technol. Manage. Res.* 22 (2), 121–140.
- Fonseca, S.B., De Souza, C.R., Redmiles, D.F., 2006. Exploring the relationship between dependencies and coordination to support global software development projects. In: *Null. IEEE*, p. 243.
- Gao, Y., Madey, G., 2007a. Network analysis of the sourceforge. net community. limerick, Ireland: sn. In: *International Conference on Open Source Systems*.
- Gao, Y., Madey, G., 2007b. Towards understanding: a study of the sourceforge. net community using modeling and simulation. In: *Proceedings of the 2007 Spring Simulation Multiconference-Volume 2. Society for Computer Simulation International*, pp. 145–150.
- Geipel, M.M., Press, K., Schweitzer, F., 2014. Communication in innovation communities: an analysis of 100 open source software projects. *Adv. Complex Syst.* 17 (07n08), 1550006. <http://dx.doi.org/10.1142/S021952591550006X>. arXiv:<https://doi.org/10.1142/S021952591550006X>.
- Gharehyazie, M., Filkov, V., 2017. Tracing distributed collaborative development in apache software foundation projects. *Empir. Softw. Eng.* 22 (4), 1795–1830.
- Gharehyazie, M., Posnett, D., Filkov, V., 2013. Social activities rival patch submission for prediction of developer initiation in oss projects. In: *Software Maintenance (ICSM), 2013 29th IEEE International Conference on*. IEEE, pp. 340–349.
- Gharehyazie, M., Posnett, D., Vasilescu, B., Filkov, V., 2015. Developer initiation and social interactions in oss: A case study of the apache software foundation. *Empir. Softw. Eng.* 20 (5), 1318–1353.
- Gilbert, E., Karahalios, K., 2007. Codesaw: A social visualization of distributed software development. In: *IFIP Conference on Human-Computer Interaction. Springer*, pp. 303–316.
- Gloor, P.A., Laubacher, R., Dynes, S.B., Zhao, Y., 2003. Visualization of communication patterns in collaborative innovation networks-analysis of some w3c working groups. In: *Proceedings of the Twelfth International Conference on Information and Knowledge Management. ACM*, pp. 56–60.
- González-Barahona, J.M., Lopez, L., Robles, G., 2004. Community structure of modules in the apache project. In: *Proceedings of 4th Workshop on Open Source Software Engineering. IET*.
- Gote, C., Scholtes, I., Schweitzer, F., 2019. Git2net - mining time-stamped co-editing networks from large git repositories. In: *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*. pp. 433–444. <http://dx.doi.org/10.1109/MSR.2019.00070>.
- Hahn, J., Moon, J.Y., Zhang, C., 2006. Impact of social ties on open source project team formation. In: *IFIP International Conference on Open Source Systems. Springer*, pp. 307–317.
- Hahn, J., Moon, J.Y., Zhang, C., 2008. Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. *Inf. Syst. Res.* 19 (3), 369–391.
- Hajiakhond Bidoki, N., Sukthankar, G., 2018. Network semantic segmentation with application to github. In: *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*. pp. 1281–1284. <http://dx.doi.org/10.1109/CSCI46756.2018.00247>.
- Hannemann, A., Klammar, R., 2013. Community dynamics in open source software projects: Aging and social reshaping. In: *IFIP International Conference on Open Source Systems. Springer*, pp. 80–96.
- He, P., Li, B., Huang, Y., 2012. Applying centrality measures to the behavior analysis of developers in open source software community. In: *Cloud and Green Computing (CGC), 2012 Second International Conference on*. IEEE, pp. 418–423.
- Herbold, S., Amirfallah, A., Trautsch, F., Grabowski, J., 2020. Sherbold/replication-kit-jss-2020-dsn-mapping-study: release of the replication kit. <http://dx.doi.org/10.5281/zenodo.3996934>.
- Herbsleb, J.D., Mockus, A., 2003. An empirical study of speed and communication in globally distributed software development. *IEEE Trans. Softw. Eng.* 29 (6), 481–494.
- Hinds, P., McGrath, C., 2006. Structures that work: social structure, work structure and coordination ease in geographically distributed teams. In: *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work. ACM*, pp. 343–352.
- Hong, Q., Kim, S., Cheung, S.C., Bird, C., 2011. Understanding a developer social network and its evolution. In: *Software Maintenance (ICSM), 2011 27th IEEE International Conference on*. IEEE, pp. 323–332.
- Honsel, V., Honsel, D., Grabowski, J., 2014. Software process simulation based on mining software repositories. In: *2014 IEEE International Conference on Data Mining Workshop*. pp. 828–831. <http://dx.doi.org/10.1109/ICDMW.2014.35>.
- Honsel, V., Honsel, D., Grabowski, J., Waack, S., 2015. Developer oriented and quality assurance based simulation of software processes. In: *SATToSE*. pp. 5–13.
- Honsel, V., Honsel, D., Herbold, S., Grabowski, J., Waack, S., 2015. Mining software dependency networks for agent-based simulation of software evolution. In: *2015 30th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW)*. pp. 102–108. <http://dx.doi.org/10.1109/ASEW.2015.9>.
- Hossain, L., Wu, A., Chung, K.K., 2006. Actor centrality correlates to project based coordination. In: *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work. ACM*, pp. 363–372.
