On the Adoption of Social Network Analysis Methods in CSCL Research – A Network Analysis

Marielle Dado, Tobias Hecking, Daniel Bodemer, and H. Ulrich Hoppe
marielle.dado@uni-due.de, hecking@collide.info, bodemer@uni-due.de, hoppe@collide.info
University of Duisburg-Essen

Abstract: Originating from mathematical sociology, social network analysis (SNA) is a method for analyzing and representing relational structures in online communities. SNA applications in learning settings and CSCL scenarios are growing in popularity, which is in-line with new trends in learning analytics. For the CSCL research community, the adoption of SNA techniques as part of the methodological repertoire requires adequate understanding of the core concepts, their potential contributions, and limitations. We started from the hypotheses that (1) most applications of SNA in CSCL research make use of a small set of basic methods; and (2) the discourse related to SNA is partly inadequate or imprecise. To further analyze and corroborate these "issue hypotheses" we have used network analysis techniques in order to reveal relations between SNA measures and specific aspects of CSCL research (activities, contexts, research methods) based on a corpus of 90 published studies. Based on the results we pinpoint specific issues and outline new opportunities.

Introduction

The methodological foundation of CSCL creates a genuine interest in computational methods to analyze and formally represent relevant characteristics of learning groups and communities based on large amounts of digital data (e.g., log files) collected in technology-enhanced learning settings (Jeong, Hmelo-Silver, & Yu, 2014). Rooted in sociological studies of communities, social network analysis (SNA) provides a well-defined and elaborate mathematical apparatus that resonates with theoretical models based on actor-actor and actor-artefact relations (Wasserman & Faust, 1994). The adoption of SNA in CSCL started more than 15 years ago (Nurmela, Lehtinen & Palonen, 1999; Reffay & Chanier, 2003). Originally, networks derived from email and discussion boards were the most prominent type studied, such as the study of cohesion in learning groups using a shared forum (Reffay & Chanier, 2003). Martínez et al. (2003) present an evaluation method that combines SNA with traditional sources of data and analyses in blended collaborative learning scenarios. More recently, the interest in network analysis techniques related to the study of learning with CSCL and online learning as particular cases has been strengthened by the emergence of learning analytics as a new research paradigm (Haya et al., 2015).

Given the existing usage of SNA in CSCL research, it is not surprising that such methods have been subject to meta-level analyses. One example is a co-citation network analysis of CSCL studies from 2006 to 2013 by Tang, Tsai, & Lin (2014) which found that the most widely cited works in the field pertain to issues regarding CSCL communication and interaction patterns. Social interaction has been referred to as the "key" to collaborative learning: "If there is collaboration then social interaction can be found in it… if there is no social interaction then there is also no real collaboration" (Kreijns, Kirschner, & Jochems, 2003, p. 338). Interactions characterized as "collaborative" do not merely refer to how frequently peers in a joint activity interact, but also to how influential these interactions are in their cognitive processes (Dillenbourg, 1999). Suthers and colleagues consider the fundamental basis of CSCL interactions as the relationship present when one actor's learning activities builds upon that of another actor (Suthers et al., 2010).

Similarly, from the social network perspective, social relations are more important for understanding the behavior of groups and communities than individual attributes. In a group, actions and beliefs are strongly determined by social contexts and conditions (Wasserman & Faust, 1994). In this sense, the perspectives of CSCL and SNA coincide in focusing on social relations that go beyond individual characteristics. CSCL considers learning as a social endeavor, occurring as a result of relationships among learners and between learners and objects in the learning environment (Jones, 2015). By interacting, communicating, and sharing knowledge via computers, learners form "computer-supported social networks" from which learning emerges through constructive information exchange (Haythornthwaite, 1999). A social network approach to CSCL would thus help researchers and practitioners answer the questions: How is information distributed among learners? How much does the group share its information? What media supports this collaboration?

