# Dear Author

Here are the proofs of your article.

- You can submit your corrections **online,** via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For **fax** submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
- Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

**Please note**

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL:

`http://dx.doi.org/10.1007/s11412-018-9285-y`

If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information, go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us, if you would like to have these documents returned.

# Metadata of the article that will be visualized in OnlineFirst

| | | |
|---|---|---|
| Article Title | **Using big data techniques for measuring productive friction in mass collaboration online environments** | |
| Article Sub-Title | | |
| Article Copyright Year | **The Author(s) 2018** **(This will be the copyright line in the final PDF)** | |
| Journal Name | **International Journal of Computer-Supported Collaborative Learning** | |
| Corresponding Author | Family Name | **Holtz** |
| | Particle | |
| | Given Name | **Peter** |
| | Suffix | |
| | Division | Knowledge Construction Lab |
| | Organization | Leibniz-Institut für Wissensmedien IWM / Knowledge Media Research Center |
| | Address | Schleichstraße 6, 72076 Tübingen, Germany |
| | e-mail | p.holtz@iwm-tuebingen.de |
| Author | Family Name | **Kimmerle** |
| | Particle | |
| | Given Name | **Joachim** |
| | Suffix | |
| | Division | Knowledge Construction Lab |
| | Organization | Leibniz-Institut für Wissensmedien IWM / Knowledge Media Research Center |
| | Address | Schleichstraße 6, 72076 Tübingen, Germany |
| | Organization | Eberhard Karls Universität Tübingen |
| | Address | Tübingen, Germany |
| Author | Family Name | **Cress** |
| | Particle | |
| | Given Name | **Ulrike** |
| | Suffix | |
| | Division | Knowledge Construction Lab |
| | Organization | Leibniz-Institut für Wissensmedien IWM / Knowledge Media Research Center |
| | Address | Schleichstraße 6, 72076 Tübingen, Germany |
| | Organization | Eberhard Karls Universität Tübingen |
| | Address | Tübingen, Germany |
| Schedule | Received | 1 August 2017 |
| | Revised | |
| | Accepted | 9 October 2018 |

AUTHOR'S PROOF

Abstract

The advent of the social web brought with it challenges and opportunities for research on learning and knowledge construction. Using the online-encyclopedia Wikipedia as an example, we discuss several methods that can be applied to analyze the dynamic nature of knowledge-related processes in mass collaboration environments. These methods can help in the analysis of the interactions between the two levels that are relevant in computer-supported collaborative learning (CSCL) research: The individual level of learners and the collective level of the group or community. In line with constructivist theories of learning, we argue that the development of knowledge on both levels is triggered by productive friction, that is, the prolific resolution of socio-cognitive conflicts. By describing three prototypical methods that have been used in previous Wikipedia research, we review how these techniques can be used to examine the dynamics on both levels and analyze how these dynamics can be predicted by the amount of productive friction. We illustrate how these studies make use of text classifiers, social network analysis, and cluster analysis in order to operationalize the theoretical concepts. We conclude by discussing implications for the analysis of dynamic knowledge processes from a learning sciences perspective.

Keywords (separated by '-'): Learning - Knowledge construction - Productive friction - Big data - Wikipedia

Foot note information

Intern. J. Comput.-Support. Collab. Learn.
https://doi.org/10.1007/s11412-018-9285-y

# Using big data techniques for measuring productive friction in mass collaboration online environments

Peter Holtz[1] • Joachim Kimmerle[1,2] • Ulrike Cress[1,2]

Received: 1 August 2017 / Accepted: 9 October 2018


**Abstract**
The advent of the social web brought with it challenges and opportunities for research on learning and knowledge construction. Using the online-encyclopedia Wikipedia as an example, we discuss several methods that can be applied to analyze the dynamic nature of knowledge-related processes in mass collaboration environments. These methods can help in the analysis of the interactions between the two levels that are relevant in computer-supported collaborative learning (CSCL) research: The individual level of learners and the collective level of the group or community. In line with constructivist theories of learning, we argue that the development of knowledge on both levels is triggered by productive friction, that is, the prolific resolution of socio-cognitive conflicts. By describing three prototypical methods that have been used in previous Wikipedia research, we review how these techniques can be used to examine the dynamics on both levels and analyze how these dynamics can be predicted by the amount of productive friction. We illustrate how these studies make use of text classifiers, social network analysis, and cluster analysis in order to operationalize the theoretical concepts. We conclude by discussing implications for the analysis of dynamic knowledge processes from a learning sciences perspective.

**Keywords** Learning · Knowledge construction · Productive friction · Big data · Wikipedia

## Introduction Q1

Learning and knowledge construction are often the result of learners' overcoming particular challenges, such as successfully dealing with new, unexpected, or contradicting information. In the tradition of Piaget, the term *socio-cognitive conflict* has been used (e.g., Mugny and Doise

✉ Peter Holtz
p.holtz@iwm-tuebingen.de

[1] Knowledge Construction Lab, Leibniz-Institut für Wissensmedien IWM / Knowledge Media Research Center, Schleichstraße 6, 72076 Tübingen, Germany

[2] Eberhard Karls Universität Tübingen, Tübingen, Germany

1978) to indicate that disturbances of individuals' cognitive systems can, among others, result from other people's differing cognitive schemas. For positive consequences of conflicts in large-scale social settings with many collaborators (see Jeong et al. 2017), we suggest using the term *productive friction* that originally has been introduced in organization science (Hagel 3rd and Brown 2005; for an application in the educational context see Ward et al. 2011). Productive friction refers to the process of overcoming obstacles in a productive way that, in turn, leads to individual learning and collaborative knowledge construction (Kimmerle et al. 2015). Friction between different people (for example in the form of contradicting knowledge, schemas, scripts, values, meanings, and other cognitive structures) can incur costs in the short term. But overcoming this friction in a productive way may lead to innovation and new knowledge in the long run (Hagel 3rd and Brown 2005).

In the reasoning that we present here, we use the term productive friction in a broad sense for all kinds of discrepancies among different individuals' thoughts, ideas, values, and attitudes or between an individual's schemas and the general view of a social group, which can be resolved in a way that productively fosters epistemic growth and development (Bientzle et al. 2013; Kimmerle et al. 2017a, b). We argue that the same friction that drives learning processes Q2 on the level of individuals may trigger knowledge construction processes on the social level as well. If the friction is too low, the chances for learning and knowledge construction are limited, because the knowledge that already exists is by and large sufficient to solve the tasks at hand. If the friction is too high on the individual level, learners will most likely fail to adapt to the challenges. On the social level, too much friction among the thoughts and ideas of individuals in the community can either lead to a situation where ideas are ignored or to a rift in the community. Earlier research on the role and importance of socio-cognitive conflicts for learning processes focused on individual learners and changes in their cognitive schemas (e.g., Darnon et al. 2007; Doise et al. 1975; Mugny and Doise 1978). The position that we present here focuses on the interrelatedness between individual processes on the one hand, and meaning negotiation in communities on the other.