- Hossain, L., Zhou, D., 2008. Measuring oss quality through centrality. In: *Proceedings of the 2008 International Workshop on Cooperative and Human Aspects of Software Engineering. ACM*, pp. 65–68.
- Hossain, L., Zhu, D., 2009. Social networks and coordination performance of distributed software development teams. *J. High Technol. Manage. Res.* 20 (1), 52–61.
- Howison, J., Inoue, K., Crowston, K., 2006. Social dynamics of free and open source team communications. In: *IFIP International Conference on Open Source Systems. Springer*, pp. 319–330.



- Hu, Y., Wang, S., Ren, Y., Choo, K.-K.R., 2018. User influence analysis for github developer social networks. *Expert Syst. Appl.* 108, 108–118.
- Hu, W., Wong, K., 2013. Using citation influence to predict software defects. In: *Mining Software Repositories (MSR)*, 2013 10th IEEE Working Conference on. IEEE, pp. 419–428.
- Hu, Y., Zhang, J., Bai, X., Yu, S., Yang, Z., 2016. Influence analysis of github repositories. *SpringerPlus* 5 (1), 1–19.
- Hu, D., Zhao, J.L., 2008. A comparison of evaluation networks and collaboration networks in open source software communities. In: *AMCIS 2008 Proceedings*. p. 277.
- Hu, D., Zhao, J., Cheng, J., 2012. Reputation management in an open source developer social network: An empirical study on determinants of positive evaluations. *Decis. Support Syst.* 53 (3), 526–533.
- Huang, K., Choi, N., 2011. Relating and clustering free/libre open source software projects and developers: A social network perspective. In: *2011 44th Hawaii International Conference on System Sciences*. pp. 1–10. <http://dx.doi.org/10.1109/HICSS.2011.359>.
- Huang, S.-K., Liu, K.-m., 2005. Mining version histories to verify the learning process of legitimate peripheral participants. *ACM SIGSOFT Softw. Eng. Notes* 30 (4), 1–5.
- Ichimura, T., Uemoto, T., 2015. Analysis of the social community based on the network growing model in open source software community. In: *Computational Intelligence and Applications (IWCI/A)*, 2015 IEEE 8th International Workshop on. IEEE, pp. 149–153.
- Iyer, D.G., Lyytinen, K., 2019. Requirements engineering (re) effectiveness in open source software: the role of social network configurations and requirements properties. *Requir. Eng. (RE)* 5, 15–2019.
- Jarczyk, O., Jaroszewicz, S., Wierzbicki, A., Pawlak, K., Jankowski-Lorek, M., 2018. Surgical teams on github: Modeling performance of github project development processes. *Inf. Softw. Technol.* 100, 32–46. <http://dx.doi.org/10.1016/j.infsof.2018.03.010>, URL <http://www.sciencedirect.com/science/article/pii/S095058491730304X>.
- Jeong, G., Kim, S., Zimmermann, T., 2009. Improving bug triage with bug tossing graphs. In: *Proceedings of the 7th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering*. ACM, pp. 111–120.
- Jermakovics, A., Sillitti, A., Succi, G., 2011. Mining and visualizing developer networks from version control systems. In: *Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering*. ACM, pp. 24–31.
- Jermakovics, A., Sillitti, A., Succi, G., 2013. Exploring collaboration networks in open-source projects. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 97–108.
- Jiang, J., Zhang, L., Li, L., 2013. Understanding project dissemination on a social coding site. In: *2013 20th Working Conference on Reverse Engineering (WCORE)*. pp. 132–141. <http://dx.doi.org/10.1109/WCORE.2013.6671288>.
- Joblin, M., Apel, S., Hunsen, C., Mauerer, W., 2017a. Classifying developers into core and peripheral: An empirical study on count and network metrics. In: *Software Engineering (ICSE)*, 2017 IEEE/ACM 39th International Conference on. IEEE, pp. 164–174.
- Joblin, M., Apel, S., Mauerer, W., 2017b. Evolutionary trends of developer coordination: A network approach. *Empir. Softw. Eng.* 22 (4), 2050–2094.
- Joblin, M., Mauerer, W., Apel, S., Siegmund, J., Riehle, D., 2015. From developer networks to verified communities: a fine-grained approach. In: *Proceedings of the 37th International Conference on Software Engineering—Volume 1*. IEEE Press, pp. 563–573.
- Kakimoto, T., Kamei, Y., Ohira, M., Matsumoto, K., 2006. Social network analysis on communications for knowledge collaboration in oss communities. In: *Proceedings of the International Workshop on Supporting Knowledge Collaboration in Software Development (KCS'D'06)*. Citeseer, pp. 35–41.
- Kamei, Y., Matsumoto, S., Maeshima, H., Onishi, Y., Ohira, M., Matsumoto, K.-i., 2008. Analysis of coordination between developers and users in the apache community. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 81–92.
- Kavaler, D., Filkov, V., 2017. Stochastic actor-oriented modeling for studying homophily and social influence in oss projects. *Empir. Softw. Eng.* 22 (1), 407–435.