CSCL research has a vibrant, interdisciplinary research tradition that incorporate both qualitative and quantitative techniques (Jeong et al, 2014). SNA itself is also not exclusively one or the other, a trait that has been an attractive incentive for educational researchers with regards to quantifying social phenomena (Carolan, 2014). Nevertheless, the adoption of SNA methods requires a certain degree of familiarity with mathematical concepts from graph theory and network algorithms, similar to foundations of quantitative-empirical research in statistics. Yet, even after overcoming this first barrier, the new concepts have to be appropriated and integrated with the "CSCL epistemology" so that "technical" SNA are smoothly inter-related with pedagogical interpretations. The bridges that we use in this context should be sound and precise, not just surface-level shortcuts. One example of a problematic bridge is the notion of "interaction pattern" that is frequently mentioned in conjunction with SNA studies. Also, without a deeper understanding of the formal-analytic background a large part of the potential of SNA may remain unexploited.

To empirically study and corroborate these "issue hypotheses", we have applied network analysis techniques to a corpus of publications at the intersection of CSCL and SNA to detect and visualize relations between SNA measures and procedures and specific aspects of CSCL research. These networks encode concept-concept relations based on co-occurrence extracted from the abstracts and the results of publications which were previously subjected to a qualitative literature review. The concepts have been categorized, which allows for introducing a cross-category perspective in the form of multi-mode networks. The themes and trends that emerged from this analysis are then evaluated based on technical definitions of the SNA techniques and may serve as a springboard for researchers to (1) understand how CSCL research is currently viewed through the lens of SNA and (2) identify opportunities to apply underexplored SNA techniques to expand the current knowledge base in CSCL.

Data sample and qualitative findings

The application of SNA in education research has grown in the last decade. However, relative to other research methods, the use of SNA in CSCL settings is not as well-established. A recent review of SNA in a related field, e-learning research (Cela, Sicilia, & Sanchez, 2015), revealed that SNA is mostly applied to study direct interactions between learners collaborating in online discussion forums based on communication patterns; these are usually measured using density and centrality indices. SNA is also often combined with qualitative content analysis to provide a deeper understanding of the nature of learner interaction within the network. The review however is preliminary and limited in that the general search term "e-learning" may have excluded relevant studies that use more specific terminology (e.g., CSCL).

In order to uncover trends in the application of SNA in CSCL research, a qualitative literature review was conducted. Ninety full-text studies published as peer-reviewed journal articles, book chapters, and conference papers were collected from October to November 2015 using the following keywords: social network analysis" AND "computer-supported collaborative learning" online OR computer OR collaborat* (e.g., "collaborative" and "collaboration") OR learning. To be included in the analysis, studies must: (1) use primary data; (2) be set in an instructional course/program up to the postgraduate (Masters) level; (3) use SNA techniques, explicitly mentioned in the Methods section; (4) report SNA findings in the Results section; and (5) analyze collaborative learning activities between learners using computers. Information on the general methodology (research design, learning setting, collaborative activity, non-SNA methods) and SNA features (actor type, relational tie, SNA measures and analysis on SNA data) were identified in each paper and quantified using content analysis.

Similar to the findings of Cela et al (2015) in the e-learning literature, between 50% to 70% of the analyzed studies measured interaction as direct communication between learners during project or task-based activities in blended learning settings, primarily using centrality and density indices. About a quarter of papers conducted content analysis to supplement SNA findings, although analyses of SNA data in most studies were limited to a descriptive report of the SNA indices. More sophisticated SNA procedures, such as identifying network positions and detecting cliques and subgroups, appeared in less than 20% of studies. A handful of studies conducted correlational analysis or inferential statistics on SNA and learner characteristics to enhance the implications of network data on learning.

These results suggest that applications of SNA in CSCL research are rather homogenous, dominated by the basic local and global measures of centrality and density as indicators of social interaction in CSCL environments. Although SNA is a promising method for analyzing collaborative learning in computer-mediated settings, the qualitative literature review lends support to the hypothesis that the analysis of CSCL interactions using SNA may not be adequate due to the limited range of applied SNA measures. In the present paper, we extend the qualitative results by exploring which CSCL activities, contexts, and research methods are associated with which SNA measures and procedures, using network analysis as a technique for meta-level literature analysis (cf. Tang et al, 2014).

