Collaboration has long been the subject of scholarly inquiry to test the assertion that “two heads are better than one.” Characterizing how and when learning happens as people work together has vexed researchers across a number of fields, including education, psychology, and business. A contemporary understanding of the social nature of learning and the power of the Internet to connect people over time and space, coupled with the recognition that collaboration is an essential competency in the modern workforce, continues to keep this topic salient—if not essential—for an educated and productive society.

In the field of learning analytics, the context for investigating collaboration is often, unsurprisingly, collaborative learning, which has been the focus of Computer-Supported Collaborative Learning (CSCL)—a scholarly community that was launched in the 1990s to investigate collaborative learning in computer-mediated settings [10]. The intervening 30 years of CSCL has produced a wide body of research that demonstrates a diversity of methodologies intended to identify and capture the complex set of variables that determine the success of any collaborative effort. CSCL has contributed to the formation of learning analytics [51, 62] and also benefited from methods and tools developed in learning analytics.

In addition to CSCL, several other fields including Computer Supported Cooperative Work (CSCW), Human-Computer Interaction (HCI), and Social Computing are also deeply invested in investigating collaboration, particularly as it relates to the ways in which information technology is used in the workplace. CSCW is a research community that emerged in the 1980s as “an effort by technologists to learn from economists, social psychologists, anthropologists, organizational theorists, educators, and anyone else who could shed light on group activity” [20, pp. 19–20] and focused on the twin goals of (a) examining how people work in groups, and (b) how computer systems and groupware can support collaborative activities [60]. Although one of the earliest papers on computer supported collaborative learning appeared at a CSCW conference [44] until very recently, the CSCW and CHI communities were primarily interested in studying how groupware was used by adults in the context of work rather than educational systems used by students in formal and informal learning contexts. However, with the ubiquitous nature of information technology in everyday life, both the CSCW and CHI conferences now include tracks for papers where the context is learning, education, and families.

As discussed above, CSCL and CSCW overlap considerably in both research interests and design methodology. This overlap was explored in three workshops (ACM Group 2010, ACM Group 2012 and CSCL 2013) and resulted in an edited book, CSCL@Work [18]. The two communities also share a strong interest in applying computational methods to understanding and coordinating
collaborative activities. Given emergent trends in modern societies, such as the blurred boundary between learning and work [30] and the rise of learning in the openly networked settings [23, 58], it becomes important to bridge perspectives from CSCL, CSCW, HCI, Social Media, and other fields where collaboration is explored.

To harness the power of learning analytics for scholarly research at the intersection of learning, work, and collaboration, several fundamental challenges need to be addressed. Specifically, the conceptualization of collaboration varies greatly across different communities, leading to a myriad of collaboration constructs researchers theorize and investigate. While multiplicity of ideas is championed within interdisciplinary fields like CSCL and CSCW, the scarcity of cross-community exchanges can lead to a disconnect between scholarly communities, efforts wasted on “reinventing the wheel,” and missed opportunities caused by different terminologies and epistemic cultures. This chapter about collaboration analytics aims to highlight these challenges for the learning analytics community.

We first survey the conceptual landscape of collaboration and learning while attending to perspectives from CSCW and HCI. Then, we distinguish between two salient strands of collaboration analytics: (a) computational analysis of collaboration that applies computational methods to examining collaborative processes; and (b) analytics for collaboration which is primarily concerned with designing and deploying data analytics in authentic contexts to facilitate collaboration. To articulate these two distinct strands, we introduce examples and cases that represent contexts of different scale, space, and analytical frames. Finally, we conclude by discussing challenges that lie ahead for collaboration analytics and point to future directions for research.