In the following paragraphs, we will first discuss how productive friction may play out in the social web and what this implies for capturing data and applying learning analytics approaches. Then we will introduce theoretical considerations on the dynamic and collective nature of knowledge. After that, we will combine these methodological and theoretical considerations and illustrate how productive friction may occur in Wikipedia by summarizing previously published studies that represent three different methodological approaches: The automatic classification of text, social network analysis, and cluster analysis. Our main focus is on showing how both the notion of knowledge as a dynamic and collective phenomenon and the phenomenon of socio-cognitive conflicts can be operationalized in such a way that these conceptions can be used to explain and predict dynamics of knowledge construction in Wikipedia. Finally, in the conclusion, we will discuss implications for the future analysis of knowledge-related processes in CSCL environments.

## Productive friction in the social web

The digital revolution in the second half of the twentieth century and the emergence of the Internet as a mass phenomenon in the 1990s have profoundly changed how people deal with information and how they acquire knowledge (Castells 2010; Happer and Philo 2013). People receive much of their information from the Internet (e.g., Hermida et al. 2012). One example of a prominent information source on the Internet is Wikipedia, which represents a knowledge corpus that has been constructed through collaborative endeavor (Cress et al. 2016). Comprehending the underlying processes of such collaboration presents several challenges for the learning sciences—in particular from a methodological point of view. In this article we aim to provide illustrating examples of methods that can be applied to analyze or predict the dynamic knowledge processes in the mass collaboration environment Wikipedia and discuss some implications from an educational perspective.

The way knowledge develops in the social web makes it obvious that knowledge is hardly ever a purely individual and never a static phenomenon (Oeberst et al. 2016). Quite in contrast, the occurrence, progress, and dissemination of knowledge have for a long time been considered to be social and dynamic phenomena (Brown et al. 1989; Resnick 1991). The article presented here makes an attempt to shed light on the complex dynamic knowledge-related processes that take place at the individual as well as on the collective levels. In line with Piaget (1970, 1977) and other educational constructivists, we view disturbances in the form of socio-cognitive conflicts (Mugny and Doise 1978) as important triggers for knowledge dynamics. While most of the earlier studies have dealt with dyads or small groups of learners, we aim to apply the concept of socio-cognitive conflict to large groups. We argue that a certain amount of conflict is instrumental for driving complex dynamic processes in mass collaboration environments (Jeong et al. 2017; Kimmerle et al. 2015).

One side effect of the emergence of social media as a mass phenomenon is the sudden availability of massive amounts of data in the form of behavior traces, such as the edit history of a digital artifact, or the browsing histories of users. Compared to data acquired using traditional data collection techniques, such as questionnaires or the observation of behavior in a laboratory experiment, behavior trace data is most often weakly structured: Relevant constructs such as learning trajectories, for instance, first have to be derived from an abundance of data points from user-artifact interactions (e.g., D'Aquin et al. 2017).

Analysis of this kind of data has only recently become feasible through the increasing calculation capabilities of modern computers. The main applications for big data in education research have been *learning analytics*, that is, the optimization of learning environments and learning activities on the basis of analysis of data on a learner's activities and learning outcomes (Gasevic et al. 2014; Ferguson 2012; Lang et al. 2017; Siemens and Long 2011). Learning analytics may be applied to supervise, predict, and potentially improve learner performance (Dietz-Uhler and Hurn 2013). Frequently, learning analytics are used to devise recommender systems that provide learners with useful learning resources according to their learning trajectory (i.e., providing content for adaptive learning). Learning analytics are also used in the context of *predictive modeling* to identify as early as possible unwanted developments, such as a learner's displaying signs of losing motivation or interest in the respective subject. Together, these technologies can help in devising learning goals for traditional educational settings or virtual learning environments, with the advantage that they can take into account learners' individual predilections and habits as well as general abstract patterns that are indicative of positive or negative trajectories (Picciano 2012; Siemens and Long 2011). Learning analytics, however, also come with some ethical challenges, such as privacy issues (for an overview see Slade and Prinsloo 2013).

Empirical studies that use such applications need a theoretical foundation. Otherwise, researchers run the risk of getting lost in the plethora of variables and results they are dealing with. They may easily miss out on potentially relevant confounding variables, subgroups, or covariates in their analyses. In addition, without a theoretical framework, researchers may encounter difficulties in interpreting their findings and in determining to which cases their results can be generalized (Wise and Shaffer 2015). Empirical studies with large amounts of data can expand our knowledge about processes of learning and knowledge construction if their findings have theoretical importance. That is, they are useful if they can guide the testing of existing theories or if they, in a more exploratory way, support researchers in heuristically devising novel hypotheses. Therefore, in the following section, we elaborate on knowledge and learning with respect to their dynamic and collective nature.

## The dynamic and collective nature of knowledge

The *dynamic* nature of knowledge was a central aspect of Piaget's (1970, 1987) classic work: To cope with disturbances from new and changing environments, individuals first try to apply their existing internal cognitive schemas to new and changing environments (assimilation). If assimilation strategies fail, they have to change their existing schemas or acquire new ones (accommodation). Individuals constantly strive for an equilibrium between assimilation and accommodation. Too much assimilation would mean stagnation and too much accommodation would result in chaos (Piaget 1977).

Even though Piaget was fully aware of the importance of social interactions (see e.g., DeVries 1997), he focused primarily on interactions between an individual learner and the environment. So, regarding the social and *collective* nature of knowledge, one might rather refer to Vygotsky (e.g., 1934/1962) as another classical psychologist who focused more strongly on the socio-cultural underpinnings of learning. Together, Piaget and Vygotsky influenced scores of researchers in fields as different as sociology, political science, psychology, and education. Whereas one line of research further pursued the idea of education as a process of scaffolding (e.g., Bruner 1996), other researchers elaborated upon the concept of *situated learning* (e.g., Brown et al. 1989; Greeno 1997). In this approach learning is not considered as the acquisition of abstract decontextualized knowledge items, but as the development of complex skills comprised of *knowing* as well as *doing,* achieved by means of problem solving and communication activities within concrete application settings. An individual is enculturated into a community, that is, a group of people who are connected through a common activity and who share their knowledge—thereby learning from each other (Greeno 1997). The act of learning from each other provides the group also with the means to continuously refine its collective knowledge and to engage in the creation of new knowledge.

Over the last two decades, the concept of *knowledge creation* (Paavola et al. 2004) has become popular in research on learning in professional (e.g., Nonaka 1991, 1994) and educational settings (e.g., Engeström 1999). In all of these approaches, the "pursuit of newness" (Paavola et al. 2004, p. 562) is behind learning processes: Knowledge is not some *object* that can be acquired; it is the collaborative *creation* of something new by means of the communication of experiences and subsequent attempts to put ideas into practice. Hakkarainen and colleagues (Hakkarainen and Paavola 2009; Hakkarainen et al. 2009) further developed the knowledge creation concept into their *trialogical approach to learning*. Here, the key entities are learners, communities, and objects. In a collaborative knowledge creation community, different types of artifacts make the exchange of ideas possible: *conceptual objects*, for example in the form of questions, theories, and designs; *material objects*, for instance in the form of collaboratively written documents; and finally, *procedural objects,* such as certain norms and behavioral scripts. These objects may also serve a mediation function between individual and collective activities. With the trialogical approach, (digital) artifacts have increasingly moved into the focus of research on the dynamics of learning and knowledge construction.