Methodology

In network text analysis (NTA) words or concepts are linked by relations based on proximity of occurrence in a text, manual coding, or grammar relations. By extracting terms in a set of texts and constructing a network based on how these terms relate to each other, NTA aims at preserving the conceptual structure of texts. More recently, NTA has also been applied to model and visualize the conceptual structure of learners within a knowledge domain based on questions and answers (Daems, Erkens, Malzahn, & Hoppe, 2014). When used in this manner, NTA could be considered automated technique for classical content analysis (Diesner & Carley, 2005). Our NTA approach is based on the network extraction pipeline used in state of the art NTA tools, such as Tools Automap (Diesner & Carley, 2005) and ConText (Diesner, 2014), and comprises three main steps: (1) concept identification, (2) concept normalization and classification, and (3) relation extraction.

First, the abstracts and the results section of the papers were prepared for analysis by removing non-relevant words (articles, auxiliary verbs, etc.) and stemming by removing suffixes. Then, parts-of-speech identification was done to identify nouns, adverbs and adjectives that represent concepts in the literature. The extraction process produced 3,057 unigrams (single nouns) and 38,934 bi-grams (two terms, combination of nouns, adjectives and adverbs), from which the top 150 unigrams and bigrams were included in the next step.

Table 1: Codebook example

Term	Concept	Categories
knowledge construction	KNOWLEDGE_CONSTRUCTION	CSCL_ACTIVITY
construction of knowledge	KNOWLEDGE_CONSTRUCTION	CSCL_ACTIVITY
online course	ONLINE_COURSE	CSCL_CONTEXT
content analysis	CONTENT_ANALYSIS	CSCL_METHOD
betweenness centrality	BETWEENNESS_CENTRALITY	SNA

An excerpt of the codebook used to identify and classify concepts in the pool of studies is displayed on Table 1. The first column contains concrete terms occurring in the texts. To account for different spellings and synonyms, the second column maps the specific terms in the first column to a general concept. The third column assigns each concept to one of the four categories: (1) "CSCL activity" for terms associated with aspects of a collaborative learning activity (e.g., "knowledge building", "score"); (2) "CSCL context" for terms pertaining to physical/virtual settings or platforms where CSCL activities take place (e.g., "class", "forum"); (3) "CSCL method" for terms pertaining to other analysis methods applied alongside SNA (e.g., "correlation"); (4) "SNA", for terms related to SNA procedures and techniques (e.g., "centrality"). After combining synonyms and spelling variations, the final analysis included 101 concepts: 22 SNA, 45 CSCL activity, 23 CSCL context, 11 CSCL method. Figure 1 shows how the codebook is used to automatically identify and classify the concepts in text.

```
This paper took the social KNOWLEDGE_CONSTRUCTION as the
perspective and "Introduction to Educational Technology"
ONLINE_COURSE as an example to analyze
KNOWLEDGE_CONSTRUCTION level implied in those POSTS
contributed by learners with the method of
CONTENT_ANALYSIS. Meanwhile, social network analysis (SNA)
was adopted to explore the DENSITY, CENTRALITY, COHESION
in this online LEARNING_COMMUNITY and to discuss
strategies for effective collaborative learning in virtual
LEARNING_COMMUNITY. Results indicated that the entire
network CENTRALIZATION is comparatively low but still some
points with higher BETWEENNESS_CENTRALITY and some points
functioned as bridges exist in our sample network.
```

Figure 1. Concept identification, normalization and classification.
(purple: CSCL_ACTIVITY, blue: CSCL_CONCEPT, orange: CSCL_method, blue: CSCL: CONTEXT, green: SNA)

Once the CSCL and SNA concepts have been identified, the next step is to extract a concept network that reflects the associations made between the different concepts in the selected CSCL literature. First, for each publication the identified concepts from the codebook are interlinked to a fully connected network (or concept clique). Second, the concept cliques corresponding to particular publications are merged by overlaying these networks such that the result is a single network which contains all concepts and all links from the original networks. It is important to mention that in order to account for the relationships between SNA and CSCL a bipartite version of the network was used, which restricts the original one to having edges solely between SNA and CSCL concepts. The weight of each edge between two concepts corresponds to the number of concept cliques containing this edge, i.e., the number of papers in which the concepts co-occur.

