Analyzing Social Media and Learning Through Content and Social Network Analysis: A Faceted Methodological Approach

Anatoliy Gruzd
Ryerson University, Toronto, Canada
gruzd@ryerson.ca

Drew Paulin
University of California, Berkeley, USA

Caroline Haythornthwaite
University of British Columbia, Vancouver, Canada

ABSTRACT: In just a short period, social media have altered many aspects of our daily lives, from how we form and maintain social relationships to how we discover, access, and share information online. Now social media are also affecting how we teach and learn. In this paper, we discuss methods that can help researchers and educators evaluate and understand the observed and potential use of social media for teaching and learning through content and network analyses of social media texts and networks. This paper is based on a workshop given at the 2014 Learning Analytics and Knowledge conference, and presents an overview of the measures and potential of a multi-method approach for studying learning via social media. The theoretical discussion is augmented with study of the case of Twitter discussion from a cMOOC class.

Keywords: Social media, content analysis, social network analysis, networked learning

1   INTRODUCTION

Social media use has dramatically increased over the past few years. Currently, over 302 million people use Twitter each month, and over 500 million tweets are sent every day (Twitter, 2015); Facebook has over 1.44 billion active users per month (Statista, 2015); and every minute, 300 hours of video are uploaded to YouTube, with YouTube videos generating billions of views daily (YouTube, 2015). These media, along with other Internet technologies, have greatly influenced learning environments and the roles and behaviours that both learners and educators enact in creating and sharing learning experiences. Social media are at the forefront of this transformative shift, bridging the social relationships and communities in which learners participate with the discovery, sharing, filtering, and co-constructing knowledge and information that is a principal aspect of the online world.

Whether learners are in a formal learning context or are informally seeking new learning experiences under their own direction, they are turning to various social media platforms for information from individuals and communities with a shared learning interest. However, with the turn to social media, both instructors and learners are yet again challenged to develop new learning practices for constructing these collective learning spaces. Questions of concern for researchers, instructors, and learners then arise: What work do these media do in support of learning? How can we identify and evaluate learning processes through social media? What conditions, structures, exchanges, pedagogies, and practices foster and enable learning through social media?

Research and practice in this area is supported by the rich digital trails left behind as social media are used to form and maintain social relationships, and to discover, access, and share information online. These trails describe the social learning networks of who is interacting with whom, what they are talking about, and how information and resources flow and circulate in a network. From the comments, contributions, images, and videos posted by individuals, to the network structures formed through relationships, connections, ties, information flows, and exchanges, the resulting dataset can be leveraged to address questions about networked learning and the benefits that accrue for participating in these learning networks. The challenge is how best to analyze and make sense of these data to understand and support online learning.

Our experience in working with social media datasets leads us to advocate a semi-automated, multi-method approach for evaluating and understanding the observed and potential use of social media for teaching and learning. This approach relies on both content analysis and social network analysis, and allows for the exploration of multiple levels and facets of social media use for learning.

Incorporating multiple methods is key to our approach because these bring to light different facets of the phenomenon of learning through social media, leverage the strengths of two different methods of analysis, and offer a number of combinatory tactics towards exploration and understanding. In our work, we are interested in both *who* is talking with whom, and *what* they are talking about. This emphasizes our interest in both the network of social connections, and the nature of the tie that underpins these connections. Social network analysis provides the means to address questions about the structure of the social network, while content analysis allows us to focus on the nature of the tie (Gruzd & Haythornthwaite, 2008). A faceted research methodology enables exploration of how these two inquiries intersect and complement each other. For example, this approach allows a focus on network characteristics and outcomes (e.g., resource flows, roles and positions, relationships, and social structures) *in relation to* the emergence of shared language, community styles and norms, attention to specific topics, patterns of affective language, and so on. Further, in the case of formal learning contexts, one can look at any of these facets of social media use in relation to employed pedagogies, strategies, and teaching practices, in order to evaluate and inform learning design.

We also believe that the nascent nature of learning through social media engenders an exploratory rather than confirmatory approach. The framework we propose, along with the body of research that

we discuss in this paper, is not geared towards studying or measuring specific learning outcomes. Instead, we are studying behaviours and conditions that the literature associates with learning. Thus, we focus our attention on aspects of information exchange and social interaction that we believe are correlated to learning, are important contributing factors to learning processes that might be taking place in social media, and/or are indicators of environments that enable and foster the development and maintenance of learning communities and networks.

Our overarching goal is to develop and evaluate a framework of methods and strategies for learning analytics that can be used to detect and study learning processes happening on social media platforms in both formal and informal settings. The intent of this paper is to generate discussion around this broad framework, and revisit the existing tools and methods that support this kind of faceted multi-method approach to researching teaching and learning through social media.

This paper starts with a review of relevant literature that informs the landscape of social media and learning, provides the theoretical underpinning of significant methods and approaches that help form our proposed analytic framework, and integrates concepts from other knowledge domains into a learning analytics perspective. Next, this paper provides a case study that further explains and demonstrates our analytic framework. We apply our framework to a dataset collected from a Connectivist Massive Online Open Course (cMOOC), illustrate several analysis methods we rely on, and show how they can be used in a combinatory, complementary fashion to generate new insights. We then discuss our framework in relation to a number of contexts: from how it might be employed in formal educational to evaluate and optimize learning design, to how it can help detect and understand learner behaviours in informal, self-regulated learning contexts. The paper concludes with a reflection of our work in relation to the ongoing development in learning analytics research and tool development, a discussion of limitations and potential issues surrounding our framework, and a look ahead to directions for future work.

2   THE LANDSCAPE OF RESEARCH ON SOCIAL MEDIA AND LEARNING

2.1   Formal and Informal Learning Contexts

Higher education faculty recognize the value that social media can leverage in their curriculum, with over one-third of teaching faculty in the US using some form of social media in their courses, and adoption rates of social media as high as 80% in university classrooms in the US (Moran, Seaman, & Tinti-Kane, 2012). A recent EDUCAUSE study (Dahlstrom, Walker, & Dziuban, 2013; Smith & Caruso, 2010) indicates that social media are being formally integrated into institutional academic learning experiences, and being informally used by students to supplement their learning experiences. This allows students to reach wider social networks via social media while simultaneously "meeting the student population where it lives: i.e., online, in social networking sites and in the microforms of communication adopted in Twitter" and other popular online platforms (Gruzd, Haythornthwaite, Paulin, Absar, & Huggett, 2014, p. 254).

Learners use various forms of social media to bridge the gap between in-school and out-of-school learning by enabling the discovery of connections between their traditional curricula, their personal interests, and online communities that can support and further their engagement and learning (Ito et al., 2013). Traditional learning contexts and online platforms such as learning management systems (LMSs) do not often expose students to the learning opportunities afforded by social media in terms of enabling connections to peers, communities, and resources across time and space (Dabbagh & Kitsantas, 2012). To this end, learners use social media to expand their learning opportunities beyond the classroom and the LMS in a self-directed manner, enabling the personalization of their learning experiences to their own interests, their own learning goals, and their own preferences in terms of participation, online communities, and social media platforms (Mcloughlin & Lee, 2010; Siemens, 2008).

As learners progress through school and towards professional life, formal learning plays an increasingly smaller role in lifelong learning experiences while informal learning becomes integral to developing knowledge and skills (Banks et al., 2007; Chen & Bryer, 2012). Informal learning opportunities are afforded through connections and interactions with networks of peers, and with the ideas and resources made available through those networks. In this way, informal learning supports involvement in a knowledge-creating culture: developing knowledge-building competencies, understanding one's own learning in relation to, and in contribution to, a larger knowledge-building community (Scardamalia & Bereiter, 2006), shaping the (online) community of practice (Lave & Wenger, 1991; Wenger, 1998; Haythornthwaite & Andrews, 2011). Social media enable learners to pursue this kind of social, group-based learning by providing the means to create, find, organize, and share resources, and participate in networks and communities with a shared learning focus or interest (e.g., see Gruzd & Haythornthwaite, 2013). Thus, social media amplify and expand the informal learning opportunities available to learners.