1 COLLABORATIVE LEARNING

Collaboration as a term is treated differently across scholarly communities. In the fields of CSCW and HCI, collaboration is used interchangeably with cooperation, broadly meaning cooperative work in a group [11, 20]. In contrast, the CSCL community has specific ideas about what can be considered collaboration. For many CSCL researchers, collaboration necessitates having a joint problem space [41], being intersubjective [47], and making deliberate efforts to coordinate group activities [10]. Despite these differences in defining collaboration, these communities overlap on the core constructs of collaboration. For example, much attention is given to group awareness in both CSCW [11] and CSCL [34]. The same parallels could be drawn about other collaboration constructs such as joint attention, shared understanding, transactivity, and intersubjectivity [3, 10, 49].

It is desirable to interrogate these constructs as new contexts for learning, such as Twitter and Microsoft Teams, continue to emerge.

The conceptualization of learning is also multifaceted. In CSCL, multiple traditions of learning co-exist, representing cognitive views of learning that foreground individual cognition, inter-subjective views that stress interactional sensing-making, and inter-objective views that locate learning with heterogeneous networks of learners, tools, artifacts, and practices [46, 26]. These frameworks guide research on learning in various contexts and also respond to emergent contexts in which learning happens. While much attention is given to learning in formal education spaces such as classrooms, new learning paradigms in informal education and at workplace challenge traditional conceptions of learning [7, 14, 32]. For instance, networked professional learning treats work as continual problem solving and learning as an integral part of such problem solving [5]. As the boundary between learning and work gets further blurred, cross-fertilization between research communities to enrich our understanding of learning is needed.

Building on the exploration of collaboration and learning, the following two sections discuss two salient strands of collaboration analytics: (a) computational analysis of collaboration that involves the application of computational methods to examining collaborative processes; and (b) analytics for collaboration which is primarily concerned with designing and deploying data analytics in various contexts of collaboration.

2 COMPUTATIONAL ANALYSIS OF COLLABORATION

Both CSCL and CSCW communities have been applying sophisticated computational methods to analyze collaborative processes, practices, and outcomes. The rise of data science has resulted in new computational methods to cope with large datasets, assist humans in laborious analysis of complex phenomena, and offer means to examine these phenomena from novel angels. While computational methods are sometimes touted as a silver bullet, Wise & Schwartz [63] remind us that “the substantive question is not if we should embrace computational approaches to understanding collaborative learning, but how to develop practices and norms around their use that maintain the community’s commitment to theory and situational context” (p. 441).

In the CSCL literature, methodologies from various disciplines including psychology, linguistics, and anthropology are adopted to examine collaboration learning [26]. Multiple data sources and mixed methods are often used to understand complex CSCL processes (e.g., [39]). Even with the same dataset, collaboration can be examined at different levels—e.g., individuals, small groups, the whole class, a massive online community—and at various units of analysis such as verbal utterances, gestures, discussion threads, and sessions of collaboration. Methodological richness and tensions have inspired research teams to explore the potential of “productive multivocality” by applying multiple analytical methods to shared datasets [50]. Growing awareness and access to computational methods are intensifying this exploration (e.g., [16]). Below we survey the specific ways in which computational methods can be applied to investigating collaborative learning (see Table 1 for an overview).
First, the cognitivist tradition focuses on the analysis of individuals. Within this tradition, while some may view collaboration as merely stimuli for internal cognitive processes (e.g., the Piaget’s [38] theory of cognitive conflict), others recognize the situated and embodied aspects of cognition (e.g., Hutchins’s [25] theory of distributed cognition, and Greeno’s [19] theory of situatedness). As a result, computational analysis could examine the impact of participating in collaborative activities on individual learning or the extent to which cognitive content is reflected in group exchanges. For example, a collaborative intelligent tutoring system, COMET, was developed to support medical problem-based learning in small groups. This system involved student groups to collaboratively form hypotheses of medical problems by examining shared medical images and chatting via text [48]. Students’ clinical reasoning was then modeled as Bayesian networks based on their hypothesis structure and their use of medical concepts in group chats. This analysis centered on students’ reasoning and cognitive content. In another study that involved group dialogues, Howley et al. [24] examined the cognitive constructs of reasoning and transactivity. The unit of analysis is the minimum amount of text in a dialogue that can adequately express reasoning. Transactivity is captured by first identifying reasoning in discourse and then re-instantiating new instances of reasoning that build on or evaluate existing ones. Computational linguistic techniques can be applied to measure semantic overlaps between contributions; machine learning models are built using linguistic features to automatically label the transactivity of discourse contributions.