The resulting edge-weighted and bipartite network together with the category attributes of nodes is the basis for our further analyses. In particular, we analyze the network in terms of distance between concepts and cohesive clusters, which will be described in more detail in the following section.

Analysis and results

For a first overview we provide a frequency count of the number of publications mentioning the 22 SNA concepts. The result shown in Figure 2 supports the findings of the qualitative literature review that the usage of SNA in CSCL research is mostly restricted to centrality analysis. The most basic measure "degree centrality" appears in the abstracts and result sections of 70 of the 90 papers. In contrast, more advanced techniques such as modularity of sub-communities, positional analysis using blockmodels, or network simulation are rarely mentioned.

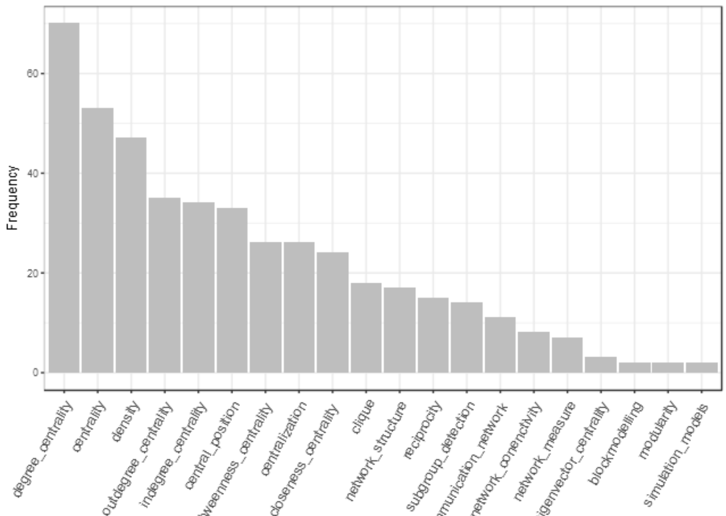

Figure 2. Number of occurrences of different SNA concepts in the 90 documents.

To further investigate the contexts in which particular SNA concepts are applied, the previous findings are extended by identifying for each SNA concept the closest CSCL concepts based on geodesic distance in the concept network. The edge weights (i.e., number of co-occurrences) of concepts were taken into account by setting the edge distance between connected concepts to the inverse of its weight.

Table 2: Profiles of SNA concepts based on proximities to different types of CSCL concepts

<table>
<tr><th>SNA</th><th>CSCL Activity</th><th>CSCL Context</th><th>CSCL Method</th></tr>
<tr><td>Degree Centrality</td><td rowspan="2">groupwork, interaction, help</td><td rowspan="5">course, message, class</td><td>interaction_pattern, correlation, questionnaire</td></tr>
<tr><td>Centrality</td><td rowspan="3">correlation, content_analysis, interaction_pattern</td></tr>
<tr><td>Density</td><td rowspan="3">groupwork, interaction, communication</td></tr>
<tr><td>Outdegree Centrality</td></tr>
<tr><td>Indegree Centrality</td><td>interaction_pattern, correlation, questionnaire</td></tr>
<tr><td>Central Position</td><td>groupwork, interaction, role</td><td rowspan="2">message, course, post</td><td>content_analysis, correlation, interaction_pattern</td></tr>
<tr><td>Betweenness Centrality</td><td>groupwork, interaction, discussion</td><td>correlation, interaction_pattern, questionnaire</td></tr>
<tr><td>Centralization</td><td>groupwork, interaction, communication</td><td rowspan="2">course, message, class</td><td>content_analysis, interaction_pattern, correlation</td></tr>
<tr><td>Closeness Centrality</td><td rowspan="2">groupwork, help, interaction</td><td rowspan="6">correlation, interaction_pattern, questionnaire</td></tr>
<tr><td>Clique</td><td>course, message, post</td></tr>
<tr><td>Network Structure</td><td rowspan="2">groupwork, interaction, communication</td><td rowspan="4">course, message, class</td></tr>
<tr><td>Reciprocity</td></tr>
<tr><td>Subgroup Detection</td><td>groupwork, interaction, help</td></tr>
<tr><td>Communication Network</td><td>communication, groupwork, interaction</td></tr>
</table>