Ziegler, Paulus, and Woodside (2014) note that research on informal learning has largely relied on retrospective accounts of learning from the learners themselves, through interview or survey data. However, asking people what they have learned, and how they have learned it, can be problematic as respondents often lack awareness of their own learning, and regard it as part of their own general capability rather than something learned (Eraut, 2010). While self-reported data provides an account of the lived experiences of individuals, dialogue and textual language that occurs during social activity provide an account of the social reality constructed by those engaged in conversation. Rorty (1992) argues that language creates, rather than represents, lived experiences. The language that comprises the exchanges and interactions on social media is a valuable source of data that can be analyzed to understand how informal learning occurs.

2.2   Text-based Content Analysis

Social media creates a vast quantity of textual data that record the history of group interaction as networks and communities form, grow, and decline. Content analysis is a method for examining patterns of text and language. Content analysis relies on systematic techniques that compress large amounts of text into fewer coded categories, enabling researchers to discover and explore the focus of

attention in text or dialogue (Krippendorff, 1980; 2012). Within educational and learning contexts, content analysis has been used to investigate asynchronous discussion to identify markers of collaboration and co-operation (De Wever, Schellens, Valcke, & Van Keer, 2006); to detect cognitive presence in online discussions (Kovanović et al., 2016); to conduct sentiment analysis to understand the relationship between sentiment expressed in discussion forums and attrition rates in a MOOC (Wen, Yang, & Rosé, 2014).

Analyses can give insight into the characteristics, interests, and priorities of a learning network, and reveal patterns of language and interaction that characterize a community and foster learning (Haythornthwaite & Gruzd, 2007). Analysis of online discussions can uncover underlying mechanisms of group interaction, and identify unique language patterns that demonstrate instances of thinking, collaboration, or learning (Strijbos, Martens, Prins, & Jochems, 2006). Further, text analysis can provide insight into concepts central to discussion or to generating interest within a learning community, the nature of exchanges occurring (i.e., informational, socially oriented, and so on), or the semantic or affective weight of language used in discussions. In determining which social processes and concepts should be examined through content analysis, researchers are led by theories and perspectives that guide understanding of learning (see De Wever et al., 2006, for a review of common concepts and processes studied, along with corresponding theories; see also Rogers, Dawson, & Gašević, 2016; Eynon, Schroeder, & Fry, 2016; Wise & Shaffer, 2015).

While there are many perspectives on what social processes and concepts are most appropriate for studying learning, most content analysis work relies on the development of *categories* that define the processes and concepts under investigation, *coding* them to identify and interpret text that falls under one or more categories (see Krippendorff, 1980; 2012). Research choices include the definition of categories and the selection of units of analyses — words, symbols, or phrases — within the text that represent or indicate a category. For example, if a category was defined by emotive expression, words such as "love" or "hate" are likely to be useful units that identify discussion contributions that can be categorized as emotive expression.

Content analysis often relies on manually finding, labelling, and interpreting categories in text. While this is manageable for smaller corpora, manual content analysis is not practical for larger datasets such as those found in social media or MOOCs. While teams could be formed to distribute the burden of manual coding and analysis, the resulting lack of consistency and agreement on interpretation of categories introduces the problem of consistency and reliability. Automated text analysis offers an alternative for such datasets.

2.2.1 Automated Text Analysis

The field of computational linguistics has developed many Natural Language Processing (NLP) algorithms and techniques to automate the analysis and representation of text. Many of these techniques provide analysis in the form of finding meaningful patterns in text through word counting, key phrase matching, or visualization of patterns of categories (Rosé et al., 2008). Tools such as Linguistic Inquiry and Word

Count (LIWC) rely on dictionary-based methods to identify organizations of words and phrases that indicate specific mental states or emotions (Pennebaker, 2003). NLP relies on lexical analysis to identify word classes (i.e., nouns, verbs, etc.) and syntactic analysis to reveal grammatical structures in text (Liddy, 1998; Rubin, Stanton, & Liddy, 2004). This allows nouns and noun phrases — considered to be the most informative elements of text (Boguraev & Kennedy, 1999; Carley & Palmquist, 1992; Carley, 1997; Corman, Kuhn, McPhee, & Dooley, 2002) — to be identified, and visualized in topic maps or world clouds (Haythornthwaite & Gruzd, 2007).

Using machine learning approaches towards NLP, semantic analysis allows for automatic analysis of text beyond dictionary-based categorization and frequency counts of words. Through a process of training a program on massive textual data sets and focusing on frequency, proximity, and many other linguistic factors, a program can learn and assign context to language. This goes beyond understanding meanings and categorizations of words towards understanding relationships between words, phrases, and ideas akin to human-like, common-sense knowledge about the world through language. Semantic analysis enables complex tasks such as word-sense disambiguation for words with multiple meanings, building systems capable of answering questions posed in plain language, or translating across languages. Table 1 presents a list of examples of currently available content analysis tools and their key features.

Table 1. Examples of content analysis tools and key features

Tool name	Key features
Netlytic	A cloud-based text and social network analysis tool that allows users to capture and import online conversational data, and find, explore, and visualize emerging themes of discussions.
LIWC	A dictionary-based text analysis program that categorizes words that reflect different emotions, cognitive styles, social and psychological states.
Atlas.ti	Software that aids qualitative analysis of unstructured data (text, multimedia, etc.) through coding, annotation, and visual structuring.
NVivo	A qualitative data analysis software package that allows users to classify, sort, and arrange unstructured data, and examine relationships within data.
LightSIDE	A text mining tool bench that leverages machine learning to enable automated analysis of conversational interactions and social aspects of text (e.g., perspective modelling, sentiment analysis, opinion mining).
RapidMiner	An analytics software platform that offers text analysis and sentiment analysis tools.
Weka	A collection of machine learning algorithms for data mining tasks, including semantic analysis and sentiment analysis.

2.3   Social Network Analysis

The emergence and growth of social media — networked tools, platforms, and their associated practices — has inspired rethinking of how we might learn in today's highly connected environment (Siemens, 2005). This line of thinking has led to the conceptualization of a personalized learning network — a collection of interoperating applications that form an ecology of social media and networks through

which individuals explore and learn (Fiedler & Väljataga, 2011). An ecosystem approach leads to a particular network-based pedagogy where learning is supported through practice, reflection, and participation in communities, and engaging in a distributed environment consisting of networks of people, services, and resources that provide learning opportunities (Downes, 2006).

While learning networks provide opportunities for the learner, the distributed, interconnected nature of the model provides challenges for educators, learning designers, and researchers interested in understanding how people learn and the effectiveness of their learning experience. Social Network Analysis (SNA) provides knowledge, perspectives, and tools that can be applied to the interpretation and design of networked learning (Haythornthwaite & de Laat, 2010; Haythornthwaite, de Laat, & Schreurs, 2016). SNA can help in understanding how and why learners in a network are connected, how they seek each other out, and how their connections, configurations, and interaction patterns support information and knowledge sharing. Thus, a network perspective can provide a number of novel ways that learning can be represented and addressed, guide efforts in evaluation, and aid in designing learning experiences and technologies that foster and support networked learning (Haythornthwaite, 2008, 2011; Daly, 2010).

SNA has been used in learning research to depict teacher and learner communication patterns from LMS data (Dawson, Bakharia, & Heathcote, 2010), to identify collaborative work patterns across different media and channels among online learners (Haythornthwaite, 1999), to identify learners who are absent or peripheral to a course's learning network in order to identify disengaged and at-risk students (Macfadyen & Dawson, 2010), and to explore how students from different cultures interact, develop friendships, and forge learning relationships within an interactional classroom (Rientes, Héliot, & Jindal-Snape, 2013; see also Haythornthwaite, de Laat, & Schreurs, 2016).

The network approach focuses on how patterns of interaction afford an environment for exchange of resources (Wasserman & Faust, 1994). This perspective views learning as social relations in a network: transactions, exchanges, and shared experiences that emerge from interaction between individuals, and engagement across a larger group that forms a community of learning. The characteristics of community learning exemplify the principles of SNA derived from graph theory, which looks at patterns of relational connections between nodes in a graph: *Actors* are seen as nodes in the network connected by *relations* that form interpersonal *ties.*

In formal educational settings, actors can be teachers, students, or administrators. In informal learning settings, actors may be interested learners, students, experts, organizations, institutions, researchers, practitioners, co-workers, or collaborators. Learning can occur through interaction with other people, through participation in events, or through experiences. Thus, learning networks may be multi-modal; actors in learning networks may be people, sources, or activities (Haythornthwaite & de Laat, 2010). The relations through which these actors interact and connect — exchanges of information, provisions of support and resources, collaborations and communication — define the kind of relationship between actors, from close personal friendships to professional acquaintances, to people who do not know each

other beyond interacting within the same network of actors (Gruzd & Haythornthwaite, 2013; Haythornthwaite, 2008).