The question is whether the existing methods that are commonly used in these research traditions are sufficient for adequately analyzing complex knowledge dynamics in mass collaboration online environments. For such an analysis, one has to simultaneously take into account individual learning trajectories and learning outcomes, macro processes of social knowledge, and the close ties and interactions between individual learning and collective knowledge construction. The co-evolution model of individual learning and collective knowledge construction (Cress and Kimmerle 2008, 2017; Kimmerle et al. 2015) combines the individual and the social levels. It predicts that conflicts between individual and collective perspectives (as presented in collaboratively created artifacts) may make the two systems involved drift over time: In the case of the cognitive system, the drift can be called learning; in the case of the social system, this drift constitutes knowledge construction. Both systems are *structurally coupled* (Maturana 1975) insofar as they co-evolve toward greater capabilities in handling complexity.

The key entities within the co-evolution of learning and knowledge construction in online environments are persons and (digital) artifacts. Persons can be related to artifacts in that they read, view, share, author, or edit the artifacts. All forms of behavior that can at least potentially be perceived by others constitute acts of communication, which are traceable in the artifacts. People who make use of a knowledge resource in the form of an electronically provided artifact by dealing with this artifact constitute a community of individuals who share an interest in the underlying topic(s), and who are willing to share their ideas with other community members, or to deal with other members' ideas. As a consequence, what may result from such active participation in collaboration is the co-construction of knowledge (Leseman et al. 2000). However, in artifact-mediated collaboration, people will not necessarily engage only in shared activities; quite the contrary, they have to spend time and energy coordinating their contributions and handling disagreement (Matusov 2001). Using the example of wikis for collaborative learning, Pifarré and Kleine Staarman (2011) have shown that this coordination and handling of conflicts can be accomplished through social interaction in a "dialogic space" (see also Wegerif 2007).

In the following section, we present three different methodological approaches to address these theoretical considerations on the dynamic and collective nature of knowledge, the interplay of the individual and the collective levels, and the role of socio-cognitive conflicts and productive friction. The overarching research question is whether learning processes on the individual level as well as knowledge construction processes on the collective level can be predicted and explained by productive friction. All of the studies make use of authentic behavior traces in Wikipedia, such as articles' and users' histories of previous edits.

The need to compliment studies that use self-reports in questionnaires as their primary data sources with studies using" actual" behavior data has been formulated frequently over the last years in different subfields of psychology (e.g., Baumeister et al. 2007; Serfass et al. 2017). In the studies that we are going to discuss, behavior traces such as edit histories are used to operationalize abstract constructs such as (productive) friction in the form of different viewpoints or different knowledge regarding a certain topic. By discussing the methods applied in these studies against this theoretical background we aim to show how these methods can be used for examining the interplay of the individual and the collective levels that drive learning as well as knowledge construction.

## Productive friction in Wikipedia

In the following paragraphs we introduce three methods of analysis for identifying productive friction in Wikipedia: (1) Statistical analyses based on automatic text classification, (2) social network analysis, and (3) clustering methods. For each of these three methods we first provide general considerations of opportunities for their use in education research and then present in more detail a Wikipedia-related study as an illustrating example for application.

### Automatic text classification

**Text classifiers in education research** Text classifiers automatically assign documents (or segments of documents) to predefined categories. Technically speaking, they are models that predict the class value of documents based upon a certain number of attributes (Hämäläinen and Vinni 2010). In unsupervised machine learning, algorithms similar to those that are used in cluster analysis are applied first to detect a reasonable number of classes (Mohri et al. 2012). In supervised machine learning, a pre-coded training set is used to derive the ideal combination of attributes to classify the texts as precisely as possible: A substantial number of documents that are comparable to the texts to be analyzed are assigned to the categories in question by trained coders or some pre-established coding rule. Classifiers then use machine learning techniques such as *support vector machines*, *decision trees*, or (*naïve*) *Bayesian classifiers* to "learn" the ideal combination of predictors which mirror the coders' assignment of documents.

In education research, automatic text classification is frequently being used in studies on the communication among learners in e-learning and blended learning environments (e.g., Chen et al. 2014; Dascalu et al. 2010; Dascalu et al. 2015). For example, Dascalu et al. (2015) developed the computational model *ReaderBench* to automatically identify different instances of collaboration from chat messages in a tool for CSCL. This automatically generated information on collaboration processes can in turn be used in learning analytics systems, for example, to facilitate an exchange of ideas. Machine learning is also used in the context of recommender systems that aim at providing learners with the most useful learning materials according to their previous learning trajectory (e.g., Drachsler et al. 2015; Kopeinik et al. 2016; Moskaliuk et al. 2011).

One challenge when employing automatic classifiers in education is to avoid the potential dangers of overfitting: The more researchers attempt to exploit all the information that is in their training set as means of increasing the accuracy of the model, the more they run into the danger of creating a highly specialized model that only works well with the data that were used for training and is not transferable to different cases. A possible counterstrategy is to limit the number of attributes and features by means of diligent feature selection and feature extraction (Hämäläinen and Vinni 2010).

### Example 1:

**Knowledge dynamics in Wikipedia articles** In a recent article, Jirschitzka et al. (2017) devised a study based on text classification investigating the role and function of productive friction in learning and knowledge construction on the Wikipedia platform. The study focused on the struggle between proponents of alternative and conventional medicine. All of the edits of all of the articles within the categories of alternative medicine and nutrition in the German language Wikipedia were sampled; a total of more than 70.000 edits from 398 articles were crawled, processed, and analyzed. For further analysis, a supervised machine learning algorithm was trained to distinguish all modifications into pro alternative medicine, pro conventional medicine, or neutral modifications. Using the terms of the trialogical approach to learning, we could say that *material objects* (edits) were used to operationalize *conceptual objects* (attitudes/knowledge) and *procedural objects* (contribution patterns) regarding alternative and conventional medicine (Hakkarainen et al. 2009). Based on these scores, a summary score for all edits (usually comprised of a number of modifications) was calculated that indicated to what extent they displayed a view in favor of or against alternative medicine, or a neutral view. Hence, edits of digital artifacts were used to reconstruct knowledge and attitudes on the level of individuals and on an aggregated social level. Based on all of the edits of an individual contributor, the respective *user's knowledge trajectory* was reconstructed. Analogously, the *knowledge trajectory of each article* was calculated by aggregating all of the contributions to the article, regardless of which authors they came from. Differences between an author's and an article's view at any given moment were used as a proxy of friction.

The studies showed a typical confirmation bias (Jonas et al. 2001): Wikipedia contributors primarily edited those articles that were consistent with their own view on the topic (see Fig. 1). However, indicators for *productive* friction could also be identified: In the case of pro alternative medicine articles, the articles were more balanced when they were edited by more heterogeneous contributors, that is, contributors who displayed both pro and contra alternative medicine attitudes. It seems that the presence of contributors from different backgrounds and with different attitudes toward medicine are a means of preventing a possible polarization of views (Bakshy et al. 2015). The so-called *echo chamber effect* (Del Vicario et al. 2016) refers to the observation that the ubiquitous availability of any kind of information on the Internet and the possibility of bonding with like-minded individuals via social networks can lead groups to isolate themselves from the rest of society. Within these groups, only information that confirms the group's beliefs is shared, which can lead in the long run to increasing polarization of group opinions and norms.