Intersubjective frameworks are oriented more to the social and cultural levels of analysis. Computational analysis in this tradition emphasizes social and linguistic interactions in often messy group processes. In an example of collaborative problem-solving, student dyads collaborated remotely to understand human brains while they were able to review a set of diagrams and communicate with each other via audio [42]. Being interested in the construct of joint visual attention, researchers designed a condition where learners could see the eye gaze of their partner on the screen while solving the problem. Using natural language processing, the researchers found higher correlations between students’ learning gains and their verbal coherence in the condition with shared eye gaze. In another case of collaborative problem-solving by triads, Spikol et al. [45] attempted to build machine learning models to predict collaboration constructs including physical engagement and synchronization based on face and hand tracking data. In a similar example from a colocated, face-to-face context, Echeverria et al. [12] investigated teamwork from four intertwined aspects including physical, social, epistemic, and affective. Using multimodal data collected from location sensors, physiology wristbands, and microphones, they instrumented a data representation named the multimodal matrix and carried out matrix operations to derive proxies of teamwork related to awareness and accountability.

The inter-objective tradition requires more attention to the mediational objects and object-related activities in collaboration. Analyses abiding to this tradition could trace the trajectories of objects and unpack nuanced human activities around them. To analyze collaborative knowledge work on a wiki-based platform named Wikiversity, Halatchilyski and colleagues [22] adopted the main path analysis to examine the dynamic relations of knowledge artifacts and map the trajectories of ideas in different domains of the platform. In another example, an analytic tool named Knowledge Building Discourse Explorer (KBDeX) is designed to represent the evolving relations among key terms in collaborative discourse [37]. Rather than linking learners based on their social interactions, KBDeX connects learners based on the co-occurrence of key terms in their discourse contributions. The intricate, dynamic evolution of collaborative discourse is then represented by network representations that center on key terms, making it possible to assess constructs of collaboration such as collective responsibility using network indices [31].

To summarize, this strand of collaboration analytics is interested in applying a variety of computational approaches toward the study of collaboration. The application of these approaches is informed by theoretical frameworks and shaped by researchers’ epistemological stances. As demonstrated by these examples, computational methods have shown promise in making laborious analysis more efficient, creating new representations of data, and offering novel means to make sense of collaboration data.

3 ANALYTICS FOR COLLABORATION

Computational analysis also makes it possible to provide timely feedback for collaboration. In this section, we locate the central concern of analytics at the translation or transformation of findings from analysis to actions in the learning analytics cycle [43]. While the analysis of collaboration is dictated by epistemological and conceptual ideas, the use of analytics for collaboration deals with the distribution of agency between human and computer, as well as a wide range of other design decisions. Below we advance a typology of analytics built for collaboration based on how they are deployed in socio-technical systems of collaboration to make an impact. We choose to articulate these two important dimensions (see Table 2) as they are central to the human-computer partnership that have concerned CSCL and CSCW since their inception.

3.1 Analytics as Partner vs. Regulator of Collaboration

The first dimension is concerned with the power distribution between analytics and humans. Along this dimension, we distinguish analytics as a regulator versus a partner of collaborative interaction.

When analytics functions as a partner of collaboration, it acts to facilitate collaboration but still turns to humans for decision-making and action-taking. For instance, analytics applications are designed to support time coordination, a surprisingly challenging task for today’s organizations and teams. To confront this challenge, HCI and CSCW
Table 1: Applying computational methods to investigating collaboration.