In closer relationships, more types of exchanges between people occur and more importance is placed on these exchanges as they often demonstrate a higher level of self-disclosure and intimacy (Granovetter, 1973). Such ties are referred to as *strong ties*, where paired actors engage in high levels of resource sharing, are often similar to each other, and tend to know and interact with similar sets of actors within a network. Trust and familiarity between close tie relationships foster environments in which learners feel comfortable asking questions and exchanging feedback. However, due to homophily in information sources and perspectives, reliance on only strong tie relationships can result in a filter bubble where new information and differing opinions are suppressed. In contrast, *weak ties* exhibit fewer exchanges, fewer different types of exchanges, and are less motivated to share resources. However, the "strength of weak ties" (Granovetter, 1973) is that they are dissimilar in terms of habits, circles of friends, etc., and thus offer greater access to different resources circulating in other domains. A learning network that provides a variety of ties across varying degrees of strength and closeness is optimal in that it provides a wealth of knowledge sources and perspectives, and a variety of interaction opportunities in which learners may engage.

SNA depicts conditions that support learning in several ways. SNA can reveal how information flows through ties in a network, and how a network's structure and configuration allows knowledge to be disseminated and created across actors (Haythornthwaite, 2011). The configuration of a network may affect learning by indicating which actors have access to information and resources, and which actors lack access. In high-density networks with many links between nodes, high degrees of sharing and access to information are more probable. Sparse networks often exhibit structural holes between clusters of highly connected nodes, where specific actors may serve as information brokers, required to bridge such gaps so that information can be shared between groups (Burt, 2004).

By viewing a network from the perspective of an individual learner, one can understand what information sources that learner has been exposed to and with whom they may be learning, along with where conflicts in their understanding may come from (i.e., opposing viewpoints or contradictory information), and may also reveal conflicting or complementary demands on individuals, particularly for adults at work (Haythornthwaite & de Laat, 2010). Viewing a network as a whole allows one to see how learning may be occurring across an entire set of people, and provides a view on the norms and character of the larger network to which individuals belong. For example, is the network collaborative, highly active, helpful, and inclusive? Is the network clustered into cliques? How do clusters tend to form? A whole network perspective allows one to understand the social conditions and relations that underpin learning behaviours within that network, and what holds the network together (Haythornthwaite & de Laat, 2010). Table 2 presents some examples of current social network analysis tools and their key features.

Table 2. Examples of social network analysis tools and key features

Tool name	Key features
Netlytic	A cloud-based text and social network analysis tool that allows users to capture and import online conversational data, and build and visualize communication networks. Netlytic can automatically build chain networks and personal name networks, based on who replies to whom and who mentioned whom. Netlytic also allows for comparison of networks across a number of centrality and other network measures.
Gephi	A network analysis and visualization package that allows for interaction and exploratory analysis of graph data that offers a number of different layouts based on force-based algorithms, and offers common SNA metrics. Gephi also allows for visualization over time so that one can see how a network evolves across a timeline.
UCINet and NetDraw	A comprehensive social network analysis and visualization tool. Allows users to include and add attribute data alongside relational data typically used in SNA. Supports matrix analysis routines and multivariate statistics.
NodeXL	A Microsoft Excel add-in and C#/.Net library for network analysis and visualization. Adds "directed graph" as a chart type to Excel spreadsheets, and offers a number of network metrics and visualization options.
R (igraph, sna, and network packages)	R contains several packages that can be used for social network analysis, including igraph, sna, and network. These represent a sample of a larger collection of network analysis and visualization packages available in R. Using R for social network analysis allows one to complement SNA work with other statistical analysis within the R environment.

3   CASE STUDY

3.1   Dataset

To provide further explanation and demonstration of our analytic strategy and framework, this section focuses on several analysis methods we rely upon and how they are used in combination to generate new insights about learning. For this case study, we use a sample of public tweets posted by participants in a 2011 cMOOC led by Stephen Downes and George Siemens, called *Connectivism and Connective Knowledge* 2011 (CCK11, http://cck11.mooc.ca/).

CCK11 ran for 12 weeks, from January to April 2011, and addressed the topic of connectivist perspectives on networked, distributed learning and construction of knowledge. Discussions and learning processes in this course were supported through the following four tasks:

1) Aggregate: Participants were given access to a wide variety of resources to read, watch, or play with.

2) Remix: Participants were encouraged to keep track of and reflect on their in-class activities using blogs or other types of online posts.

3) Repurpose: Participants were asked not just to repeat what other people have said, but also to create their own content.

4) Feed Forward: Participants were encouraged to share their work with others in the course or outside the course to spread the networked knowledge.

Course resources were distributed through a central course site, along with online seminars delivered using Elluminate. The course, however, was not restricted to a single platform or environment. Participants were free to use a variety of technologies for sharing and participating in the course, and hence the content was distributed across the web. To keep track of their learning and sharing content, participants were encouraged to create blogs using any blogging service (e.g., blogger.com or wordpress.com), use del.icio.us, discuss on Google groups forums, tweet about items on Twitter, or use any other platform such as Flickr, Second Life, Yahoo Groups, Facebook, or YouTube.

To keep track of their content, participants were asked to use the #cck11 tag in whatever content they created and shared. This tag was used by aggregators to recognize content related to the courses. The aggregated content was then displayed in an online "newsletter" created every day to highlight new content posted by learners.

To collect data for our study, we scraped the archives of the daily newsletters for each course and used automated extraction for Twitter messages, discussion threads, blog posts, and comments on blogs. The platform that generated the greatest number of posts was Twitter, followed by blogs. The sample used in the case study presented here is limited to tweets using the course hashtag #CCK11, posted between January 21 and March 10, 2011. This dataset consists of 1,617 Tweets, from 467 unique Twitter users. The methods detailed in this section are available in the cloud-based text and social network analysis tool suite called Netlytic. Along with a description of text, network analysis, and visualization techniques, this section offers potential insights and explorations facilitated by such analyses.

3.2 Text Analysis

3.2.1 Most frequently used words

The first step in our case was to build concise summaries of the communal textual discourse present in the dataset by identifying frequently used words (mostly nouns). Figure 1 shows a word cloud visualization of the top 50 most frequently used words in the #CCK11 Twitter chat over the data collection period. The search keyword (#CCK11) and other common words (also known as "stop-words") such as "of," "will," and "to" were automatically removed prior to building this visualization. The size of a word in the visualization is directly related to the number of times it appears in the dataset relative to the other words found in that same dataset. In Netlytic, this visualization allows users to click on any of the words in the cloud in order to explore the context(s) in which the word appears.
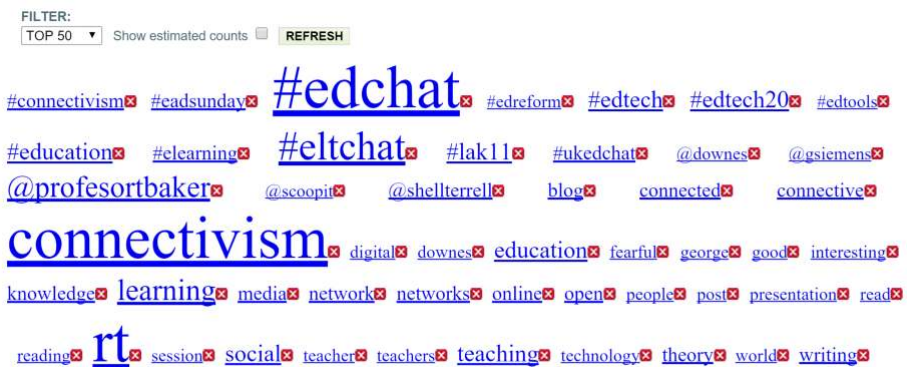
Figure 1. Top 50 most frequently used words in #CCK11 Twitter chat.