Network Connectivity	groupwork, communication, help		
Network Measure	groupwork, discussion, interaction		
Eigenvector Centrality	groupwork, help, interaction		
Blockmodelling	communication, groupwork, discussion	course, discussion_forum, message	
Modularity		class, post, course	
Simulation Models	communication, groupwork, participation	course, message, class	time_period, interaction_pattern, correlation

Table 2 lists for all SNA concepts the three closest CSCL concepts of each category (Activity, Context, Method) yielding characteristic profiles of the SNA concepts in terms of their proximity to CSCL concepts. "Groupwork", "interaction", and "communication" as CSCL activities are the closest to most SNA terms, which suggests that interactions within CSCL group activities are often characterized by communication. The proximity of the concept "help" to the node- and group-centric SNA concepts (e.g., centrality, clique) suggests that help or support are mostly associated with positions of individuals in a social network. In the CSCL method category, "correlation analysis" and "questionnaire" alongside "interaction pattern" appeared in almost every profile. As was found in the qualitative literature review, this indicates that SNA is commonly used in combination with empirical data collected from questionnaires to relate structural network properties to quantitative measures of learning. Simulation models for social networks constitute the only SNA concept that is closely related to time ("time_period" in the category CSCL Method).

Next, we have specifically analyzed the relation between SNA concepts and CSCL concepts in terms of cohesive bipartite network clusters to reveal the inherent organization of the complex network. Cohesive clusters in such a network stand for subgroups or clouds of terms that are more densely connected among each other than the average of the network. This cohesiveness can be characterized by the "modularity" measure (cf. Barber, 2007), here particularly using *bipartite modularity maximization* (Hecking, Steinert, Göhnert, & Hoppe, 2014). The clustered network representation shown in Figure 3 results from the following workflow: (1) filtering out of "weak" edges with weight below 8; (2) further reduction of the network to its 2-core to ensure that the remaining nodes are connected to at least 2 others (no singular or satellite nodes); (3) identification of mixed clusters of SNA and CSCL concepts based on bipartite modularity; (4) visualization, as presented in Figure 3. The color of the nodes indicates cluster association and node size is scaled according to its degree in the weighted network. The shapes of the nodes represent the different categories.

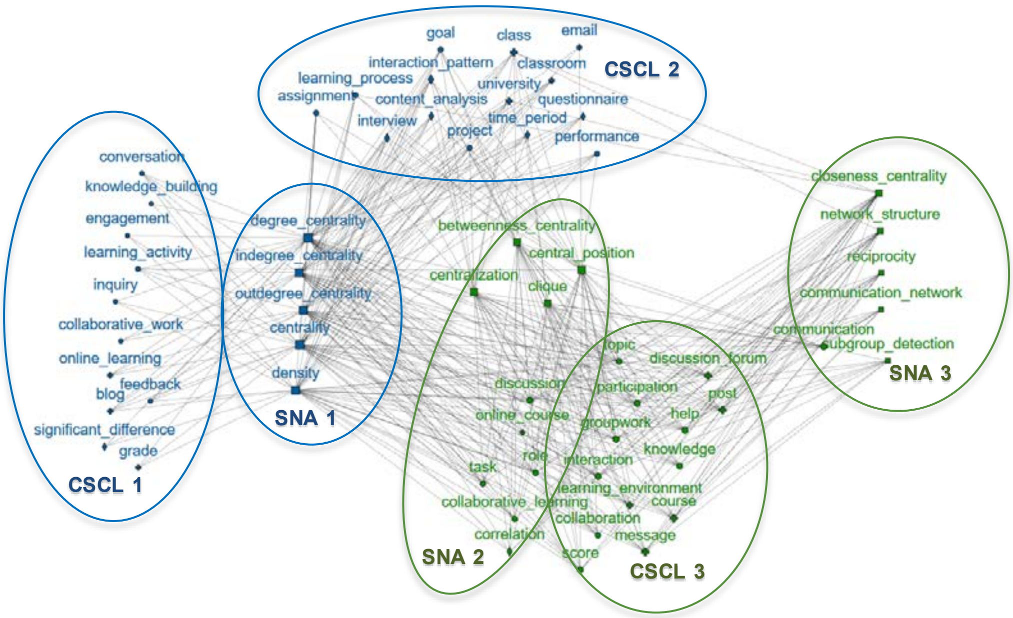

Figure 3. Result of the bipartite modularity optimization.
(• - CSCL activities, ♦ - CSCL research methods, ✚ - CSCL context, ■ - SNA concepts)