These results mirror findings from previous lab studies (Schwind et al. 2012) insofar as they show human beings' general preference for information that confirms their opinions, attitudes, and ideologies. Nevertheless, it still seems that productive friction between one's own views



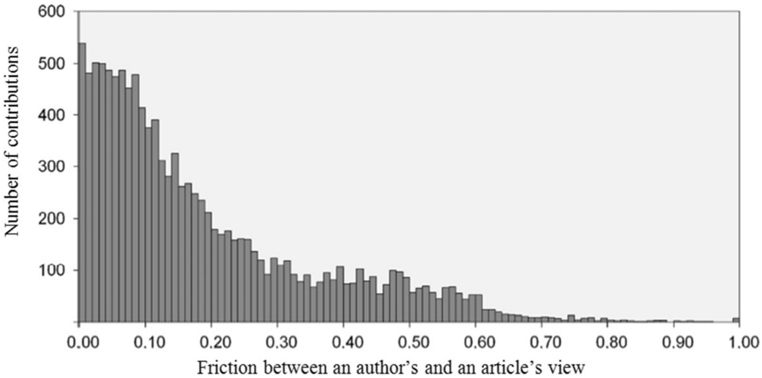


**Fig. 1** Number of contributions as a function of the friction between an author's and an article's view (adapted from Jirschitzka et al. 2017)

and others' views is necessary for learning and successful knowledge construction. Platforms such as Wikipedia that explicitly aim at incorporating different viewpoints enable some degree of exposure to others' perspectives, which may be supportive for the development of knowledge (Oeberst et al. 2014), even though it is not a guarantee for successful knowledge construction (Greenstein et al. 2016; Holtz et al. 2018).

## Social network analysis

**Social network analysis in education research** Social network analysis (SNA) is the analysis of social structures by means of studying the relations (edges, links, or ties) among individual actors—or *nodes* in the language of SNA—such as persons, or things like digital artifacts, in a linked environment. Relations between nodes can be established based on similarities (e.g., same location, similar attributes), social relations (e.g., kinship or friendship), social interactions (e.g., communication), or flows (e.g., flows of information or goods) between nodes (Borgatti et al. 2009; Wasserman and Faust 1994).

Whereas SNA emerged from more descriptive sociometric approaches in the 1930s and 1940s, more currently the emergence of modern information technology has facilitated the analysis and modeling of vast amounts of interrelated data drawing upon elaborated mathematical graph theories or network theories (Borgatti et al. 2009). This makes it possible to calculate mathematical indicators of certain network structures and properties of nodes and ties within a network. For example, the centrality of a certain node or group of nodes can be assessed to quantify the relative importance of these nodes for a given network structure (e.g., Freeman 1979). In education research, SNA has been used mainly to analyze the role and function of study networks and other interactional social structures within classrooms or educational institutions, such as schools and universities (e.g., Brunello et al. 2010; Grunspan et al. 2014). Bruun and Brewe (2013) found, for example, a correlation between students' embeddedness into communication structures and their academic achievement. CSCL has evolved as another major field of application of SNA in education (e.g., De Laat et al. 2007; Wellman 2001). In the case of collaboratively written documents with links between subsections and/or articles, there is a tri-partite network where people as well as articles or their subsections can serve as nodes; edges exist in the form of author-author relations, author-article relations, or article-article relations.

A challenge for the use of SNA in education research consists of capturing the inherent dynamics in the analyzed networks, for example, in the form of varying collaboration structures as a consequence of changing friend networks and educational settings (Borgatti et al. 2009). One possible strategy to capture changes in network structures, that is used in the following example, comprises capturing "snapshots" of structures at different points in time.

### Example 2:

**The role of boundary spanners for knowledge construction in Wikipedia** A series of studies (Halatchliyski et al. 2014; Halatchliyski and Cress 2014) used SNA to identify the places where new knowledge is created in Wikipedia and to highlight the persons creating this knowledge. The studies focused on the Wikipedia domains *psychology* and *education* and used snapshots of the link structure among related articles for the years 2006 to 2012. There is a substantial overlap between the two domains, for example in the form of articles on educational psychology. But these domains were chosen because they each also clearly represent separate entities. All of the articles analyzed (approximately 11.000) fell either into one of the two categories or into both. Figure 2 presents a visual representation of the distinctness and the partial overlap between the two domains by means of plotting all hyperlinks (edges) between the education articles (grey dots), psychology articles (white dots), and the intersection articles (black dots).

The main hypothesis was that new knowledge would be primarily created in the immediate vicinity of pivotal articles. Pivotal articles are either central articles *within* a domain, or articles that serve as boundary spanners *between* different domains. The boundary spanning articles represented friction whenever information from both domains needed to be integrated. A centrality measure was used to assess each article's importance within a domain, whereas a betweenness measure was used as a proxy for productive friction (see Freeman 1979). A first analysis (Halatchliyski and Cress 2014) revealed that new knowledge in the form of edits or newly created articles did indeed center either around articles with high levels of centrality within the respective domains, or around articles with high betweenness scores. These results show that friction can indeed be productive. New knowledge emerges at those points where friction has occurred, that is, in the nodes that span different domains. New knowledge also develops in the center of a domain, where friction supposedly results from the central article being connected to a high number of other articles within the domain, which increases the number of potential starting points for friction.

Further analysis (Halatchliyski et al. 2014) revealed that both kinds of articles, those with high centrality scores (friction within the domain) and those with high betweenness scores (friction between domains), were authored by more experienced users (authors having made many edits before). An analysis of the contributors' previous edits revealed that articles with high centrality scores, which were centerpieces of the respective domains of psychology or education, were primarily edited by *specialists* who were only active in one of the domains. In comparison, articles with high betweenness scores, which spanned the two domains, were primarily authored by *generalists* who regularly contributed to both domains.

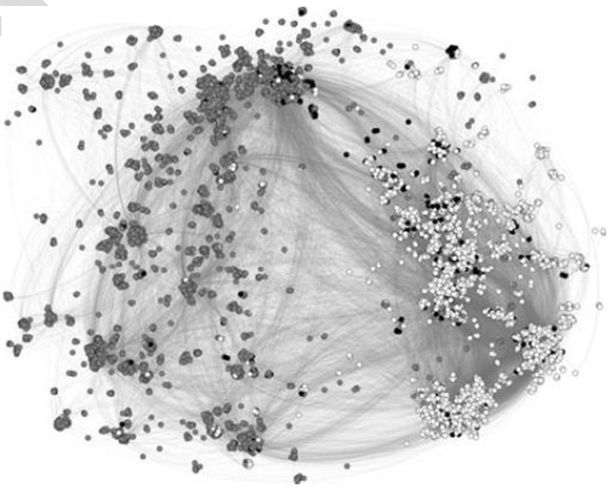

**Fig. 2** The network of Wikipedia articles in psychology and education (from Halatchliyski et al. 2014)

It seems that two kinds of knowledge construction processes could be identified here: One could be observed as an increase of knowledge in very specialized areas at the center of each respective domain. This consolidation or refinement of existing knowledge was driven particularly by specialists in the respective areas. The other knowledge construction process took the form of creation of conceptually new knowledge or enrichment of existing knowledge between the two domains by generalists who were well versed in both areas and who acted as boundary spanners (see also Tushman and Scanlan 1981). With regard to Piaget's distinction between assimilation and accommodation, the first form of knowledge construction resembles assimilation insofar as experienced specialized authors improve knowledge within a domain by means of refining central articles. The second type of knowledge construction resembles accommodation, as new knowledge is created in boundary spanning articles by boundary spanning authors, thereby resolving friction between domains in a productive way.