- Kerzazi, N., El Asri, I., 2016. Who can help to review this piece of code?. In: *Working Conference on Virtual Enterprises*. Springer, pp. 289–301.
- Kerzazi, N., El Asri, I., 2017. Knowledge flows within open source software projects: A social network perspective. In: *Advances in Ubiquitous Networking 2*. Springer, pp. 247–258.
- Kidane, Y.H., Gloor, P.A., 2007. Correlating temporal communication patterns of the eclipse open source community with performance and creativity. *Comput. Math. Organ. Theory* 13 (1), 17–27.
- Kumar, A., Desai, Y., Gandhi, M., Agarwal, S., 2019. Studying multifaceted collaboration of oss developers and its impact on their bug fixing performance.
- Kumar, A., Gupta, A., 2013. Evolution of developer social network and its impact on bug fixing process. In: *Proceedings of the 6th India Software Engineering Conference*. ACM, pp. 63–72.
- Kwan, I., Schroter, A., Damian, D., 2011. Does socio-technical congruence have an effect on software build success? a study of coordination in a software project. *IEEE Trans. Softw. Eng.* 37 (3), 307–324.
- Lee, M.J., Ferwerda, B., Choi, J., Hahn, J., Moon, J.Y., Kim, J., 2013. Github developers use rockstars to overcome overflow of news. In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 133–138.
- Leibzon, W., 2016. Social network of software development at github. In: *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, pp. 1374–1376.
- Li, N., Mo, W., Shen, B., 2016. Task recommendation with developer social network in software crowdsourcing. In: *Software Engineering Conference (APSEC)*, 2016 23rd Asia-Pacific. IEEE, pp. 9–16.
- Licorish, S.A., MacDonell, S.G., 2014. Understanding the attitudes, knowledge sharing behaviors and task performance of core developers: A longitudinal study. *Inf. Softw. Technol.* 56 (12), 1578–1596.
- Licorish, S.A., MacDonell, S.G., 2015. Communication and personality profiles of global software developers. *Inf. Softw. Technol.* 64, 113–131.
- Licorish, S.A., MacDonell, S.G., 2017. Exploring software developers' work practices: Task differences, participation, engagement, and speed of task resolution. *Inf. Manage.* 54 (3), 364–382.
- Lim, S.L., Bentley, P.J., 2011. Evolving relationships between social networks and stakeholder involvement in software projects. In: *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*. ACM, pp. 1899–1906.
- Lim, S.L., Finkelstein, A., 2012. Stakerare: using social networks and collaborative filtering for large-scale requirements elicitation. *IEEE Trans. Softw. Eng.* 38 (3), 707–735.
- Lim, S.L., Quercia, D., Finkelstein, A., 2010. Stakenet: using social networks to analyse the stakeholders of large-scale software projects. In: *Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering—Volume 1*. ACM, pp. 295–304.
- Lima, A., Rossi, L., Musolesi, M., 2014. Coding together at scale: Github as a collaborative social network. In: *Eighth International AAAI Conference on Weblogs and Social Media*.
- Linäker, J., Regnell, B., Damian, D., 2019. A method for analyzing stakeholders' influence on an open source software ecosystem's requirements engineering process. *Requir. Eng.* 1–16.
- Liu, X., Iyer, B., 2007. Design architecture, developer networks and performance of open source software projects. In: *ICIS 2007 Proceedings*. p. 90.
- Long, Y., Siau, K., 2007. Social network structures in open source software development teams. *J. Database Manage. (JDM)* 18 (2), 25–40.
- López-Fernández, L., Robles, G., Gonzalez-Barahona, J.M., Herraiz, I., 2008. Applying social network analysis techniques to community-driven libre software projects. *Int. J. Inf. Technol. Web Eng.* 1, 28–50.
- Lopez-Fernandez, L., Robles, G., Gonzalez-Barahona, J.M., et al., 2004. Applying social network analysis to the information in cvs repositories. In: *International Workshop on Mining Software Repositories*. IET, pp. 101–105.
- MacLean, A.C., Knutson, C.D., 2013. Apache commits: social network dataset. In: *Mining Software Repositories (MSR)*, 2013 10th IEEE Working Conference on. IEEE, pp. 135–138.
- MacLean, A.C., Pratt, L.J., Knutson, C.D., Ringger, E.K., 2011. Knowledge homogeneity and specialization in the apache http server project. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 106–122.
- Madey, G., Freeh, V., Tynan, R., 2002. The open source software development phenomenon: An analysis based on social network theory. In: *AMCIS 2002 Proceedings*. p. 247.
- Marczak, S., Damian, D., Stege, U., Schröter, A., 2008. Information brokers in requirement-dependency social networks. In: *International Requirements Engineering*, 2008. RE'08. 16th IEEE. IEEE, pp. 53–62.
- McDonald, D.W., 2003. Recommending collaboration with social networks: A comparative evaluation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. In: CHI '03, Association for Computing Machinery, New York, NY, USA, pp. 593–600. <http://dx.doi.org/10.1145/642611.642714>.
- Meneely, A., Corcoran, M., Williams, L., 2010. Improving developer activity metrics with issue tracking annotations. In: *Proceedings of the 2010 ICSE Workshop on Emerging Trends in Software Metrics*. ACM, pp. 75–80.