By exploring the top 50 words, we can group words into four broad categories. The first category includes words relevant to the class but not necessarily unexpected, including "learning," "education," "social," "teaching," and "knowledge." The most frequently mentioned word in this category (and in the whole dataset) is "connectivism" referring to the new learning theory at the core of this class (Siemens, 2005; Ravenscroft, 2011). While one would expect to see these words in this category, their presence is a helpful check confirming that class discussions were indeed focusing on the topics related to the class objectives. Such an observation would be useful for any instructor.

The second group of frequently used words includes Twitter hashtags: #edchat, #eltchat, and #edtech. The first hashtag, #edchat, was used to organize a Twitter community and weekly chats by educators wishing to discuss current trends in educational technologies and policies (http://edchat.pbworks.com/). The second hashtag, #eltchat, is described as a social network for English Language Teaching (ELT) professionals (primarily English language teachers), which is also used to facilitate weekly chat and continuous education (http://eltchat.org/). The third hashtag, #edtech, is frequently used in conjunction with #edchat by educators, technology bloggers, developers, and organizations interested in sharing some of the latest news and technology trends in academia. Other hashtags such as #edtech20 and #lak11 were used to connect class participants to relevant conferences on online education and teaching technologies. All these hashtags are highly relevant to the CCK11 class, considering its focus on "understanding of educational systems of the future." The prevalence of hashtags other than the one for the class #CCK11 suggests that class participants were actively connecting to other relevant communities and information on Twitter, discovering and sharing relevant resources outside the class. This exemplifies Twitter's ability to connect to other relevant people and communities, and facilitate the formation of weak ties across different communities, thereby introducing members of those communities to potentially new and diverse sources of information.

The third category includes a set of Twitter users frequently mentioned in the dataset[1] such as @profesortbaker, @downes, and @gsiemens. These were active participants and facilitators in the course. Active Twitter users will be discussed in the following section as part of the network analysis of the communication network.

The fourth category of frequent words reveals what types of online content were found to be useful and shared within the class. For example, the presence of words like "presentation," "post," "live," and "video" in the word cloud suggests that Twitter is in part being used to disseminate online presentations by instructors, students, and experts.

In addition to the four broad categories found in the dataset, we also observed the frequent use of the symbol "RT," added manually or automatically to tweets when they are "retweeted" by others. The use of RTs may indicate the extent to which class participants paid attention to what others post; the prominence here suggests frequent attention to classmates' posts with retweeting content to their own followers fulfilling the "Feed Forward" action. It is important to note that there is no suggested "optimal" ratio of retweets or replies to original posts that one might want to see in successful class discussions on Twitter. It would largely depend on the primary reasons why the social media platform, in this case Twitter, is being used in the class, and to the pedagogical approach intended by the instructor. For example, if Twitter is used as a primary forum with an intent to foster dialogue among students, then one might want to see a higher ratio of interactive-type tweets such as replies. Whatever the use and intent, we recommend the instructor establish some baseline values of the ratios based on the first couple of weeks of the class (or data from the previous iteration of the same class) and then follow the changes in ratios over time to see whether there are any sudden changes and why. In our case, there were 444 messages with RTs (27% of the total number of messages), which is comparable to that found in other Twitter communities (Suh, Hong, Pirolli, & Chi, 2010; Zhou, Bandari, Kong, Qian, & Roychowdhury, 2010; Stieglitz & Dang-Xuan, 2012).

3.2.2 Following topics over time

In addition to using computer-led, top-down text analysis, the instructor may explore how a particular topic was discussed over time. Examining the distribution of messages over time may help to confirm whether students understand a new terminology after it has been introduced in the course and whether they are incorporating this new terminology as part of their vocabulary. There are couple of ways of doing this. One way is to build a chart showing the number of tweets mentioning a particular topic over time to confirm whether it was discussed in accordance with the syllabus. For example, Figure 2 shows that the words "theory" or "theories" were only mentioned by 66 Twitter users (14% of the 467 who participated in the class discussions on Twitter). The messages about theory concentrated around the

---

[1] IDs (Twitter usernames) and associated tweets are publicly available through the CCK11 newsletters and Twitter (e.g., see http://cck11.mooc.ca/archive/11/03_01_newsletter.htm, where it says, "If you use the CCK11 tag on Twitter, your Twitter posts will be collected and listed here").

57

second week of February and at the end of the course. Knowing this, the instructor can consider whether this accords with intentions, and adjust the syllabus or time on discussion about the topic.
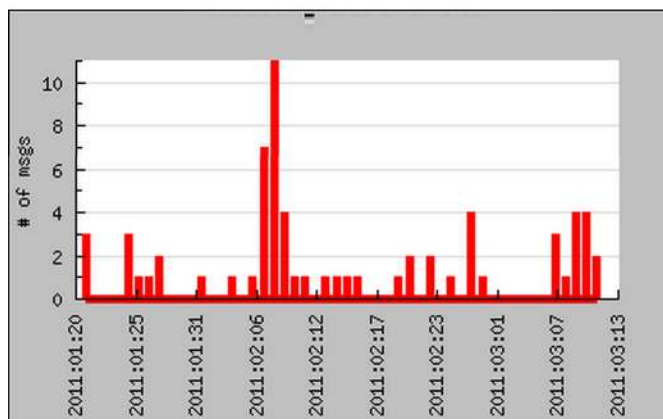

Figure 2. The number of tweets mentioning "theory" or "theories" over time.

Alternatively, the instructor may review frequently used words over time and compare them to the course outline. Figure 3 shows the patterns of frequently used terms over the span of the course. This allows instructors to see where discussion topics followed expected course topics (according to the course outline and scheduled readings for each week), and where discussion topics diverged from expected topics. For example, week 6 of the course focused on personal learning environments and networks, and yet these terms are largely absent from the dataset. Such an analysis could be used by instructors to review curriculum for that week to identify why discussion strayed far from the topic, and perhaps provide further scaffolding or engagement for student discussion to prompt further exploration of these concepts.
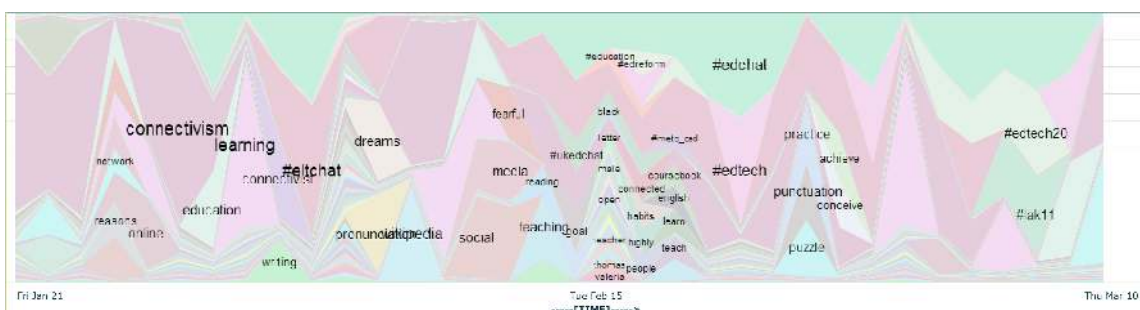

Figure 3. The relative number of tweets mentioning the top 100 frequently used words over time.

The visualization in Figure 3 potentially also allows instructors to discover patterns and relationships between concepts that emerge from learner discussions and that may influence future design of the course. For example, instructors may choose to re-sequence or potentially merge sections of the course based on how concepts and discussions co-occur or re-emerge in relation to the course design.

Overall, these simple forms of text analysis allow for the confirmation that topics of study are present in dialogue between learners, and discovery of potential relationships between concepts, course structures and sequences, and individuals and communities within a network.

Using these methods, a number of implications can be derived towards research and instructional practice. As Lockyer, Heathcote, and Dawson (2013) note, the field of learning analytics should be concerned with establishing a contextual framework that helps teachers interpret information provided by analytics to facilitate pedagogical action. By tracking the frequently used words over the course, instructors get an immediate sense of whether discussion within a course aligns with their intentions and expectations: Are students discussing the topics that instructors feel they ought to be? This allows educators and researchers to explore discussion in further detail to gain understanding of why the focus of discussion has followed or diverted from expectations of course developers, to inform decisions around changes to course curriculum, and to provide instructors with insight on when and what type of intervention and involvement in course discussions are necessary.