## Clustering techniques

**Clustering techniques in education research** Big data is by definition weakly structured, and cluster analytic techniques are often used in a first exploratory step to identify patterns in these data structures. This is done by using one of several algorithms to identify relatively homogeneous agglomerations of cases (for example, learners or learning resources), based on shared attributes (Antonenko et al. 2012; Kalota 2015). The identification of clusters of learners can help to understand the learning processes in different sub-groups of a learner population, which otherwise could be missed by the researchers (Wise and Shaffer 2015). For example, Rysiewicz (2008) used cluster analysis to identify subgroups of 13-year-old foreign language students who reacted differently to a range of teaching approaches. Kizilcec et al. (2013) used cluster analytic techniques to identify different subgroups of learners in online courses with regard to the degree of their engagement and activity trajectories. These results can in turn be used to set up an automatic classification algorithm that can identify potentially detrimental disengagement processes. This same algorithm can be applied to deploy communicative counter-measures before they lead to unwanted outcomes, such as a student leaving the course.

When applying clustering methods, researchers should be aware that automatic classifiers, like all clustering methods, can only be as valid as the variables that are entered into the analyses. Researchers should also be aware that the number and the composition of the clusters can easily change dramatically with entered or dismissed variables and cases (Sarstedt and Mooi 2019).

Cluster analytic methods are also being used to group learning resources that share common features for their use in recommender systems and intelligent curricula (Kalota 2015). For example, Wang et al. (2008) used semantic features to cluster e-learning materials to facilitate topic-related searches. Kumar et al. (2015) recently proposed algorithms for the clustering of social tags as a means of facilitating the classification and partly automatized retrieval of digital artifacts. SNA, as described above, can also be used to identify clusters of entities that have particular connections in common, such as documents which share a common link structure (e.g., Wan and Yang 2008). This last approach was used in our third example below in order to Q4 identify and visualize the development of clusters in collaborative knowledge construction over a period of several years.

**Example 3:**

**Identifying clusters of Wikipedia articles** The following study (Kimmerle et al. 2010) addressed the topic of possible social or biological causes of schizophrenia (see also Harrer et al. 2008; Moskaliuk et al. 2009, 2012). The aim of this study was to identify similarities and differences—and thereby friction—among the respective Wikipedia articles by means of clustering the results of an SNA. For the analysis, the link structures of the Wikipedia article on schizophrenia and related articles that represented various explanatory approaches (social, biological, and psychoanalytical) were analyzed in a series of six annual cross-sectional snapshots for the years 2003–2008. The *Weaver* software (Harrer et al. 2007) was then used to calculate SNA traits, such as centrality and density, for all of the Wikipedia pages that were linked at that time to these articles. Furthermore, SNA also allowed for the calculation of scores of the individual contributors with regard to the closeness or distance of their linking activities to the nodes that represented different views on the causes of schizophrenia. This process operationalized friction for individuals as instances where editors contributed to articles that represented a different opinion than their own.

Overall, the network constantly became more complex from 2003 to 2008, that is, the number of links between the pages increased continuously. In 2005, two clusters of closely interlinked articles appeared: One cluster was related to the psychoanalytic theory of schizophrenia and one was related to biological explanations of schizophrenia. In 2007, a third cluster representing social aspects emerged. Nevertheless, as can be seen in Fig. 3, the social cluster was closely related to the biological cluster via a number of *boundary-spanning*



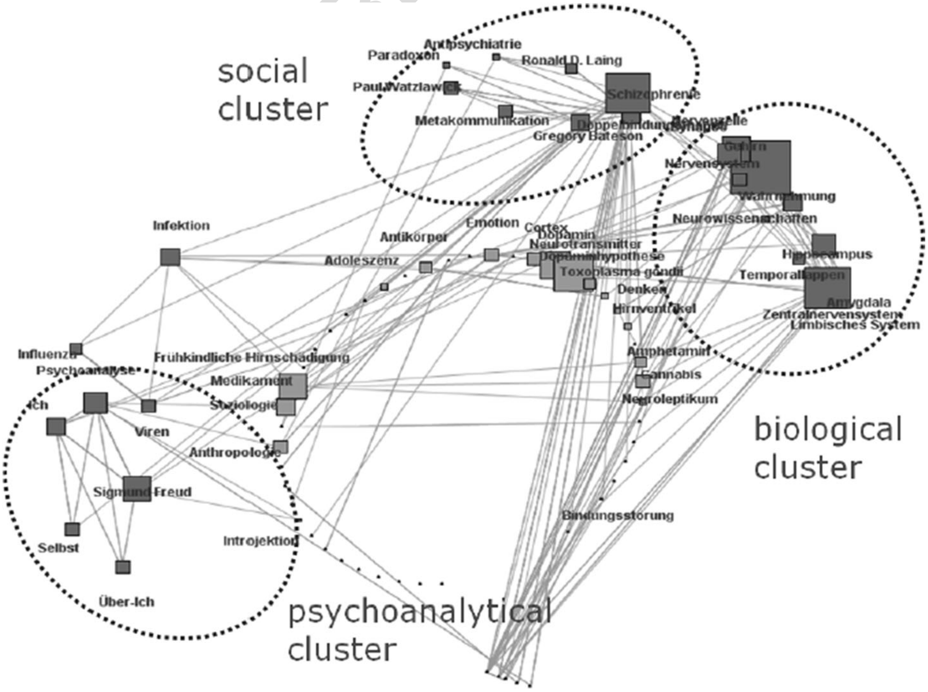


**Fig. 3** Clusters of Wikipedia articles representing a psychoanalytical, a biological, and a social perspective Q5

articles. Later, in 2008, the social and the biological clusters even merged into one common cluster which mirrored the so-called diathesis-stress model, currently the most accepted explanation model for the genesis of schizophrenia. This model explicitly tries to incorporate biological and social causes of schizophrenia into a single model (Walker and Diforio 1997). In comparison, the psychoanalytical cluster remained separated and isolated throughout the observation period with relatively few boundary-spanning articles. As for the authors, it was found that those who endorsed social or biological explanations shifted more and more toward the integrative diathesis-stress model, whereas those who endorsed the psychoanalytical explanations kept editing primarily the articles within their own cluster. Thus, these findings also indicate that development of collective knowledge may take place particularly when people with diverse backgrounds interact in a way that encourages response to productive friction. However, when people solely stick to their pre-existing beliefs and refuse to take other perspectives into account, it is not likely they will make any relevant contribution to the development of knowledge.