- Meneely, A., Williams, L., 2009. Secure open source collaboration: an empirical study of linux' law. In: *Proceedings of the 16th ACM Conference on Computer and Communications Security*. ACM, pp. 453–462.
- Meneely, A., Williams, L., 2010a. On the use of issue tracking annotations for improving developer activity metrics. *Adv. Softw. Eng.* 2010.
- Meneely, A., Williams, L., 2010b. Strengthening the empirical analysis of the relationship between linux' law and software security. In: *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. In: ESEM '10, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/1852786.1852798>.
- Meneely, A., Williams, L., 2011. Socio-technical developer networks: Should we trust our measurements?. In: *Software Engineering (ICSE)*, 2011 33rd International Conference on. IEEE, pp. 281–290.

- Meneely, A., Williams, L., Snipes, W., Osborne, J., 2008. Predicting failures with developer networks and social network analysis. In: *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM, pp. 13–23.
- Mergel, I., 2015. Open collaboration in the public sector: The case of social coding on github. *Gov. Inf. Q.* 32 (4), 464–472.
- Mikawa, S.P., Cunningham, S.K., Gaskins, S.A., 2009. Removing barriers to trust in distributed teams: understanding cultural differences and strengthening social ties. In: *Proceedings of the 2009 International Workshop on Intercultural Collaboration*. ACM, pp. 273–276.
- Miransky, A., Caglayan, B., Bener, A., Cialini, E., 2014. Effect of temporal collaboration network, maintenance activity, and experience on defect exposure. In: *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM, p. 27.
- Mockus, A., 2010. Organizational volatility and its effects on software defects. In: *Proceedings of the Eighteenth ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM, pp. 117–126.
- Nakakoji, K., Yamada, K., Giaccardi, E., 2005. Understanding the nature of collaboration in open-source software development. In: *Software Engineering Conference, 2005. APSEC'05. 12th Asia-Pacific*. IEEE, pp. 8–pp.
- Ngamkajornwiwat, K., Zhang, D., Koru, A.G., Zhou, L., Nolker, R., 2008. An exploratory study on the evolution of oss developer communities. In: *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*. IEEE, p. 305.
- Nguyen, T., Wolf, T., Damian, D., 2008. Global software development and delay: Does distance still matter?. In: *Global Software Engineering, 2008. ICGSE 2008. IEEE International Conference on*. IEEE, pp. 45–54.
- Nia, R., Bird, C., Devanbu, P., Filkov, V., 2010. Validity of network analyses in open source projects. In: *Mining Software Repositories (MSR), 2010 7th IEEE Working Conference on*. IEEE, pp. 201–209.
- Nzeko'o, A.J.N., Latapy, M., Tchuente, M., 2015. Social network analysis of developers' and users' mailing lists of some free open source software. In: *Big Data (BigData Congress), 2015 IEEE International Congress on*. IEEE, pp. 728–732.
- Ogawa, M., Ma, K.-L., Bird, C., Devanbu, P., Gourley, A., 2007. Visualizing social interaction in open source software projects. In: *Visualization, 2007. APVIS'07. 2007 6th International Asia-Pacific Symposium on*. IEEE, pp. 25–32.
- Ohira, M., Ohoka, T., Kakimoto, T., Ohsugi, N., Matsumoto, K.-i., 2005a. Supporting knowledge collaboration using social networks in a large-scale online community of software development projects. In: *Software Engineering Conference, 2005. APSEC'05. 12th Asia-Pacific*. IEEE, 6–pp.
- Ohira, M., Ohsugi, N., Ohoka, T., Matsumoto, K.-i., 2005b. Accelerating cross-project knowledge collaboration using collaborative filtering and social networks. In: *ACM SIGSOFT Software Engineering Notes, Vol. 30*. ACM, pp. 1–5.
- Omoronyia, I., Ferguson, J., Roper, M., Wood, M., 2009. Using developer activity data to enhance awareness during collaborative software development. *Comput. Support. Coop. Work (CSCW)* 18 (5–6), 509.
- Orsila, H., Geldenhuys, J., Ruokonen, A., Hammouda, I., 2009. Trust issues in open source software development. In: *Proceedings of the Warm Up Workshop for ACM/IEEE ICSE 2010*. In: *WUP '09, Association for Computing Machinery*. New York, NY, USA, pp. 9–12. <http://dx.doi.org/10.1145/1527033.1527037>.
- Ortu, M., Destefanis, G., Kassab, M., Marchesi, M., 2015. Measuring and understanding the effectiveness of jira developers communities. In: *Proceedings of the Sixth International Workshop on Emerging Trends in Software Metrics*. IEEE Press, pp. 3–10.
- Palomba, F., Tamburri, D.A.A., Arcelli Fontana, F., Oliveto, R., Zaidman, A., Serebrenik, A., 2018. Beyond technical aspects: How do community smells influence the intensity of code smells?. *IEEE Trans. Softw. Eng.* 1. <http://dx.doi.org/10.1109/TSE.2018.2883603>.
- Panichella, S., Bavota, G., Di Penta, M., Canfora, G., Antoniol, G., 2014a. How developers' collaborations identified from different sources tell us about code changes. In: *Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on*. IEEE, pp. 251–260.