3.3   Network Analysis

3.3.1   Network Discovery and Visualization
The next step is to explore the social connections underlying the online conversations being examined. Studying online classes from a network perspective allows us to see how knowledge is being co-constructed. In this step, we first discover how online participants are connected to one another (e.g., who is talking to whom), and then apply SNA to analyze the discovered networks. SNA allows us to judge whether the communication networks formed as part of the class are effectively supporting processes known to contribute to successful learning, such as information sharing, community building, and collaboration.

To proceed with SNA, we built two types of communication networks: Name and Chain networks. The Name network shows connections between online participants based on direct interactions such as *replies* or indirect interactions such as *mentions* or *retweets*. In other words, two Twitter accounts will be connected in the Name network if one replies to, retweets, or mentions another in his/her message. By including indirect interactions such as *mentions* in addition to counting *replies*, we are able to capture instances when one person learns something from another as demonstrated by that person's retweets ("endorsement") or mentions ("acknowledgment"). The Chain network connects participants based on their posting behaviour and usually includes only direct interactions. In the case of Twitter, the Chain network is a subset of the Name network because it only connects people if one replied to another. Following the Twitter convention, this would be equivalent to starting a post with one's username, such as "*@gruzd Thank you for sharing this link*." Both Name and Chain networks have been validated and applied in different contexts, including online threaded discussions (Gruzd, 2009) and Twitter communities (Gruzd, Wellman, & Takhteyev, 2011). Gruzd (2009) found that name networks were a useful diagnostic tool for educators to evaluate and improve teaching models. These networks allowed for the identification of students who needed further support and attention from instructors, students

who were successful and/or took on leadership roles and were likely to be good candidates for peer support "learner-leaders," and students who were likely to be successful in working together on projects.

Another type of social media communication network that could be examined (but was not in this study for reasons discussed below) is a "Friends" or "Followers" network that consists of self-reported connections of *who is a friend/follower of whom*. This is a potentially useful network type; however, data to generate such networks are often inaccessible to researchers or hard to collect. Even if collected, it may not be the most useful data for studying learning networks. This is because self-reported networks are often incomplete, inaccurate, and may (and often do) reinforce pre-existing connections (Freeman & Romney, 1987; Bernard & Killworth, 1977; Bernard, Killworth, & Sailer, 1981; Marsden, 1990) that may or may not be activated during learning processes. In other words, two people do not need to be "friends" on Twitter for one person to read or even retweet other person's posts. (For a more in-depth discussion of how different social networks can be discovered from online data see Gruzd, 2014, and Gruzd & Haythornthwaite, 2011).

Figure 4 shows the Name and Chain networks built from the #CCK11 dataset. The node colours are assigned automatically (based on the "Fast Greedy" community detection algorithm; Clauset, Newman, & Moore, 2004). Each colour represents a group of nodes more likely to be connected to each other than with the rest of the network. In this manner, networks can be grouped into subsets, where each subset is densely connected internally relative to other nodes in the network. Such clustering can be useful in further research as communities correspond to clusters of nodes that may share common properties, interests, or have a similar role within a network (see Fortunato & Castellano, 2012).

Based on the visual inspection of the networks, it is clear that the Chain network is less dense with fewer nodes. This is somewhat expected since it only represents direct replies between online participants. The Name network is denser and shows a number of overlapping groups of nodes (clusters) that highlight potentially interesting areas of the network to focus on in more detail. The clustering and network fragmentation aspects are discussed later in this section.
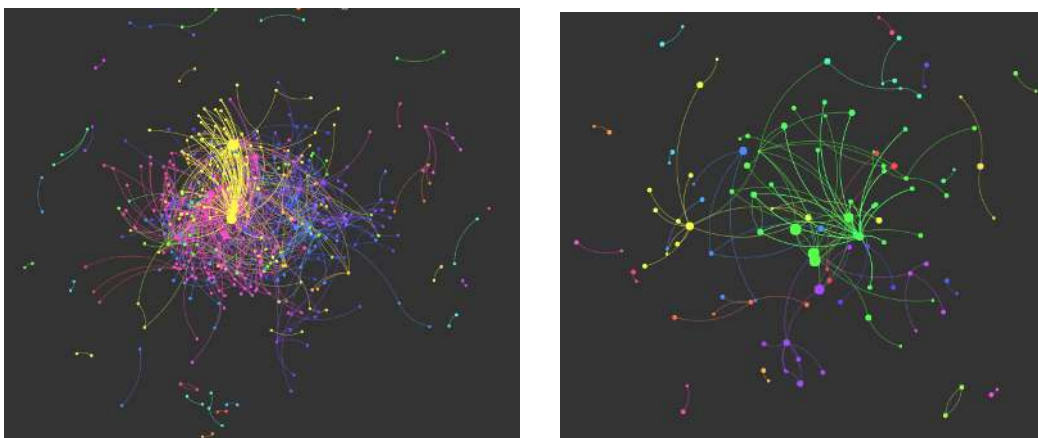

Figure 4. Name network (on the left) and Chain network (on the right).

Once the networks are discovered, we can use SNA to make sense of the emerging connections among online participants. With SNA, one can look at both micro- and macro-level measures to examine class interactions: micro-level measures provide insights at the individual node level; and macro-level measures capture the overall state of the network.

3.3.2 Micro-level SNA Measures

By calculating micro-level measures, such as various *centrality* measures, we can determine the most connected members in the class, showing who is influencing information flow in online discussions. Different centrality measures show different types of "influence." The three most used measures are *in-degree*, *out-degree,* and *betweenness* centrality (Dubois & Gaffney, 2014; Xu, Sang, Blasiola, & Park, 2014). *In-degree* suggests "prestige," highlighting the most mentioned or replied Twitter users; *out-degree* reveals active Twitter users with a good awareness of others in the network and who promote information to others; finally, *betweenness* shows actors located on the greatest number of information paths and who often connect different groups of users in the network. Table 3 shows the top 10 users based on these three measures for the Name network. (Due to the size limitation of this article, we will focus on the micro-level measures for the Name network only.)

Table 3. Top 10 Twitter users in the Name network based on centrality measures

IN-DEGREE	OUT-DEGREE	BETWEENNESS
participant1(m)	**cck11feeds**	**participant1(m)**
participant2(f)	**participant8(m)**	**participant8(m)**
gsiemens	**web20education**	**cck11feeds**
downes	**participant9(f)**	**participant11(f)**
guestLecturer3(m)	**participant6(m)**	guestLecturer12(f)
web20education	**participant1(m)**	**participant4(f)**
participant4(f)	participant10(f)	**participant9(f)**
participant5(m)	**participant4(f)**	**web20education**
participant6(m)	**participant7(m)**	**participant7(m)**
participant7(m)	**participant11(f)**	**gsiemens**
participant8(m)		participant13(m)

Notes: Users who appear in more than one column are in bold. The *in-degree* and *betweenness* lists contain 11 users instead of ten because the last two users in these lists share the 10th position. Course organizer and organization account usernames have been left intact, while individual learner participants and guest lecturer usernames have been replaced with pseudonyms, followed by their gender in parentheses.

The "in-degree" influencers include active participants: class facilitators/instructors (@gsiemens, @downes); educators (@participant1(m)/@participant5(m) [same person], @participant2(f), @participant6(m), @participant4(f), @participant7(m), @participant8(m)); bloggers and online resources (@web20education, @scoopit [ranked 15th]); guest speakers (@guestLecturer3(m), @guestLecturer12(f) [ranked 16th], @guestLecturer14(f) [ranked 19th]).