## Discussion

### Summary and conclusion

In the tradition of constructivists such as Piaget and Vygotsky, we argue that learning in the *age of social media* is still often the result of learners successfully handling disturbances in the form of solving socio-cognitive conflicts. We have also pointed out that complex forms of learning and knowledge construction can only be explained by taking the interplay between individuals and their social environment into account. In line with earlier studies on knowledge creation (e.g., Engeström 1999; Hakkarainen et al. 2009; Nonaka 1991, 1994), we assume that exchanges between people holding differing attitudes toward a given topic or having different knowledge about something are what drives the construction of new knowledge. These processes on the social level are intrinsically entwined with individuals' learning processes.

The rise of social media has opened up the way for new research methods as a means of studying actual behavior traces in digital environments. In this article, we have discussed a series of studies that used various analysis methods to deepen our understanding of the role and the function of productive friction in Wikipedia as an eminent real-world collaborative knowledge construction environment. The studies that we have presented here have used different methods to operationalize the concept of productive friction: The first method was to categorize texts and authors automatically by identifying different semantics used in the artifacts (via automatic text classification); the second approach applied different measures of centrality in a network of artifacts and authors; and the third method identified which clusters articles and authors belonged to within a network of interrelated articles. All methods allowed for describing the dynamics (i.e., the development over time) involved on the individual level as well as on the level of the artifacts that represents the collective level of knowledge construction. Based on the co-evolution model, all of the sample studies cited aimed to identify the role of productive friction for learning and knowledge construction, and examined how these processes may be predicted through the occurrence of productive friction.

## Implications for research in the learning sciences and CSCL

Socio-cognitive conflicts and productive friction are indispensable for learning and knowledge construction. Consequently, the emergence of new knowledge frequently occurs along the fault-lines of knowledge structures; in websites for collaborative knowledge construction, for example, knowledge is often developed as a consequence of experts from different domains discussing their respective ideas (Halatchliyski et al. 2014). Furthermore, a certain degree of variety in terms of attitudes and knowledge seems to alleviate unwanted biases and radicalization processes in such platforms (Jirschitzka et al. 2017).

These findings are relevant as well for the future development of learning analytics and the design of interactive learning environments: In line with the aforementioned approaches by, among others, Scardamalia and Bereiter (1994) and Hakkarainen and colleagues (e.g., Hakkarainen et al. 2009), fostering, and if necessary, enforcing socio-cognitive conflicts that lead to a productive solution can be a fruitful strategy to optimize learning outcomes. A similar idea has been investigated recently in the realm of mathematics teaching under the term *productive failure* (e.g., Kapur 2008): In the acquisition of mathematical knowledge it is often important for learners to first experience certain typical failures as means of setting the ground for fully understanding the correct solution. Enabling students to first try out collaboratively different (most often false) solutions before facilitating the "discovery" of the correct solution can be a very effective way to convey mathematical knowledge. Future research will have to show in how far similar approaches can be useful for large scale learning and knowledge construction environments as well.

In collaborative knowledge construction and learning environments, the prevention of the emergence of closely knit communities with very homogeneous attitudes toward certain topics can be a means of reducing biases by increasing productive friction (Jirschitzka et al. 2017). More research is needed to further understand the dynamics that unfold in homo- and heterogeneous knowledge communities and to disentangle the specific effects of certain forms of digital environments. Learning in digital environments often occurs as casual and partially unintended "everyday learning "in social media (D'Aquin et al. 2017). Certainly, many of the theories and constructs that were developed within more traditional learning environments will also be valid and useful for learning in the age of social media. Still, the newly available data sources and analysis techniques will allow for refinement and further development of these theories and approaches. Additionally, modern information environments also create challenges and problems that require new approaches and solutions. For example, we can safely assume that different forms of automatic recommender systems strongly impact learning outcomes (e.g., Kimmerle et al. 2017a, b). Here as well, studies analyzing the effects of singular elements of such recommender systems in isolation in laboratory studies can and should be combined with studies using large amounts of actual behavior data that is available from online learning and knowledge construction platforms such as Wikipedia.

In view of the increasing importance of social media resources for learning, there is a growing need for theories and models that help to explain, predict, and counteract the development of distortions, such as biases or misinformation. Another aspect that should be taken into account more strongly in future studies is the role emotions might play in cognitive processes. For example, recent studies found that Wikipedia articles on man-made disasters contain more and different emotional words than articles on natural catastrophes (Greving et al. 2018). We assume that processes of emotion regulation are also intrinsically entwined with the socio-cognitive processes that we have outlined. The relationship between emotional and socio-cognitive processes should be addressed more explicitly in further CSCL research.

The incredible reams of data that are available from platforms such as Wikipedia enable learning scientists and other researchers in the field of education to test their theories in real-life, large-scale scenarios and thus not only in laboratory settings or small groups. Research in real-life scenarios is needed to ensure the external validity of findings from laboratory studies (Cook and Campbell 1979). Some phenomena, like the structural coupling of individual (learning) and social or societal processes (knowledge construction) over substantial periods of time, cannot be easily mapped into laboratory experiments without the risk of missing out on the very essence of what is going on. With digital behavior traces such as editing or browsing histories, even very complex temporal trajectories can be restored at any given point in time as long as the data is available. Still, the results of such studies can only contribute to the growth of scientific knowledge if they can be used to test existing theories, or if they facilitate the development of new and better theories.

Important methods to consider in this context that have been discussed in the paper presented here are methods based on machine learning text classification, SNA, and SNA-based clustering. One problem that this kind of research faces is that social scientists are often not trained in their university studies to use these methods properly, whereas data scientists similarly have little training in relevant educational, psychological, and social scientific theories. Hence, for the moment, strategic cooperation between these two groups of experts is needed to tackle the relevant questions of our day with the most suitable methods (Harlow and Oswald 2016).

**Acknowledgements** This research was supported in part by a grant from the European Commission's Research and Innovation Programme 'Horizon 2020' (AFEL – Analytics for Everyday Learning; 687916).





## References

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 60*, 383–398.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*, 1130–1132.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396–403.

Bientzle, M., Cress, U., & Kimmerle, J. (2013). How students deal with inconsistencies in health knowledge. *Medical Education, 47*, 683–690.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*, 892–895.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Brunello, G., De Paola, M., & Scoppa, V. (2010). Peer effects in higher education: Does the field of study matter? *Economic Inquiry, 48*, 621–634.

Bruner, J. S. (1996). *The culture of education*. Cambridge: Harvard University Press.

Bruun, J., & Brewe, E. (2013). Talking and learning physics: Predicting future grades from network measures and force concept inventory pretest scores. *Physical Review Special Topics - Physics Education Research, 9*, 020109.

Castells, M. (2010). *End of millennium: The information age: Economy, society, and culture*. Oxford: Wiley-Blackwell.

Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies, 7*, 246–259.

Coase, R. H. (1937). The nature of the firm. *Economica, 4*, 386–405.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally College Publishing Company.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning, 3*, 105–122.