<table>
<thead>
<tr>
<th>Traditions</th>
<th>Studies</th>
<th>Constructs</th>
<th>Data</th>
<th>Computational techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>[48]</td>
<td>Clinical reasoning</td>
<td>Chat logs, graph-based hypotheses</td>
<td>Bayesian network modeling</td>
</tr>
<tr>
<td>Cognitive</td>
<td>[24]</td>
<td>Reasoning and transactivity</td>
<td>Learner dialogues</td>
<td>Computational linguistic techniques</td>
</tr>
<tr>
<td>Intersubjective</td>
<td>[45]</td>
<td>Physical engagement, synchronization, and</td>
<td>Face and hand tracking data</td>
<td>Machine learning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>individual accountability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intersubjective</td>
<td>[42]</td>
<td>Joint visual attention</td>
<td>Eye tracking data</td>
<td>Natural language processing; eye gaze analysis</td>
</tr>
<tr>
<td>Intersubjective</td>
<td>[12]</td>
<td>The physical, social, epistemic, and affective</td>
<td>Temporal interaction data; multimodal data</td>
<td>Multimodal matrix; Quantitative ethnography</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dimensions of group activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interobjective</td>
<td>[22]</td>
<td>Trajectories of ideas</td>
<td>Log data of wiki edits</td>
<td>Main path analysis</td>
</tr>
<tr>
<td>Interobjective</td>
<td>[31]</td>
<td>Collective responsibility in knowledge building</td>
<td>Learner dialogues</td>
<td>Socio-semantic network analysis [37]</td>
</tr>
</tbody>
</table>

Table 2: Two dimensions of collaboration analytics.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>As Partner</th>
<th>As Regulator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loosely Coupled</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Idea Thread Mapper [65]</td>
<td>• Sociometric badges and feedback [27]</td>
</tr>
<tr>
<td></td>
<td>• Reactive conversational assistants [61]</td>
<td>• CSCL teacher dashboard [28]</td>
</tr>
<tr>
<td><strong>Tightly Coupled</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• A.I. scheduling assistant</td>
<td>• Software agents in scripted collaborative inquiry [54]</td>
</tr>
<tr>
<td></td>
<td>• Group formation in MOOCs and WikiProjects [59, 67]</td>
<td>• Conversational agents for collaborative problem-solving [52]</td>
</tr>
<tr>
<td></td>
<td>• Proactive conversational assistants [61]</td>
<td>• Wikipedia ClueBot NG [66]</td>
</tr>
</tbody>
</table>

researchers have created A.I. scheduling assistants that act just like human agents to schedule meetings [9, 36]. Based on a combination of heuristics, machine learning, and natural language processing, such A.I. assistants are trained to extract meeting information, such as meeting subject, time, and attendees, from emails and engage in back-and-forth messages to coordinate meetings [9]. In this case, the A.I. agent serves as a partner delegated to solve the mundane and yet non-trivial task of time coordination.

Analytics can be a partner for team formation in large-scale collaboration settings. NovoEd is a social learning environment that supports team formation processes in massive online classes. Teams can be formed algorithmically based on instructor-specified factors such as size of the team and geographical location of the members [40]. Another analytics-based team formation approach draws on discussion data and algorithmically assigns learners to teams based on their transactive interaction with each other [59]. On Wikipedia, a variety of algorithms are designed to recommend newcomers into WikiProjects based on their interests in or relationships with project topics; human agents including project leaders remain “in the loop” to carry out the action of inviting newcomers [67].

Besides temporal coordination and team formation, analytics can also be a partner that provides content-specific support relevant to the task. For instance, Winkler et al. [61] developed a smart personal assistant using Alexa to facilitate collaborative problem-solving by providing proactive structured facilitation and reactive help for humans’ content-specific questions. When analytics act as a partner in such cases, they provide important affordances that contribute to key constructs of collaboration but do not evaluate collaboration or prescribe actions on the human’s behalf.