What is common across these users is that they were posting content that others in the class found relevant. However, there are different types of "influencers" and this can be seen by plotting the number of posts mentioning these users over time. Figure 5 shows what such a plot can reveal for two

61

sample users: @guestLecturer3(m) and @guestLecturer12(f). Both were guest speakers in the class whose content (such as presentation slides) was shared by class participants. While @guestLecturer3(m)'s posts resonated throughout the class, @guestLecturer12(f)'s impact on class discussions was most concentrated around a relatively small window of time closer to the end of the course. Arguably, the first type of "influencer" may be more desirable as the contribution sustains engagement throughout the course. (At the same time, some other guest speakers were not even active in Twitter discussions, and were only mentioned once or twice.) From another study of a Twitter community, we know that guest speakers or moderators are most effective in engaging the group if they are able to join the group conversation at least a couple of weeks prior to their own presentation (Gruzd & Haythornthwaite, 2013).
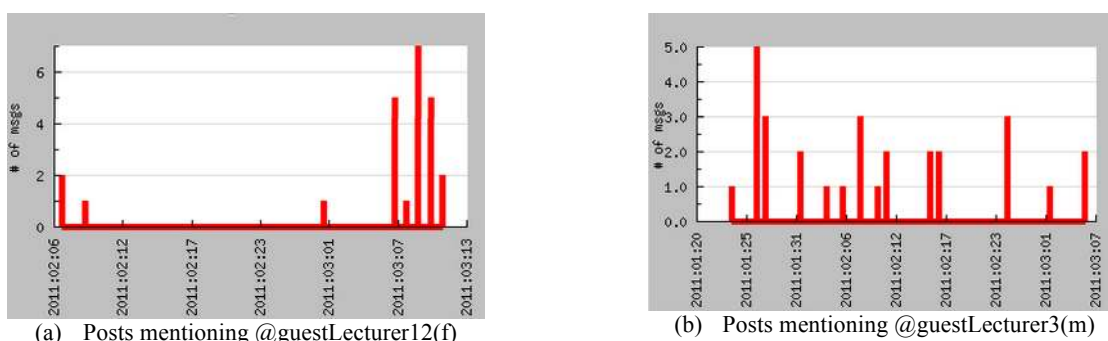

(a)   Posts mentioning @guestLecturer12(f)   (b)   Posts mentioning @guestLecturer3(m)

Figure 5. Number of Posts over Time.

Reviewing the top "out-degree" and "betweenness" accounts, a strong overlap can be seen between the users in the two lists, as well as with those who appear on the "in-degree" list. (Even the course's main account @cck11feeds, which is ranked high on both "out-degree" and "betweenness" lists, is in the top 20 based on the "in-degree" centrality.) We take this to be a good sign, indicating that most people who are influencing class discussions (ranked higher on the "in-degree" list) are also actively connecting with others in the class by engaging them in conversation and reposting their content (ranking high on the "betweenness" and "out-degree" lists).

The reason for the strong overall similarity between the "out-degree" and "betweenness" lists can also be explained by the observation that the Name network is not very fragmented. Even though there are some densely connected clusters (communities) formed in the Name network, as evident by the presence of different colour nodes in the network, there is a strong overall connectivity between these clusters. As a result, the measure of "betweenness" designed to identify users bridging communities is primarily showing highly connected users from the core of the network in the case of the #CCK11 dataset.

3.3.3   Macro-level SNA Measures
Macro-level measures found to be useful when analyzing and comparing different social networks include density, reciprocity, centralization, and modularity (Gruzd & Tsyganova, 2015). Table 4 summarizes the values of these measures for both the Name and Chain networks.

Table 4. Macro-level SNA Measures

	Name Network	**Chain Network**
Nodes	498*	122
Edges	761	125
Density	0.0031	0.0085
Diameter	38	7
Reciprocity	0.089	0.176
Centralization	0.070	0.075
Modularity	0.67	0.77

* The number of nodes is higher than the number of Twitter posters in the dataset because the Name network includes both those who posted using the #CCK11 hashtag and those who did not post using the class hashtag but were mentioned by others.

Density indicates the overall connectivity in the network (the total number of connections divided by the total number of possible connections); it is equal to 1 when everyone is connected to everyone. In our case, the Chain network is almost three times denser than the Name network, but both networks have less than 1% of the total number of possible connections. Although it is generally useful to see how dense a particular network is, caution is needed when interpreting this measure because with an increasing number of nodes in the social network, the density value often drops because it is much harder to maintain many connections in larger networks.

Diameter gives a general idea of how "wide" the network is; in other words, how many nodes information has to travel through between the two farthest nodes in the network. In mathematical terms, diameter is the longest of the shortest paths between any two nodes in the network. Smaller values for the diameter indicate a more highly connected network. The diameter measure is related to density; if *density* increases, we can expect *diameter* to reduce since there will be more paths for information to travel, thus potentially reducing the distance between online participants. In our case, the diameter is especially high and equal to 38 in the Name network. This means that it may take up to 38 connections for information to travel from one side of the network to the other. As a class facilitator, one may wish to keep the diameter low to ensure that information spreads efficiently in the network; however, when analyzing communication networks on social media, larger values of diameter may suggest that information originating inside the class also reaches people and communities far outside its core group of participants, which may be a positive sign.

Like density, we need to exercise caution in interpreting the benefits of low diameter values, and, indeed, the two-mode nature of ties — strong for sharing, weak for new information — suggests the utility of both forms (Haythornthwaite, 2002; 2015).

Reciprocity shows how many online participants are having two-way conversations. In a scenario when everyone replies to everyone, the reciprocity value will be 1. However, that almost never happens in social media conversations with hundreds or more online participants. The reciprocity of the CCK11 networks is discussed in more detail below.

Centralization indicates whether a network is dominated by only a few central participants (where centralization values are closer to 1), or whether more people are contributing to discussion and information dissemination (where centralization values are closer to 0). Communication networks that promote collaborative learning and knowledge co-creation might be expected to exhibit lower values of centralization than those with a lecturer and audience organization. Centralization values in both Name and Chain networks appear to be closer to 0, suggesting that both networks contain a number of influential participants but power is not concentrated in the hands of the few.

Finally, *modularity* provides an estimate of whether a network consists of one coherent group of participants engaged in the same conversation and paying attention to each other (modularity values closer to 0); or whether a network consists of different conversations and communities with a weak overlap (modularity values closer to 1). For more formal collaborative classes, the goal might be to achieve a network structure with a lower modularity value — i.e., everyone on the same topic attending to everyone else — potentially leading to a higher sense of community. At the same time, especially when designing a network to support informal learning, a network with a moderate number of overlapping communities (modularity values around 0.5) may be more desired as it would potentially expose participants to diverse sources of information, exercising the strength of weak ties while still maintaining the sense of community (Shen, Nuankhieo, Huang, Amelung, & Laffey, 2008). In the case of #CCK11, the Name network consists of both weak and strong ties as suggested by a moderate value of modularity (0.67). However, the modularity value of the Chain network is a bit higher and closer to 1 (0.77), suggesting that there are different groups of people having different conversations in the class. Higher values of modularity may be a sign of underlying homophilic tendencies of people to connect with other like-minded individuals. A class facilitator could follow this measure to gauge the extent of fragmentation of discussions into smaller groups and evaluate this in relation to class design (e.g., group project discussions).

Based on the discussion above, it is clear that some measures such as *centralization* and *modularity* can be interpreted relatively easily; however, other measures, such as *diameter* or *reciprocity,* are more difficult to explain without a point of reference. To help with the interpretation, we can compare our values to the values of the same measures calculated for other Twitter networks of a similar size. We will use reciprocity as an example. The Name network's reciprocity level is 0.089, which means that about 9% of the total number of ties is reciprocal (or bi-directional). The Chain network's reciprocity is 0.176 (or about 18% of the total number of ties). It is expected that the Chain network will be more reciprocal since it includes only connections when one person replies to another. However, is 9% or 18% few or many? To answer this question, a simulation can be run to generate a number of random networks with similar characteristics to test whether the observed values of reciprocity are likely to appear by chance alone. Such simulations and testing can be done, for example, using Exponential Random Graph Models (ERGM; Hunter, Handcock, Butts, Goodreau, & Morris, 2008).