Cress, U., & Kimmerle, J. (2017). The interrelations of individual learning and collective knowledge construction: A cognitive-systemic framework. In S. Schwan & U. Cress (Eds.), *The psychology of digital learning* (pp. 123–145). Cham: Springer International Publishing.

Cress, U., Feinkohl, I., Jirschitzka, J., & Kimmerle, J. (2016). Mass collaboration as co-evolution of cognitive and social systems. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education* (pp. 85–104). Cham: Springer International Publishing.

D'Aquin, M., Adamou, A., Dietze, S., Fetahu, B., Gadiraju, U., Hasani-Mavriqi, I., Holtz, P., Kimmerle, J., Kowald, D., Lex, E., López Sola, S., Maturana, R. A., Sabol, V., Troullinou, P., & Veas, E. (2017). AFEL: Towards measuring online activities contributions to self-directed learning. In *Proceedings of the 7th Workshop on Awareness and Reflection in Technology Enhanced Learning (ARTEL).*

Darnon, C., Doll, S., & Butera, F. (2007). Dealing with a disagreeing partner: Relational and epistemic conflict elaboration. *European Journal of Psychology of Education, 22*, 227–242.

Dascalu, M., Rebedea, T., & Trausan-Matu, S. (2010). A deep insight in chat analysis: Collaboration, evolution and evaluation, summarization and search. In *International conference on artificial intelligence: Methodology, systems, and applications* (pp. 191–200). Berlin: Springer.

Dascalu, M., Trausan-Matu, S., McNamara, D. S., & Dessus, P. (2015). ReaderBench: Automated evaluation of collaboration based on cohesion and dialogism. *International Journal of Computer-Supported Collaborative Learning, 10*, 395–423.

De Laat, M., Lally, V., Lipponen, L., & Simons, R. J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis. *International Journal of Computer-Supported Collaborative Learning, 2*, 87–103.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., et al. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences, 113*, 554–559.

DeVries, R. (1997). Piaget's social theory. *Educational Researcher, 26*(2), 4–17.

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning, 12*, 17–26.

Doise, W., Mugny, G., & Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5*, 367–383.

Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of recommender systems to support learning. In F. Ricci, F. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 421–451). New York: Springer.

Engeström, Y. (1999). Expansive visibilization of work: An activity-theoretical perspective. *Computer Supported Cooperative Work (CSCW), 8*, 63–93.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*, 304–317.

Fosnot, C. T., & Perry, R. S. (1996). Constructivism: A psychological theory of learning. *Constructivism: Theory, Perspectives, and Practice, 2*, 8–33.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215–239.

Gasevic, D., Rosé, C., Siemens, G., Wolff, A., & Zdrahal, Z. (2014). Learning analytics and machine learning. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 287–288). New York: ACM.

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*, 5–17.

Greenstein, S., Gu, Y., & Zhu, F. (2016). *Ideological segregation among online collaborators: Evidence from Wikipedians (no. w22744)*. Washington DC: National Bureau of Economic Research.

Greving, H., Oeberst, A., Kimmerle, J., & Cress, U. (2018). Emotional content in Wikipedia articles on negative man-made and nature-made events. *Journal of Language and Social Psychology, 37*, 267–287.

Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE Life Sciences Education, 13*, 167–178.

Hagel 3rd, J., & Brown, J. S. (2005). Productive friction: How difficult business partnerships can accelerate innovation. *Harvard Business Review, 83*, 82–91.

Hakkarainen, K., & Paavola, S. (2009). Toward a trialogical approach to learning. In B. Schwarz, T. Dreyfus, & R. Hershkowitz (Eds.), *Transformation of knowledge through classroom interaction* (pp. 65–80). London and New York: Routledge.

Hakkarainen, K., Engeström, R., Paavola, S., Pohjola, P., & Honkela, T. (2009). Knowledge practices, epistemic technologies, and pragmatic web. In A. Paschke, H. Weigand, W. Behrendt, K. Tochtermann, & T. Pellegrini (Eds.), *Proceedings I-SEMANTICS 5* (pp. 683–694). Graz: TU Graz.

Halatchliyski, I., & Cress, U. (2014). How structure shapes dynamics: Knowledge development in Wikipedia: A network multilevel modeling approach. *PLoS One, 9*, e111958.

Halatchliyski, I., Moskaliuk, J., Kimmerle, J., & Cress, U. (2014). Explaining authors' contribution to pivotal artifacts during mass collaboration in the Wikipedia's knowledge base. *International Journal of Computer-Supported Collaborative Learning, 9*, 97–115.

Hämäläinen, W., & Vinni, M. (2010). Classifiers for educational data mining. In C. Romero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of Educational Data Mining,* Chapman & Hall/CRC Data Mining and Knowledge Discovery Series (pp. 57–71). Boca Raton: CRC press.

Happer, C., & Philo, G. (2013). The role of the media in the construction of public belief and social change. *Journal of Social and Political Psychology, 1*, 321–336.

Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology: Introduction to the special issue. *Psychological Methods, 21*, 447–457.

Harrer, A., Zeini, S., Ziebarth, S., & Münter, D. (2007). *Visualisation of the dynamics of computer-mediated community networks*. Paper presented at the International Sunbelt Social Network Conference.

Harrer, A., Moskaliuk, J., Kimmerle, J., & Cress (2008). Visualizing wiki-supported knowledge building: Co-evolution of individual and collective knowledge. *Proceedings of the International Symposium on Wikis* 2008 *(Wikisym)*. New York: ACM Press. Q8

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend: Decoding the social media news consumer. *Journalism Studies, 13*, 815–824.

Holtz, P., Fetahu, B., & Kimmerle, J. (2018). Effects of contributor experience on the quality of health-related Wikipedia articles. *Journal of Medical Internet Research, 20*, e171.

Jeong, H., Cress, U., Moskaliuk, J., & Kimmerle, J. (2017). Joint interactions in large online knowledge communities: The A3C framework. *International Journal of Computer-Supported Collaborative Learning, 12*, 133–151.

Jirschitzka, J., Kimmerle, J., Halatchliyski, I., Hancke, J., Meurers, D., & Cress, U. (2017). A productive clash of perspectives? The interplay between articles' and authors' perspectives and their impact on Wikipedia edits in a controversial domain. *PLoS One, 12*, e0178985.

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80*, 557–571.

Kalota, F. (2015). Applications of big data in education. *International Journal of Social, Behavioral, Educational, Economic, Business and. Industrial Engineering, 9*, 1570–1575.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*, 379–424.

Kimmerle, J., Moskaliuk, J., Harrer, A., & Cress, U. (2010). Visualizing co-evolution of individual and collective knowledge. *Information, Communication & Society, 13*, 1099–1121.

Kimmerle, J., Moskaliuk, J., Oeberst, A., & Cress, U. (2015). Learning and collective knowledge construction with social media: A process-oriented perspective. *Educational Psychologist, 50*, 120–137.

Kimmerle, J., Bientzle, M., & Cress, U. (2017a). "Scientific evidence is very important for me": The impact of behavioral intention and the wording of user inquiries on replies and recommendations in a health-related online forum. *Computers in Human Behavior, 73*, 320–327.