When analytics acts as a regulator, in contrast, it takes responsibility in monitoring the status of collaboration and taking actions to shape the ongoing progress of collaboration. One example is the awareness lantern designed by Alavi & Dillenbourg [1]. Combining colors, lightness, and blinking, the lantern creates an ambient display of the status of collaborative groups designed to attract the tutor’s attention. Student teams can press the lantern to call...
for help and the lantern blinks and adjusts the blinking frequency based on the wait time. In this case, the lantern directly mirrors the status of collaborative groups and regulates the help-seeking process in a classroom. In another example, sociometric badges are used to collect and analyze data from geographically distributed teams and provide instant feedback about team participation [27]. Based on interaction patterns captured by sociometric badges, feedback is provided each team to promote active and balanced participation and frequent turn transitions [27]. In classrooms where multiple collaborative teams are in action, teacher dashboards are designed to capture multiple group indicators (e.g., task progress, participation balance) and alert the teacher when a group deviates from a norm [57]. In these cases, analytics provide evaluative information about collaboration to different analytics “consumers” (the teacher, participants, software) for them to take regulatory actions towards collaboration.

3.2 Action-taking Being Closely vs. Loosely Coupled with Collaboration

The second dimension is about the ways in which analytics are integrated with collaboration processes. Here we distinguish analytics that are closely vs. loosely coupled with collaborative actions. This distinction is concerned with the relation between analytics-based action-taking and the other components of a collaboration workflow.

On one side of the continuum, analytic outputs present merely outcomes of computational analysis of collaboration and it is up to humans to choose whether, when, and how to act upon the presented information. On Wikipedia, quality management in the editorial process increasingly relies on algorithmic agents or “bots” [17]. For instance, the SuggestBot applies a combination of text analysis, collaborative filtering, and hyperlink following to suggest editing tasks to Wikipedia editors based on their edit histories; suggestions are made directly to an editor who would decide how to react [8]. In this case, analytics is loosely coupled with any individual or collaborative editing efforts. In contrast, the ClueBot NG is designed to automatically detect vandalism based on a machine learning approach and autonomously revert vandalism as soon as it is discovered [66]. While both bots act as partners (see Dimension 1), they differ in how closely their analytic actions are coupled with the overall editing process on Wikipedia.

In knowledge building classrooms, teachers and students have had access to analytics tools embedded in the Knowledge Forum since the ’90s [4, 53]. Much like teacher dashboards in CSCL classrooms (van Leeuwen, Wise & Teasley, this volume), these analytics, such as social network and lexical analysis tools, are loosely coupled with the knowledge-building workflow. A more recently developed “meta-discourse” tool known as the Idea Thread Mapper shares the same characteristic [65]. With assistance from topic modeling techniques, this tool helps learners identify “idea threads” in their Knowledge Forum dialogues and then reflect on their collective progress [65]. Similar to the Wikipedia SuggestBot, the Idea Thread Mapper is also loosely coupled with students’ knowledge work and it is up to the humans to trigger its use during knowledge building.

On the other side of the continuum, analytic actions are deeply embedded in collaboration processes. Analytic tools embody ideas about how actions should be taken in response to a collaborative situation. In scripted collaboration, software agents can be specially designed to process student interactions in real-time in response to both pre-specified scripts and emergent collaborative scenarios. For example, in a “smart learning space” designed to facilitate sophisticated collaborative inquiry, high-school students work together as a community to address science problems [54]. Tablets, large displays, multi-touch tables, and the teacher play distinct roles in supporting the inquiry. In particular, multiple real-time software agents are present to sort students into groups, monitor whether groups have achieved consensus, and track individual, group, and class-wide progress. Drawing from various computational techniques, these software agents automate important parts of the collaboration scripts and help the teacher make orchestral decisions [54]. The roles played by these software agents are akin to the operators in an orchestration graph [6, 21]. Analytic actions (such as distributing student-generated post-it notes based on groups and topics) are embodied by these operators, setting the condition for the next collaboration activity (such as making sense of the assigned post-it notes as a group). Here, analytics are tightly coupled with predefined collaboration scenarios or workflows.