However, the average instructor might not be equipped with the expertise or proper computing resources to run such tests. Therefore, as a lightweight analytical approach, one can consider comparing

64

the SNA values calculated using the observed networks to the values from other networks of a similar size built using the same method (either Name or Chain). For example, Figure 6 shows the scatter plot of the number of nodes versus the reciprocity values for about 100 communication networks built from various Twitter datasets. The plots reveal that in both cases, Name and Chain networks, the values for the CCK11 class (marked with the red star), is somewhat higher than in the majority of other networks. This means that the CCK11 class is reaching or exceeding the level of reciprocity that would normally be expected in Twitter data, a reassuring sign for the class facilitators that they are on the right track in terms of engaging class participants in two-way conversations on Twitter.
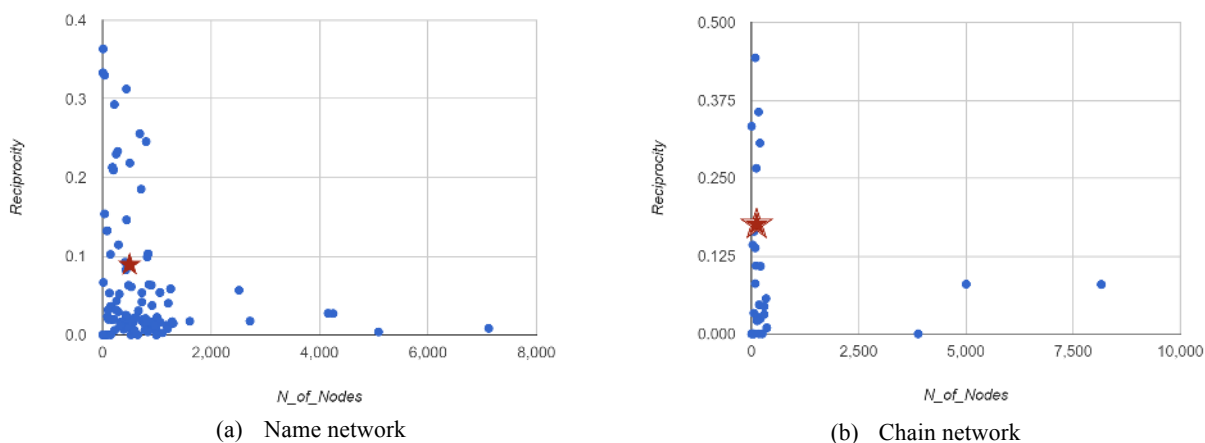

(a) Name network   (b) Chain network

Figure 6. Reciprocity versus network size in Twitter networks.

4   CONCLUSIONS

This chapter has described approaches to learning network analytics that open up possibilities for understanding designed and emergent online learning practices as supported through social media. The use of social media, and its implementation in teaching and learning is new, but advancing rapidly. Unlike earlier waves on online education, both Twitter and MOOC environments are appearing within the context of social media practice. Learners are immersed already in the presence and use of social media, and thus come to learning via social media as an additional means of information search and acquisition, learning community support and engagement, and knowledge building.

The challenge is to come to a nuanced understanding of the multiple facets of learning online via social media, exploring both the pros and cons of social network high and low density, reach, and reciprocity, and the merits or not of coherence on topic discussion. For formal settings, it is necessary to consider the intent of the instructor and to examine network and discussion formation in light of the match to intended and desired communication and pedagogical outcomes. For informal settings, we may be more interested in the societal level impact of mass learning, massively distributed learning, and just-in-time learning associated with social media exchanges and how these are balanced with the development of sustained learning communities.

Overall, as we have outlined here, we find the multi-method approach that looks at the combined effects of social network and topic discussion a promising one for discovery on new learning practices. Combined with understanding of local contexts and patterns of behaviour across multiple contexts, we expect to see important research contributions to come that contribute to our understanding of 21st century learning practices.

ACKNOWLEDGEMENTS

REFERENCES

Banks, J., Au, K., Ball, A., Bell, P., Gordon, E., Gutierrez, K., Heath, S., et al. (2007). *Learning in and out of school in diverse environments* (Consensus Report). Learning in Informal and Formal Environments (LIFE) Center. Retrieved from http://life-slc.org/docs/Banks_etal-LIFE-Diversity-Report.pdf

Bernard, H. R., & Killworth, P. D. (1977). Informant accuracy in social network data II. *Human Communication Research*, *4*(1), 3–18.

Bernard, H. R., Killworth, P., & Sailer, L. (1981). Summary of research on informant accuracy in network data and the reverse small world problem. *Connections*, *4*(2), 11–25. http://dx.doi.org/10.1111/j.1468-2958.1977.tb00591.x

Boguraev, B., & Kennedy, C. (1999). Salience-based content characterisation of text documents. *Advances in Automatic Text Summarization*, 99–110. Cambridge, MA: MIT Press.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, *110*(2), 349–399.

Carley, K. M. (1997). Network text analysis: The network position of concepts. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 79–100). London: Routledge.

Carley, K., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social Forces*, *70*(3), 601–636. http://dx.doi.org/10.1093/sf/70.3.601

Chen, B., & Bryer, T. (2012). Investigating instructional strategies for using social media in formal and informal learning. *The International Review of Research in Open and Distributed Learning, 13*(1), 87–104. http://dx.doi.org/10.19173/irrodl.v13i1.1027

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, *70*(6). https://dx.doi.org/10.1103/PhysRevE.70.066111

Corman, S. R., Kuhn, T., McPhee, R. D., & Dooley, K. J. (2002). Studying complex discursive systems. *Human Communication Research*, *28*(2), 157–206. http://dx.doi.org/10.1111/j.1468-2958.2002.tb00802.x

Dabbagh, N., & Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and Higher Education*, *15*(1), 3–8. http://dx.doi.org/10.1016/j.iheduc.2011.06.002

Dahlstrom, E., Walker, J. D., & Dziuban, C. (2013). *ECAR study of undergraduate students and information technology, 2013*. Louisville, CO: EDUCAUSE Center for Analysis and Research.

Daly, A. J. (Ed.) (2010). *Social network theory and educational change*. Cambridge, MA: Harvard Education Press.

Dawson, S., Bakharia, A., Heathcote, E. (2010). SNAPP: Realizing the affordances of real-time SNA within networked learning environments. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (NLC 2010), 3–4 May 2010, Aalborg, Denmark (pp. 125–133).

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, *46*(1), 6–28. http://dx.doi.org/ 10.1016/j.compedu.2005.04.005

Downes, S. (2006). Learning networks and connective knowledge. *Instructional Technology Forum*, (Paper 92). Retrieved from http://itforum.coe.uga.edu/paper92/paper92.html

Dubois, E., & Gaffney, D. (2014). The multiple facets of influence: Identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*, *58*(10), 1260–1277. http://dx.doi.org/10.1177/0002764214527088

Eraut, M. (2010). Informal learning in the workplace. *Studies in Continuing Education*, *26*, 247–273. http://dx.doi.org/10.1080/158037042000225245

Eynon, R., Schroeder, R., & Fry, J. (2016). The ethics of learning and technology research. In C. Haythornthwaite, R. Andrews, J. Fransman, & E. Meyers (Eds.), *The SAGE Handbook of E-learning Research* (pp. 211–231). London: SAGE.

Fiedler, S., & Väljataga, T. (2011). Personal learning environments: Concept or technology? *International Journal of Virtual and Personal Learning Environments*, *2*(4), 1–11. http://dx.doi.org/10.4018/jvple.2011100101

Fortunato, S., & Castellano, C. (2012). Community structure in graphs. In R.A. Meyers (Ed.), *Computational complexity* (pp. 490–512). New York: Springer. http://dx.doi.org/10.1007/978-1-4614-1800-9_33

Freeman, L., & Romney, A. (1987). Words, deeds and social structure: A preliminary study of the reliability of informants. *Human Organization*, *46*(4), 330–334. http://dx.doi.org/10.17730/humo.46.4.u122402864140315

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Gruzd, A. (2009). Studying collaborative learning using name networks. *Journal of Education for Library & Information Science*, *50*(4), 237–247.

Gruzd, A. (2014). Online communities. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining.* New York: Springer. http://dx.doi.org/10.1007/978-1-4614-6170-8_81

Gruzd, A., & Haythornthwaite, C. (2008). The analysis of online communities using interactive content-based social networks. *Proceedings of the American Society for Information Science and Technology Conference* (ASIS&T 2008), 24–29 October 2008, Columbus, OH, USA (pp. 523–527). http://dx.doi.org/10.1002/meet.2008.1450450318

Gruzd, A., & Haythornthwaite, C. (2011). Networking online: Cybercommunities. In J. Scott & P. Carrington (Eds.), *Handbook of social network analysis* (pp. 449–487). London: Sage.