- Panichella, S., Canfora, G., Di Penta, M., Oliveto, R., 2014b. How the evolution of emerging collaborations relates to code changes: an empirical study. In: *Proceedings of the 22nd International Conference on Program Comprehension*. ACM, pp. 177–188.
- Peng, G., 2019. Co-membership, networks ties, and knowledge flow: An empirical investigation controlling for alternative mechanisms. *Decis. Support Syst.* 118, 83–90.
- Peng, G., Yu, F.G., Peng, Y., 2018. Co-membership, networks ties, and oss success: An investigation controlling for alternative mechanisms for knowledge flow.
- Pinzger, M., Nagappan, N., Murphy, B., 2008. Can developer-module networks predict failures?. In: *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM, pp. 2–12.
- Pohl, M., Diehl, S., 2008. What dynamic network metrics can tell us about developer roles. In: *Proceedings of the 2008 International Workshop on Cooperative and Human Aspects of Software Engineering*. ACM, pp. 81–84.
- Qiu, H.S., Nolte, A., Brown, A., Serebrenik, A., Vasilescu, B., 2019. Going farther together: The impact of social capital on sustained participation in open source. In: *ICSE*. IEEE.
- Robertson, J., Hann, I.-H., Slaughter, S., 2006. Communication networks in an open source software project. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 297–306.
- Rossi, R., Neville, J., 2010. Modeling the evolution of discussion topics and communication to improve relational classification. In: *Proceedings of the First Workshop on Social Media Analytics*. In: *SOMA '10, Association for Computing Machinery*. New York, NY, USA, pp. 89–97. <http://dx.doi.org/10.1145/1964858.1964871>.
- dos Santos, T.A., de Araujo, R.M., Magdaleno, A.M., 2011. Bringing out collaboration in software development social networks. In: *Proceedings of the 12th International Conference on Product Focused Software Development and Process Improvement*. ACM, pp. 18–21.
- Sapkota, H., Murukannaiah, P.K., Wang, Y., 2020. A network-centric approach for estimating trust between open source software developers. *PLOS ONE* 14 (12), 1–30. <http://dx.doi.org/10.1371/journal.pone.0226281>.
- Sarker, S., Ahuja, M., Sarker, S., Kirkeby, S., 2011. The role of communication and trust in global virtual teams: A social network perspective. *J. Manage. Inf. Syst.* 28 (1), 273–310.
- Sarker, S., Sarker, S., Kirkeby, S., Chakraborty, S., Path to “stardom” in globally distributed hybrid teams: An examination of a knowledge-centered perspective using social network analysis. *Decis. Sci.* 42 (2), 339–370. [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5915.2011.00314.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5915.2011.00314.x), URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.2011.00314.x>, <http://dx.doi.org/10.1111/j.1540-5915.2011.00314.x>.
- Sarma, A., Maccherone, L., Wagstrom, P., Herbsleb, J., 2009. Tesseract: Interactive visual exploration of socio-technical relationships in software development. In: *Software Engineering, 2009. ICSE 2009. IEEE 31st International Conference on*. IEEE, pp. 23–33.
- Schröter, A., 2010. Predicting build outcome with developer interaction in jazz. In: *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering—Volume 2*. ACM, pp. 511–512.
- Schwind, M., Schenk, A., Schneider, M., 2010. A tool for the analysis of social networks in collaborative software development. In: *2010 43rd Hawaii International Conference on System Sciences*. IEEE, pp. 1–10.
- Schwind, M., Wegmann, C., 2008. Synnat: Measuring collaboration in software development networks. In: *E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services, 2008 10th IEEE Conference on*. IEEE, pp. 97–104.
- Schwind, M., Wegmann, C., Wendt, O., 2008. Unveiling collaboration structures in software development projects. In: *ECIS*. pp. 1488–1499.
- Sharma, V.S., Kaulgud, V., 2011. Studying team evolution during software testing. In: *Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering*. ACM, pp. 72–75.
- Sharma, P.N., Savarimuthu, B.T.R., Stanger, N., 2017. Boundary spanners in open source software development: A study of python email archives. In: *Asia-Pacific Software Engineering Conference (APSEC), 2017 24th*. IEEE, pp. 308–317.
- Simpson, B., 2013. Changeset based developer communication to detect software failures. In: *Software Engineering (ICSE), 2013 35th International Conference on*. IEEE, pp. 1468–1470.
- Singh, P.V., 2010. The small-world effect: The influence of macro-level properties of developer collaboration networks on open-source project success. *ACM Trans. Softw. Eng. Methodol. (TOSEM)* 20 (2), 6.
- Singh, P.V., Tan, Y., 2010. Developer heterogeneity and formation of communication networks in open source software projects. *J. Manage. Inf. Syst.* 27 (3), 179–210.
- Singh, P.V., Tan, Y., Mookerjee, V., 2011. Network effects: The influence of structural capital on open source project success. *MIS Q.* 35 (4), 813–829, URL <http://www.jstor.org/stable/41409962>.
- de Souza, C.R., Quirk, S., Trainer, E., Redmiles, D.F., 2007. Supporting collaborative software development through the visualization of socio-technical dependencies. In: *Proceedings of the 2007 International ACM Conference on Supporting Group Work*. ACM, pp. 147–156.