Kimmerle, J., Moskaliuk, J., Brendle, D., & Cress, U. (2017b). All in good time: Knowledge introduction, restructuring, and development of shared opinions as different stages in collaborative writing. *International Journal of Computer-Supported Collaborative Learning, 12*, 195–213.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the third International Conference on Learning Analytics and Knowledge* (pp. 170–179). New York: ACM Press.

Kopeinik, S., Kowald, D., & Lex, E. (2016). Which algorithms suit which learning environments? A comparative study of recommender systems in TEL. In *European conference on technology enhanced learning* (pp. 124–138). New York: Springer International Publishing.

Kumar, S. S., Inbarani, H. H., Azar, A. T., & Hassanien, A. A. (2015). Rough set-based meta-heuristic clustering approach for social e-learning systems. *International Journal of Intelligent Engineering Informatics, 3*, 23–41.

Lang, C., Siemens, G., Wise, A., & Gašević, D. (2017). *Handbook of learning analytics – First edition*. Online publication: https://solaresearch.org/hla-17/; last retrieved July 3 2017.

Leseman, P. P. M., Rollenberg, L., & Gebhardt, E. (2000). Co-construction in kindergartners' free play: Effects of social, individual and didactic factors. In H. Cowie & G. Van der Aalsvoort (Eds.), *Social interaction in learning and instruction: The meaning of discourse for the construction of knowledge*. Amsterdam: Pergamon/Elsevier Science Inc.

Maturana, H. R. (1975). The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies, 7*, 313–332.

Matusov, E. (2001). Intersubjectivity as a way of informing teaching design for a community of learners classroom. *Teaching and Teacher Education, 17*, 383–402.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. Cambridge: MIT Press.

Moskaliuk, J., Kimmerle, J., & Cress. (2009). Wiki-supported learning and knowledge building: Effects of incongruity between knowledge and information. *Journal of Computer Assisted Learning, 25*, 549–561.

Moskaliuk, J., Rath, A., Devaurs, D., Weber, N., Lindstaedt, S., Kimmerle, J., & Cress, U. (2011). Automatic detection of accommodation steps as an indicator of knowledge maturing. *Interacting with Computers, 23*, 247–255.

Moskaliuk, J., Kimmerle, J., & Cress, U. (2012). Collaborative knowledge building with wikis: The impact of redundancy and polarity. *Computers & Education, 58*, 1049–1057.

Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology, 8*, 181–192.

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review, 69*(6), 96–104.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*, 14–37.

Oeberst, A., Halatchliyski, I., Kimmerle, J., & Cress, U. (2014). Knowledge construction in Wikipedia: A systemic-constructivist analysis. *Journal of the Learning Sciences, 23*, 149–176.

Oeberst, A., Kimmerle, J., & Cress, U. (2016). What is knowledge? Who creates it? Who possesses it? The need for novel answers to old questions. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education* (pp. 105–124). Cham: Springer International Publishing.

Paavola, S., Lipponen, L., & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research, 74*, 557–576.

Piaget, J. (1970). *Structuralism*. New York: Basic Books.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. New York: Viking.

Piaget, J. (1987). *Possibility and necessity.* Minneapolis: University of Minnesota Press.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9–20.

Pifarré, M., & Kleine Staarman, J. (2011). Wiki-supported collaborative learning in primary education: How a dialogic space is created for thinking together. *International Journal of Computer-Supported Collaborative Learning, 6*, 187–205.

Resnick, L. B. (1991). Shared cognition: Thinking as social practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 1–20). Washington, DC: American Psychological Association.

Rysiewicz, J. (2008). Cognitive profiles of (un)successful FL learners: A cluster analytical study. *The Modern Language Journal, 92*, 87–99.

Sarstedt, M., & Mooi, E. (2019). Cluster analysis. In M. Sarstedt & E. Mooi (Eds.), *A concise guide to market research* (301–354). Berlin & Heidelberg: Springer.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*, 265–283.

Schwind, C., Buder, J., Cress, U., & Hesse, F. W. (2012). Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education, 58*, 787–796.

Serfass, D., Nowak, A., & Sherman, R. (2017). Big data in psychological research. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational social psychology (pp. 332–348). New York*. London: Routledge.

Shook, N. J., & Fazio, R. H. (2009). Political ideology, exploration of novel stimuli, and attitude formation. *Journal of Experimental Social Psychology, 45*, 995–998.

Siemens, G., & d Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 252–254). New York: ACM Press.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–32.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*, 1510–1529.

Tushman, M. L., & Scanlan, T. J. (1981). Boundary spanning individuals: Their role in information transfer and their antecedents. *The Academy of Management Journal, 24*, 289–305.

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems, 5*, 187–196.

Vygotsky, L. (1962/1934). *Thought and language.* Cambridge: MIT Press.

Walker, E. F., & Diforio, D. (1997). Schizophrenia: A neural diathesis-stress model. *Psychological Review, 104*, 667–685.

Wan, X., & Yang, J. (2008). Multi-document summarization using cluster-based link analysis. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 299–306). New York: ACM Press.

Wang, W. H., Zhang, H. W., Wu, F., & Zhuang, Y. T. (2008). Large scale of e-Learning resources clustering with parallel affinity propagation. In J. Fong, E. Kwan, & F. Lee (Eds.), *Proceedings of the International Conference on Hybrid Learning (ICHL)* (pp. 30–35). Hong Kong: ICHL.

Ward, C. J., Nolen, S. B., & Horn, I. S. (2011). Productive friction: How conflict in student teaching creates opportunities for learning at the boundary. *International Journal of Educational Research, 50*, 14–20.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications (Vol. 8)*. Cambridge: Cambridge University Press.

Wegerif, R. (2007). *Dialogic education and technology*. New York: Springer.

Wellman, B. (2001). Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research, 25*(2), 227–252.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics, 2*(2), 5–13.

# AUTHOR QUERIES

**AUTHOR PLEASE ANSWER ALL QUERIES.**

Q1. Please check if the section headings are assigned to appropriate levels.

Q2. The citation "Kimmerle et al. 2017" has been changed to "Kimmerle et al., 2017a, b" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.

Q3. The citation "Hämäläinen and Vinni 2011" has been changed to "Hämäläinen and Vinni, 2010" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.

Q4. The citation "Wan and Yan 2008" has been changed to "Wan and Yang, 2008" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.

Q5. Figure 3 contains poor quality of text inside the artwork. Please do not re-use the file that we have rejected or attempt to increase its resolution and re-save. It is originally poor, therefore, increasing the resolution will not solve the quality problem. We suggest that you provide us the original format. We prefer replacement figures containing vector/editable objects rather than embedded images. Preferred file formats are eps, ai, tiff and pdf.

Q6. Please check captured Acknowledgements if appropriate.

Q7. References [Coase, 1937, Fosnot & Perry, 1996, Scardamalia, 2002, Shook & Fazio, 2009, Siemens & Baker, 2012, Varela et al, 1974] were provided in the reference list; however, this was not mentioned or cited in the manuscript. As a rule, all references given in the list of references should be cited in the main body. Please provide its citation in the body text.

Q8. Please check captured author names of this reference if correct.