Conversational agents developed to facilitate peer collaboration can also embody analytic supports tightly within the flow of collaborative conversations. For example, MentorChat asks learners to collaborate on open-ended learning tasks through online chats. Drawing on the Accountable Talk framework that details productive classroom discussion practices and norms [33], MentorChat processes each dialogue contribution, updates students’ domain models, decides whether an intervention is desirable, and if so, delivers its intervention verbally using a text-to-speech engine [52]. Analytics, including semantic analysis based on WordNet, directly responds to the unfolding student dialogue; the agent directly intervenes and hereby triggers further student conversations [52]. In contrast with the Alexa-based conversational agent that acts as a partner who answers student questions [61], MentorChat serves a regulatory role by monitoring students’ domain understanding and directly intervening when necessary.

In summary, we have identified two important dimensions of analytics for supporting collaboration: analytics as regulator vs. partner, and analytic actions being tightly vs. loosely coupled with collaborative interaction. This typology can provide a roadmap for future development of collaboration analytics. It is important to note that these two dimensions function as continuums and, as illustrated in these aforementioned cases, one analytics application could serve multiple roles that cut across multiple areas of the space.
4 CONCLUSIONS AND FUTURE DIRECTIONS

Collaboration is widely considered to be an important competency in modern society. As educators and researchers, we actively theorize what collaborative learning means, debate where collaboration sits in the curriculum, and develop interventions to facilitate collaboration at all levels of education. Given the importance of collaboration, coupled with the emerging quest for human-computer or human-A.I. partnerships, analytics and computation are destined to play an essential role in future efforts to facilitate collaboration in all domains of human activity.

Because analytics can be used to both examine collaborative processes and support the design of systems to facilitate collaboration, analytics can be leveraged to make progress on two essential questions: How do successful collaborations work? How can we design supports to promote collaboration? Learning analytics has the potential to inform the research on collaboration by contributing to good learning design, effective pedagogy and increasing learner self-awareness [13]. To do so, we see several important challenges and future directions in the area of collaboration analytics. First, more efforts need to be invested in bridging research communities that have been actively investigating collaboration from distinct but overlapping theoretical viewpoints. A number of projects are ongoing to bridge perspectives from CSCL, CSCW, HCI, Social Computing, and Learning Analytics (e.g., [12]). Such work would alleviate the scarcity of theory underlying learning analytics since its earliest days [15, 64]. At the same time, learning analytics has the opportunity to contribute to our theoretical understanding of successful collaboration by creatively integrating sources of data (such as demographic information, physiological data, and behavioral data) and modeling collaboration processes [2].

Second, as the world is increasingly connected, it is important to consider the factor of scale and ways to harness scale in collaboration. In CSCL, scale is considered from both group size and time but heavily focused on small group collaboration within a limited timeframe [10], typically in single classrooms, after-school clubs, and museums. By contrast, CSCW and social computing researchers have a more expansive coverage given their stronger interests in open online communities such as Wikipedia [67] and software development projects [35]. Compared to small-scale collaboration scenarios in highly controlled educational contexts (e.g., collaboration scripting software, intelligent tutoring systems), the mechanisms or interactive processes to support collaboration may be different in open, large-scale environments where the participants have very different motivations to collaborate than do students. The ubiquity of the Internet has not only created new opportunities for geographically unbounded interactions, the rise of “Web 2.0” technologies have also blurred the lines between school, home, and the workplace. Following Bransford’s notion of “lifelong and life-wide” learning [29], we need to utilize learning analytics to conceptualize collaborative learning whenever and wherever it occurs. This remains a challenge for the field of learning analytics where the research has to-date been conducted primarily in formal educational settings, particularly higher education and professional training.

Third, the distribution of agency between humans and analytics is a critical and contentious issue that needs to be carefully navigated when designing and deploying collaboration analytics. In Wikipedia, the delicate relations between human editors and bots, as well as among bots, are especially illuminating [17, 56]. The learning analytics community needs robust design approaches to help us cope with value tensions and ethical dilemmas in a learning analytics system [55, 67]. As human activities are shaped by various analytics tools, we need to critically examine the structures (temporal, spatial, social, material, conceptual) created for collaboration, and the ways in which human and computer agents are collectively shaping these structures.

REFERENCES