- Sowe, S.K., Cerone, A., Settas, D., 2014. An empirical study of foss developers patterns of contribution: challenges for data linkage and analysis. *Sci. Comput. Program.* 91, 249–265.
- Sowe, S., Stamelos, I., Angelis, L., 2006. Identifying knowledge brokers that yield software engineering knowledge in oss projects. *Inf. Softw. Technol.* 48 (11), 1025–1033.
- Spinellis, D., 2006. Global software development in the freebsd project. In: *Proceedings of the 2006 International Workshop on Global Software Development for the Practitioner*. ACM, pp. 73–79.
- Sun, X., Yang, H., Xia, X., Li, B., 2017. Enhancing developer recommendation with supplementary information via mining historical commits. *J. Syst. Softw.* 134, 355–368. <http://dx.doi.org/10.1016/j.jss.2017.09.021>, URL <http://www.sciencedirect.com/science/article/pii/S0164121217302091>.

- Sureka, A., Goyal, A., Rastogi, A., 2011. Using social network analysis for mining collaboration data in a defect tracking system for risk and vulnerability analysis. In: *Proceedings of the 4th India Software Engineering Conference*. ACM, pp. 195–204.
- Surian, D., Liu, N., Lo, D., Tong, H., Lim, E.-P., Faloutsos, C., 2011. Recommending people in developers' collaboration network. In: *Reverse Engineering (WCRE), 2011 18th Working Conference on*. IEEE, pp. 379–388.
- Surian, D., Lo, D., Lim, E.-P., 2010. Mining collaboration patterns from a large developer network. In: *Reverse Engineering (WCRE), 2010 17th Working Conference on*. IEEE, pp. 269–273.
- Surian, D., Tian, Y., Lo, D., Cheng, H., LIM, E.P., 2013. Predicting Project Outcome Leveraging Socio-Technical Network Patterns. *IEEE Computer Society*.
- Syed, M.M., Hammouda, I., 2013. Socio-technical congruence in oss projects: Exploring conway's law in freesb. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 109–126.
- Tamburri, D.A., Palomba, F., Serebrenik, A., Zaidman, A., 2019. Discovering community patterns in open-source: a systematic approach and its evaluation. *Empir. Softw. Eng.* 24 (3), 1369–1417.
- Tan, Y., Mookerjee, V., Singh, P., 2007. Social capital, structural holes and team composition: Collaborative networks of the open source software community. In: *ICIS 2007 Proceedings*. p. 155.
- Teixeira, J., Robles, G., González-Barahona, J.M., 2015. Lessons learned from applying social network analysis on an industrial free/libre/open source software ecosystem. *J. Internet Serv. Appl.* 6 (1), 14.
- Thung, F., Bissyande, T.F., Lo, D., Jiang, L., 2013. Network structure of social coding in github. In: *Software Maintenance and Reengineering (Csmr), 2013 17th European Conference on*. IEEE, pp. 323–326.
- Toral, S.L., Martínez-Torres, M.d.R., Barrero, F., 2010. Analysis of virtual communities supporting oss projects using social network analysis. *Inf. Softw. Technol.* 52 (3), 296–303.
- Tymchuk, Y., Mocci, A., Lanza, M., 2014. Collaboration in open-source projects: Myth or reality?. In: *Proceedings of the 11th Working Conference on Mining Software Repositories*. ACM, pp. 304–307.
- Urdangarin, R., Fernandes, P., Avritzer, A., Paulish, D., 2008. Experiences with agile practices in the global studio project. In: *Global Software Engineering, 2008. ICGSE 2008*. IEEE International Conference on. IEEE, pp. 77–86.
- Valetto, G., Helander, M., Ehrlich, K., Chulani, S., Wegman, M., Williams, C., 2007. Using software repositories to investigate socio-technical congruence in development projects. In: *Proceedings of the Fourth International Workshop on Mining Software Repositories*. IEEE Computer Society, p. 25.
- Van Antwerp, M., Madey, G., 2010. The importance of social network structure in the open source software developer community. In: *2010 43rd Hawaii International Conference on System Sciences*. pp. 1–10. <http://dx.doi.org/10.1109/HICSS.2010.385>.
- Van Antwerp, M., Madey, G., 2010a. The importance of social network structure in the open source software developer community. In: *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*. IEEE, pp. 1–10.
- Van Antwerp, M., Madey, G., 2010b. Open source software developer and project networks. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 407–412.
- Wagstrom, P., Herbsleb, J., Carley, K., 2005. A social network approach to free/open source software simulation. In: *Proceedings First International Conference on Open Source Systems*. pp. 16–23.
- Wan, Y., Chen, L., Xu, G., Zhao, Z., Tang, J., Wu, J., 2018. Scsminer: mining social coding sites for software developer recommendation with relevance propagation. *World Wide Web* 21 (6), 1523–1543.
- Wang, J., 2012. Survival factors for free open source software projects: A multi-stage perspective. *Eur. Manage. J.* 30 (4), 352–371. <http://dx.doi.org/10.1016/j.emj.2012.03.001>, URL <http://www.sciencedirect.com/science/article/pii/S0263237312000199>.