The applied clustering algorithm identifies two clusters. The network is very densely connected, however interesting relational patterns between different parts of the network are salient. The blue cluster can be further divided into three groups of concepts indicated by the ellipses in Figure 3. The group "CSCL 1" contains general CSCL activities that are almost solely linked to the most common SNA concepts (SNA 1) and not to third blue group (CSCL 2), which in turn is strongly connected to the SNA concepts in its cluster (SNA 1) as well as in the other cluster (SNA 2). A similar pattern can be discovered for the green cluster. Here there is a group of SNA concepts (SNA 3) that is almost completely separated from the blue cluster. The concepts in this group can be considered as less common compared to the other SNA concepts in the network. The green cluster further contains SNA concepts (SNA 2) and CSCL concepts (CSCL 3) that act as bridges between the two large concept clusters. The distinction between groups of concepts based on "strong" connections to groups within the same cluster and "weak" connections to groups of the other cluster can be used to reveal the overall relational structure of SNA and CSCL concepts in the investigated literature. This macro-structure is shown in Table 3, which is also called an image matrix (Wasserman & Faust, 1994). Image matrices depict the presence or absence of pre-defined relations (here weak, strong, or absent) between different parts of a network and can be considered as a characteristic to help interpret the inherent organization of a given network

Table 3: Relationships between different parts of the concept network as image matrix

	CSCL 1 (blue cluster)	CSCL 2 (blue cluster)	CSCL 3 (green cluster)
SNA 1 (blue cluster)	strong	strong	weak
SNA 2 (green cluster)	absent	weak	strong
SNA 3 (green cluster)	absent	absent	strong

SNA concepts in SNA 1 (indegree/outdegree centrality, centrality, degree centrality) and CSCL 3 (e.g., course, post, collaborative_learning, groupwork), are interlinked to most parts of the network, and can be considered as the conceptual core of the usage of SNA in CSCL research. Note that the terms in CSCL 3 are the same as the CSCL terms that appear in the proximity analysis in Table 2. The least connected CSCL group (CSCL 2) pertain to specific environments (e.g., blogs), pedagogical approaches (e.g., knowledge building), or processes (e.g., engagement); these are studied using most common SNA indices of centrality and density (SNA 1). Similarly, the least connected SNA group (SNA 3) contains techniques that are less frequently applied; this cluster is only associated with the most prominent CSCL concepts (CSCL 3).

Discussion

Overall the analyses presented in this paper demonstrate a visible relational structure and conceptual core of SNA and CSCL activities, contexts, and methods that are present in the CSCL literature. The profiles generated using network analysis techniques corroborate the findings of the qualitative literature review, namely that SNA applications in CSCL research aim to understand CSCL "interaction patterns" based on communication. This was the case for all SNA indices included in the analysis and not only the most frequently used indices of centrality and density. The results also show that there have been efforts to associate these interaction patterns with learning-related variables using statistical methods (correlation), which could be seen as a way of bridging research perspectives between CSCL and education traditions (Carolan, 2014). In this section we critically discuss the core CSCL concept of "interaction patterns" in relation to the technical underpinnings of the core SNA indices.

Given the high connectivity of "interaction" and "communication" to the most prominent SNA concepts in our pool of studies, one might think that "interaction pattern" is a technical term derived from SNA. However, the basic SNA measures that were used frequently in our sample, especially centrality measures but also subgroup detection methods, only characterize relational attributes based on the connectivity of single nodes (actors), sub-networks, or the entire network. Centralization and reciprocity are global measures that indeed reveal certain general network characteristics and can characterize the topology of communication in networked learning environments. This can be conceived as a certain type of structural-relational pattern, but it is not about repeated concrete constellations, especially not in a temporal sense. The SNA technique of blockmodeling would be a means of detecting roles and role models based on consistent network structures and relational similarity. However, as our results show, it has been rarely used in CSCL studies.