Gruzd, A., & Haythornthwaite, C. (2013). Enabling community through social media. *Journal of Medical Internet Research*, *15*(10), e248. http://dx.doi.org/10.2196/jmir.2796

Gruzd, A., Haythornthwaite, C., Paulin, D., Absar, R., & Huggett, M. (2014, March). Learning analytics for the social media age. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 254–256. http://dx.doi.org/10.1145/2567574.2576773

Gruzd, A., & Tsyganova, K. (2015). Information wars and online activism during the 2013/2014 crisis in Ukraine: Examining the social structures of pro- and anti-Maidan groups. *Policy & Internet*, *7*(2), 121–158. http://dx.doi.org/10.1002/poi3.91

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an imagined community. *American Behavioral Scientist, 55*(10), 1294–1318. http://dx.doi.org/10.1177/0002764211409378

Haythornthwaite, C. (1999). Collaborative work networks among distributed learners. *Proceedings of the 32nd Hawaii International Conference on System Science* (HICSS-32), 4–8 January 1999, Maui, HI, USA (pp. 16–32). IEEE Computer Society. http://dx.doi.org/10.1109/HICSS.1999.772707

Haythornthwaite, C. (2002). Building social networks via computer networks: Creating and sustaining distributed learning communities. In K. A. Renninger & W. Shumar, *Building virtual communities: Learning and change in cyberspace* (pp. 159–190). Cambridge, UK: Cambridge University Press.

Haythornthwaite, C. (2008). Learning relations and networks in web-based communities. *International Journal of Web Based Communities*, *4*(2), 140–158. http://dx.doi.org/10.1504/IJWBC.2008.017669

Haythornthwaite, C. (2011). Learning networks, crowds and communities. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 18–22. https://doi.org/10.1145/2090116.2090119

Haythornthwaite, C. (2015). Rethinking learning spaces: Networks, structures and possibilities for learning in the 21st century. *Communication, Research and Practice, 1*(4), 292–306. http://dx.doi.org/10.1080/22041451.2015.1105773

Haythornthwaite, C., & Andrews, R. (2011). *E-learning theory and practice*. London: Sage.

Haythornthwaite, C., & de Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (NLC 2010), 3–4 May 2010, Aalborg, Denmark (pp. 183–190). Retrieved from

http://www.lancaster.ac.uk/fss/organisations/netlc/past/nlc2010/abstracts/PDFs/Haythornwaite.pdf

Haythornthwaite, C., de Laat, M., & Schreurs, B. (2016). A social network analytic perspective on e-learning. In C. Haythornthwaite, R. Andrews, J. Fransman, & E. Meyers, *Handbook of e-learning research, 2nd ed.* (pp. 251–269). London: Sage.

Haythornthwaite, C., & Gruzd, A. (2007). A noun phrase analysis tool for mining online community conversations. In *Communities and Technologies 2007: Proceedings of the Third Communities and Technologies Conference, Michigan State University 2007* (pp. 67–86). London: Springer. http://dx.doi.org/10.1007/978-1-84628-905-7_4

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ERGM: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3).

Ito, M., Gutierrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salen, K., & Watkins, S. C. (2013). *Connected learning: An agenda for research and design*. Digital Media and Learning Research Hub.

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 15–24. https://doi.org/10.1145/2883851.2883950

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Newbury Park, CA: Sage.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology (3rd Ed)*. Thousand Oaks, CA: Sage.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Liddy, E. D. (1998). Enhanced text retrieval using natural language processing. *Bulletin of the American Society for Information Science and Technology*, *24*(4), 14–16. http://dx.doi.org/10.1002/bult.91

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action aligning learning analytics with learning design. *American Behavioral Scientist*, *57*(10), 1439–1459. http://dx.doi.org/10.1177/0002764213479367

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2), 588–599. http://dx.doi.org/10.1016/j.compedu.2009.09.008

Marsden, P. V. (1990). Network data and measurement. *Annual review of sociology*, 435–463.

Mcloughlin, C., & Lee, M. J. W. (2010). Personalised and self-regulated learning in the Web 2.0 era: International exemplars of innovative pedagogy using social software. *Australasian Journal of Educational Technology*, *26*(1), 28–43.

Moran, M., Seaman, J., & Tinti-kane, H. (2012). *How today's higher education faculty use social media*. Boston, MA: Pearson. Available at http://www.pearsonlearningsolutions.com/higher-education/social-media-survey.php

Pennebaker, J. W. (2003). The social, linguistic, and health consequences of emotional disclosure. In J. Suls & K. A. Wallston (Eds.) *Social psychological foundations of health and illness* (pp. 288–313). Malden, MA: Blackwell.

Ravenscroft, A. (2011). Dialogue and connectivism: A new approach to understanding networked learning. *International Review of Research in Open and Distance Learning*, *12*(3), 139–160. http://dx.doi.org/10.19173/irrodl.v12i3.934

Rienties, B., Héliot, Y., & Jindal-Snape, D. (2013). Understanding social learning relations of international students in a large classroom using social network analysis. *Higher Education*, *66*(4), 489–504. http://dx.doi.org/10.1007/s10734-013-9617-9

Rogers, T., Dawson, S., & Gašević, D. (2016). Learning analytics and the imperative for theory driven research. In C. Haythornthwaite, R. Andrews, J. Fransman, & E. Meyers (Eds.), *The SAGE handbook of e-learning research* (pp. 232–250). London: SAGE.

Rorty, R. (1992). *The linguistic turn: Essays in philosophic methods*. Chicago, IL: University of Chicago Press.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., et al. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer Supported Collaborative Learning, 3*(3), 237–271. http://dx.doi.org/10.1007/s11412-007-9034-0

Rubin, V. L., Stanton, J. M., & Liddy, E. D. (2004). Discerning emotions in texts. In *The AAAI Symposium on Exploring Attitude and Affect in Text* (AAAI-EAAT), (pp.124–128). Retrieved from http://surface.syr.edu/cgi/viewcontent.cgi?article=1041&context=istpub

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.)*, Cambridge handbook of the learning sciences* (pp. 97–118). New York: Cambridge University Press.

Shen, D., Nuankhieo, P., Huang, X., Amelung, C., & Laffey, J. (2008). Using social network analysis to understand sense of community in an online learning environment. *Journal of Educational Computing Research*, *39*(1), 17–36. http://dx.doi.org/10.2190/EC.39.1.b

Siemens, G. (2005, January). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*. Retrieved from http://www.itdl.org/journal/jan_05/article01.htm

Siemens, G. (2008). Learning and knowing in networks: Changing roles for educators and designers. *ITFORUM for Discussion*, 28 January 2008, 1–26.

Smith, S. D., & Caruso, J. B. (2010). *The ECAR study of undergraduate students and information technology, 2010: Key findings*. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED514182

Statista. (2015). *Statistics and facts about Facebook*. Retrieved from http://www.statista.com/topics/751/facebook/

Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through microblogging: An empirical analysis of sentiment in Twitter messages and retweet behavior. *Proceedings of the*

70

45th Hawaii International Conference on System Sciences (HICSS-45), 4–7 January 2012, Maui, HI, USA (pp. 3500–3509). IEEE Computer Society. http://dx.doi.org/10.1109/HICSS.2012.476

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers in Education, 46*(1), 29–48. http://dx.doi.org/10.1016/j.compedu.2005.04.002

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010, August). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. *Proceedings of the 2nd International Conference on Social Computing* (SocialCom 2010), 20–22 August 2010, Minneapolis, MN, USA (pp. 177–184). IEEE Computer Society. http://dx.doi.org/10.1109/SocialCom.2010.33

Twitter. (2015). *Twitter usage and company facts*. Retrieved from https://about.twitter.com/company

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

Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, S. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM 2014), 4–7 July, London, UK (pp. 130–137): International Educational Data Mining Society.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems Thinker*, *9*(5), 2–3.

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. http://dx.doi.org/10.18608/jla.2015.22.2

Xu, W. W., Sang, Y., Blasiola, S., & Park, H. W. (2014). Predicting opinion leaders in twitter activism networks: The case of the Wisconsin recall election. *American Behavioral Scientist*, *58*(10), 1278–1293. http://dx.doi.org/10.1177/0002764214527091

YouTube. (2015). *Statistics.* https://www.youtube.com/yt/press/statistics.html

Zhou, Z., Bandari, R., Kong, J., Qian, H., & Roychowdhury, V. (2010). Information resonance on Twitter: Watching Iran. *Proceedings of the 1st Workshop on Social Media Analytics, at the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, (pp. 123–131). New York: ACM. https://dx.doi.org/10.1145/1964858.1964875

Ziegler, M. F., Paulus, T., & Woodside, M. (2014). Understanding informal group learning in online communities through discourse analysis. *Adult Education Quarterly*, *64*(1), 60–78. http://dx.doi.org/10.1177/0741713613509682