- Wang, D., Cao, J., Qian, S., Qi, Q., 2019. Investigating cross-repository socially connected teams on github. In: *2019 26th Asia-Pacific Software Engineering Conference (APSEC)*. pp. 490–497. <http://dx.doi.org/10.1109/APSEC48747.2019.00072>.
- Wang, J., Hu, M.Y., Shanker, M., 2012. Human agency, social networks, and foss project success. *J. Bus. Res.* 65 (7), 977–984. <http://dx.doi.org/10.1016/j.jbusres.2011.04.014>, URL <http://www.sciencedirect.com/science/article/pii/S0148296311001317>.
- Wang, Y., Redmiles, D., 2016. The diffusion of trust and cooperation in teams with individuals' variations on baseline trust. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. In: *CSCW '16, Association for Computing Machinery*, New York, NY, USA, pp. 303–318. <http://dx.doi.org/10.1145/2818048.2820064>.
- Wang, J., Wang, Q., 2016. Analyzing and predicting software integration bugs using network analysis on requirements dependency network. *Requir. Eng.* 21 (2), 161–184.
- Wang, S., Zhang, W., Yang, Y., Wang, Q., 2013. Devnet: exploring developer collaboration in heterogeneous networks of bug repositories. In: *Empirical Software Engineering and Measurement, 2013 ACM/IEEE International Symposium on*. IEEE, pp. 193–202.
- Weiss, M., Moroiu, G., Zhao, P., 2006. Evolution of open source communities. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 21–32.
- Wiggins, A., Howison, J., Crowston, K., 2008. Social dynamics of floss team communication across channels. In: *IFIP International Conference on Open Source Systems*. Springer, pp. 131–142.
- Wolf, T., Nguyen, T., Damian, D., 2008. Does distance still matter?. *Softw. Process: Improv. Pract.* 13 (6), 493–510.
- Wolf, T., Schroter, A., Damian, D., Nguyen, T., 2009a. Predicting build failures using social network analysis on developer communication. In: *Proceedings of the 31st International Conference on Software Engineering*. IEEE Computer Society, pp. 1–11.
- Wolf, T., Schröter, A., Damian, D., Panjer, L.D., Nguyen, T.H., 2009b. Mining task-based social networks to explore collaboration in software teams. *IEEE Softw.* 26 (1), 58–66.
- Wu, J., Goh, K.-Y., Li, H., Luo, C., Zheng, H., 2016. The effects of communication patterns on the success of open source software projects: an empirical analysis from social network perspectives. *J. Glob. Inf. Manage. (JGIM)* 24 (4), 22–44.
- Wu, H., Liu, H., Ma, Y., 2018. Empirical study on developer factors affecting tossing path length of bug reports. *IET Softw.* 12 (3), 258–270. <http://dx.doi.org/10.1049/iet-sen.2017.0159>.
- Wu, W., Zhang, W., Yang, Y., Wang, Q., 2011. Drex: Developer recommendation with k-nearest-neighbor search and expertise ranking. In: *Software Engineering Conference (APSEC), 2011 18th Asia Pacific*. IEEE, pp. 389–396.
- Xu, J., Christley, S., Madey, G., 2005a. The open source software community structure. In: *NAACOS2005, Notre Dame, IN*.
- Xu, J., Gao, Y., Christley, S., Madey, G., 2005b. A topological analysis of the open source software development community. In: *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*. IEEE, p. 198a.
- Xu, J., Madey, G., 2004. Exploration of the open source software community. In: *Proceedings of North American Association for Computational Social and Organizational Science (NAACOS)*, Pittsburgh, PA, USA.
- Xuan, Q., Devanbu, P., Filkov, V., 2016. Converging work-talk patterns in online task-oriented communities. *PLoS One* 11 (5), e0154324.
- Xuan, Q., Gharehyazie, M., Devanbu, P.T., Filkov, V., 2012a. Measuring the effect of social communications on individual working rhythms: A case study of open source software. In: *Social Informatics (SocialInformatics), 2012 International Conference on*. IEEE, pp. 78–85.
- Xuan, J., Jiang, H., Ren, Z., Zou, W., 2012b. Developer prioritization in bug repositories. In: *Software Engineering (ICSE), 2012 34th International Conference on*. IEEE, pp. 25–35.
- Yang, X., 2014. Social network analysis in open source software peer review. In: *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*. In: *FSE 2014, Association for Computing Machinery*, New York, NY, USA, pp. 820–822. <http://dx.doi.org/10.1145/2635868.2661682>.
- Yang, G., Zhang, T., Lee, B., 2014. Utilizing a multi-developer network-based developer recommendation algorithm to fix bugs effectively. In: *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. In: *SAC '14, Association for Computing Machinery*, New York, NY, USA, pp. 1134–1139. <http://dx.doi.org/10.1145/2554850.2555008>.
- Yu, L., Ramaswamy, S., 2007. Mining cvs repositories to understand open-source project developer roles. In: *Proceedings of the Fourth International Workshop on Mining Software Repositories*. IEEE Computer Society, p. 8.
- Yu, L., Ramaswamy, S., 2013. A study of sourceforge users and user network. In: *2013 AAAI Fall Symposium Series*.