Furthermore, single instances of social networks do not represent time-dependent relations, but rather capture and aggregate relations harvested during a given time window. Hoppe, Harrer, Göhnert, Hecking (2016) have made the point that the choice of different time windows as a step prior to the network generation can have systematic effects on the ensuing analysis results. While SNA research has developed several ways to handle dynamic networks that evolve over time (Aggarwal & Subbian, 2014), such techniques are not widely adopted in

CSCL: the only indication of considering time dependencies in combination with CSCL concepts was related to network simulation models, which is one of the least common SNA methods in our results. Since it has been argued that the explicit consideration of temporal processes is crucial to make sense of data produced in CSCL environments (Reimann, 2009), a future advancement of SNA in CSCL can be to consider dynamic network analysis methods based on time series of graphs.

All this indicates that the notion of "interaction patterns" as measured in SNA is indeed not very specific in a technical sense: what constitutes "interaction patterns" is not strictly operationalized in SNA. Thus, the actual technical definitions of SNA measures should be clarified when interpreting SNA findings. The SNA concept of "density", the third frequent in our list of SNA concepts is another example for potential issues: There is a general caveat concerning the usage of density as a comparative measure applied to networks of different sizes (including growing networks). *Density* as a general measure for graphs is equal to the *average degree divided by the number of nodes* in the graph. On the other hand, in most naturally evolving networks the *average degree* will grow (if at all) at a much lower rate than the number of nodes. For scale-free networks (Barabasi & Bonabeau, 2003) the average degree is inherently constant. This means that the smallest networks will have the highest density. Hoppe, Engler & Weinbrenner (2012) have discussed this effect when studying student-generated concept maps. This makes it difficult to definitively identify an "optimal" density level of communication in CSCL research, where units of analysis tend to vary in size (Stahl, 2015). Density can be reasonably used as a comparative measure only with networks that have an identical number of nodes, otherwise the ratio of *number of edges per number of nodes* (which is proportional to the *average degree*) should be used.

In sum, the results indicate mismatches between the intended aims of applying SNA in CSCL (i.e., to investigate interaction patterns) and the actual technical definitions of SNA concepts even at the most basic level. SNA *describes* characteristics of network structures based on several common indices, and those descriptions do not necessarily capture consistent patterns of interaction that persist over time and onto other contexts. A number of sophisticated SNA techniques that are able to accomplish this, such as blockmodeling, measures on time series of networks, and network simulations, are largely underexplored.

Conclusion

The adoption of SNA techniques in CSCL is very much focused on understanding and modeling "interaction patterns". However, we also argue that CSCL "interaction patterns" do not necessarily become apparent from the basic and most commonly used SNA indices. However, despite SNA originating from a different analytical tradition, its use in combination with statistical analysis shows how SNA is able to cut through disciplinary boundaries. We hope that interested CSCL researchers will use our analysis as a basis for expanding the current knowledge body to include advanced SNA techniques for exploring other network dimensions such as time. This challenge would not only contribute to a more nuanced analysis on CSCL interactions, but it would also enrich the interdisciplinarity of research methods and the skill sets of CSCL researchers. As long as the CSCL community is interested in the study of social interactions, there is a place for SNA in this research field.

References

Aggarwal, C., & Subbian, K. (2014). Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, *47*(1), 10. doi:10.1145/2601412.

Barabási, A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, *288*(5), 50-59.

Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E, 76*(6), 066102. https://doi.org/10.1103/PhysRevE.76.066102

Carolan, B. V. (2014). *Social network analysis and education: Theory, methods & applications*. California: Sage Publications.

Cela, K. L., Sicilia, M. A., & Sanchez, S. (2015). Social network analysis in e-Learning environments: A preliminary systematic review. *Educational Psychology Review*, *27*(1), 219–246. doi:10.1007/s10648-014-9276-0.