- Yu, L., Ramaswamy, S., Zhang, C., 2008. Mining email archives and simulating the dynamics of open-source project developer networks. In: *Fourth International Workshop on Enterprise and Organizational Modeling and Simulation: Proceedings of the Fourth International Workshop on Enterprise and Organizational Modeling and Simulation (France), Montpellier*. pp. 17–31.
- Yu, Y., Yin, G., Wang, H., Wang, T., 2014. Exploring the patterns of social behavior in github. In: *Proceedings of the 1st International Workshop on Crowd-Based Software Development Methods and Technologies*. ACM, pp. 31–36.
- Zanetti, M.S., Sarigol, E., Scholtes, I., Tessone, C.J., Schweitzer, F., 2012. A quantitative study of social organisation in open source software communities. *arXiv preprint arXiv:1208.4289*.
- Zanetti, M.S., Scholtes, I., Tessone, C.J., Schweitzer, F., 2013. Categorizing bugs with social networks: a case study on four open source software communities. In: *Proceedings of the 2013 International Conference on Software Engineering*. IEEE Press, pp. 1032–1041.
- Zanetti, M.S., Scholtes, I., Tessone, C.J., Schweitzer, F., 2013. The rise and fall of a central contributor: Dynamics of social organization and performance in the gentoo community. In: *2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*. pp. 49–56. <http://dx.doi.org/10.1109/CHASE.2013.6614731>.
- Zhang, W., Cheung, S.-C., Chen, Z., Zhou, Y., Luo, B., 2019. File-level socio-technical congruence and its relationship with bug proneness in oss projects. *J. Syst. Softw.* 156, 21–40.



- Zhang, W., Han, G., Wang, Q., 2014a. Butter: An approach to bug triage with topic modeling and heterogeneous network analysis. In: 2014 International Conference on Cloud Computing and Big Data. pp. 62–69. <http://dx.doi.org/10.1109/CCBD.2014.14>.
- Zhang, T., Lee, B., 2012. An automated bug triage approach: A concept profile and social network based developer recommendation. In: International Conference on Intelligent Computing. Springer, pp. 505–512.
- Zhang, Y., Shen, B., Chen, Y., 2014b. Mining developer mailing list to predict software defects. In: Software Engineering Conference (APSEC), 2014 21st Asia-Pacific, Vol. 1. IEEE, pp. 383–390.
- Zhang, W., Wang, S., Yang, Y., Wang, Q., 2013. Heterogeneous network analysis of developer contribution in bug repositories. In: Cloud and Service Computing (CSC), 2013 International Conference on. IEEE, pp. 98–105.
- Zhang, H., Wu, Y., Wu, W., 2015. Analyzing developer behavior and community structure in software crowdsourcing. In: Information Science and Applications. Springer, pp. 981–988.
- Zhang, T., Yang, G., Lee, B., Lua, E.K., 2014b. A novel developer ranking algorithm for automatic bug triage using topic model and developer relations. In: 2014 21st Asia-Pacific Software Engineering Conference, Vol. 1. pp. 223–230. <http://dx.doi.org/10.1109/APSEC.2014.43>.
- Zhang, W., Yang, Y., Wang, Q., 2011. Network analysis of oss evolution: an empirical study on argouml project. In: Proceedings of the 12th International Workshop on Principles of Software Evolution and the 7th Annual ERCIM Workshop on Software Evolution. ACM, pp. 71–80.
- Zhang, W., Yang, Y., Wang, Q., 2012. An empirical study on identifying core developers using network analysis. In: Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies. ACM, pp. 43–48.
- Zhang, Z., Yoo, Y., Wattal, S., Zhang, B., Kulathinal, R., 2014. Generative diffusion of innovations and knowledge networks in open source projects. In: Proceedings of the 35th International Conference on Information Systems.
- Zhou, M., Mockus, A., 2011. Does the initial environment impact the future of developers?. In: Proceedings of the 33rd International Conference on Software Engineering. ACM, pp. 271–280.

**PD Dr. Steffen Herbold** manages a Machine Learning professorship at the Institute of Computer Science of the University of Goettingen. His research is focused on the application of data science methods and their applications in software engineering as well as software engineering to data science methods.

**Aynur Amirfallah** works as Data Scientists at adesso SE. She received her B.Sc. degree in computer science from the Amirkabir University of Technology and her M.Sc. degree in applied computer science from the University of Göttingen. Her research interests include developer social networks and the identification of developer identities.

**Fabian Trautsch** works as IT Application Specialist at the Sartorius AG. Previously, he worked as a PostDoc in the Software Engineering for Distributed Systems group at the Institute of Computer Science of the University of Goettingen. He received the B.Sc. degree in applied computer science from the University of Goettingen in 2013, the subsequent M.Sc. degree in 2015 and his doctorate in 2019. His research interests include mining software repositories, software evolution, software testing, and empirical software engineering.

**Jens Grabowski** is professor at the University of Goettingen. He is the dean of the Faculty of Mathematics and Computer Science and is heading the Software Engineering for Distributed Systems Group at the Institute. Prof. Grabowski is one of the developers of the standardized testing languages TTCN-3 and UML Testing Profile. The current research interests of Prof. Grabowski are directed towards model-based development and testing, managed software evolution, and empirical software engineering.