Dillenbourg, P. (1999). What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Oxford, UK: Elsevier.

Daems, O., Erkens, M., Malzahn, N., & Hoppe, H. U. (2014). Using content analysis and domain ontologies to check learners' understanding of science concepts. *Journal of Computers in Education, 1*(2), 113-131. doi.10.1007/s40692-014-0013-y

Diesner, J. (2014). *ConText: Software for the integrated analysis of text data and network data.* Paper presented at the Social and Semantic Networks in Communication Research. Preconference at Conference of International Communication Association (ICA), Seattle, WA.

Diesner, J., & Carley, K. M. (2005). Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis. In V.K. Narayanan and D. Armstrong (eds.), *Causal mapping for information systems and technology research* (p. 81-108). Hershey, PA: IGI Global.

Haythornthwaite, C. (1999). Collaborative work networks among distributed learners. In R. Sprague (ed.), *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, 1999. HICSS-32, Hawaii, US* (pp. 1023-39). Virginia, US: IEEE Computer Society Press. doi:10.1109/HICSS.1999.772707

Haya, P. A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H. U. (2015). Analysing content and patterns of interaction for improving the learning design of networked learning environments. *British Journal of Educational Technology*, *46*(2), 300-316. doi:10.1111/bjet.12264

Hecking, T., Steinert, L., Göhnert, T., & Hoppe, H. U. (2014). Incremental clustering of dynamic bipartite networks. In *Proceedings of the 2nd European Network Intelligence Conference (ENIC 2014) Wroclaw, Poland* (pp. 9-16). Los Alamitos, CA : Conference Publishing Services. doi:10.1109/ ENIC.2014.15

Hoppe, H. U., Engler, J., & Weinbrenner, S. (2012). The impact of structural characteristics of concept maps on automatic quality measurement. In J. van Aalst, K. Thompson, M. Jacobson, and P. Reimann (eds.), *Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012), Sydney, Australia* (pp. 308-316). Sydney: International Society of the Learning Sciences. www.isls.org/icls2012

Hoppe, H. U., Harrer, A., Göhnert, T., & Hecking, T. (2016). Applying network models and network analysis techniques to the study of online communities. In U. Cress, J. Moskaliuk, and H. Jeong (eds.), *Mass Collaboration and Education* (pp. 347-366). Switzerland: Springer International Publishing.

Jeong, H., Hmelo-Silver, C. E., & Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *International Journal of Computer-Supported Collaborative Learning*, *9*(3), 305–334. doi:10.1007/s11412-014-9198-3.

Jones, C. (2015). *Networked learning: an educational paradigm for the age of digital networks*. Switzerland: Springer International Publishing.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*(3), 335–353. doi:10.1016/S0747-5632(02)00057-2.

Martínez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & De La Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, *41*(4), 353-368. doi:10.1016/j.compedu.2003.06.001

Nurmela, K., Lehtinen, E., & Palonen, T. (1999). Evaluating CSCL log files by social network analysis. In C. Hoadley and J. Roschelle (eds.), *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning*, article no. 42. International Society of the Learning Sciences. http://dl.acm.org/citation.cfm?id=1150294.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 5–42. doi:10.1007/s11412-009-9081-9.

Stahl, G. (2015). A decade of CSCL. *International Journal of Computer-Supported Collaborative Learning*, *10*(4), 337–344. doi:10.1007/s11412-015-9222-2.

Tang, K. Y., Tsai, C. C., & Lin, T. C. (2014). Contemporary intellectual structure of CSCL research (2006–2013): a co-citation network analysis with an education focus. *International Journal of Computer-Supported Collaborative Learning*, *9*(3), 335–363. doi:10.1007/s11412-014-9196-5.

Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In B. Wasson, S. Ludvigsen, and H. U. Hoppe (eds.), *Designing for change in networked learning environments* (pp. 343-352). Springer Netherlands. doi:10.1007/978-94-017-0195-2

Reimann, P. (2009). Time is precious: Variable-and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 239-257. doi:10.1007/s11412-009-9070-z.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 1). Cambridge University Press.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. GRK 2167, Research Training Group "User-Centred Social Media".