

WORKING THE BACKCHANNEL: ANALYSIS OF A LITERACY ORGANIZATION'S CONFERENCE HASHTAG

Peggy Semingson *

Henry Anderson

Elizabeth Powers

Pete Smith

The University of Texas at Arlington

**Address all correspondence to Peggy Semingson, Department of Curriculum and Instruction, The University of Texas at Arlington, 502 Yates St., Arlington, TX, 76019; peggys@uta.edu*

Abstract

Can social network analysis (SNA) combined with natural language processing (NLP) provide greater insights into complex online discussions? Using this combined approach, these researchers worked to better understand the complex roles and professional identities of participants in large-scale, online professional discussion: the NCTE 2016 annual conference Twitter backchannel.

Professional educational organizations are increasingly seeking ways to foster digital connectivity and informal learning among membership through social media platforms and tools. Professional organizations frequently use Twitter and a conference hashtag to foster participatory and networked approaches to learning while developing an online community around the event, yet these backchannels are seldom studied. This paper applies an integrative approach to exploration of the Twitter backchannel of the National Council of Teachers of English (NCTE) 2016 national conference. A social network analysis (SNA) was utilized to better understand the digital connectivity of key professional players and influencers at this literacy conference, and in a parallel thread, Linguistic Inquiry and Word Count (LIWC) software was used to document the tone, clout, authenticity, and analytical thinking of participants in the conference Twitter data set.

SNA findings indicated distinct classes of educational practitioners within the Twitter conversation, such as those from a major publishing house, practitioner-oriented authors, children-oriented authors, and consultants. Subsequent content analysis using LIWC indicated that these practitioners' conversations were decidedly analytical, positive in emotional tone, and contained language indicative of experts and expertise. This combined approach yielded insights which SNA or NLP alone did not provide.

KEY WORDS: social media, Twitter, networked learning, professional development, social network analysis

1. INTRODUCTION

A networked participatory scholarship (Veletsianos and Kimmons, 2012) offers participants access to engage with a professional learning network (PLN), to both share and consume knowledge related to a knowledge domain. “Professional learning network” is sometimes used interchangeably with the term “Professional learning community” (PLC) in the literature. A PLC is “[a] group of people sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way” (Stoll, et al., 2006, p. 223).

Two such online PLN/PLC practices are scheduled Twitter chats (Megele, 2014) and conference-specific Twitter backchannels, where participants virtually join with others at a designated and limited time frame, with a specific hashtag, to collaboratively discuss the chat or conference themes. During Twitter chats and conference-specific Twitter backchannels, participants contribute knowledge in varying forms, both theoretical and practitioner-based. Megele (2014) provides a definition of a Twitter chat, noting that “Twitter chat is a thematic multilogue (i.e., a many-to-many conversation focused on a given theme/topic) often situated within a community of practice (CoP) and/or community of interest (CoI)...” (p. 47).

As a case study of one such organization (Stake, 2005), this paper examines the emerging and exploratory use of several methods of linguistic and computational analysis of a moderately large set of Twitter posts connected to the November 2016 conference of the National Council of Teachers of English (NCTE), a prominent literacy organization, which used the official hashtag #NCTE2016, and a commonly-used variant, #NCTE16.

The current study originated from a shared interest in how professional educational organizations regulate discourse among their members in explicitly discussion-oriented contexts, such as Twitter, and what methodological approaches are appropriate to a comprehensive analysis of such discourse. This was realized via a two-part investigation into NCTE's Twitter backchannel. First, the researchers identified the need to examine the networked and connected learning (Siemens, 2005) that was taking place within the organization, a question that lends itself to approaches from Social Network Analysis (SNA). Second, the research team was interested in examining the language used by participants, and identified tools from Natural Language Processing (NLP) as the most appropriate way to accomplish this. The researchers then concluded that a methodology that fully integrates both the more quantitative SNA tools and the more qualitative NLP

tools is necessary to fully understand the nature of such discourse. In this study, Gephi (Bastian et al., 2009) was selected for the SNA portion; Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015a) was used to perform a computational analysis of Twitter-based corpora.

To collect the data, the second author (Anderson) used Twitter's publicly available application programming interfaces (APIs) to collect tweets using the #NCTEChat ("National Council of Teachers of English chat") hashtag, and the #NCTE2016 and #NCTE16 hashtags (the 2016 official conference hashtags). During collection, we discovered that there were far more conference tweets than NCTEChat tweets, and we shifted focus to the larger data set.

During preliminary investigations, we hypothesized that influential individuals within the organization would be influential in the discussions. This led us to the field of SNA, where a range of tools and algorithms exist to identify prominent and influential nodes within a network.

1.1 Guiding Research Questions

The researchers used two guiding questions to shape the case study of NCTE's Twitter backchannel and guide the development of the methodology:

1. What insights can be gained about organizationally regulated online discussions by combining established approaches and methodologies from the fields of social network analysis, linguistics, and natural language processing—insights that might not be available when only drawing on SNA or NLP alone?
2. Can such integrative approaches help researchers to better understand the complex roles and professional identities of key influencers in large-scale, online professional discussions, such as the NCTE 2016 annual conference Twitter backchannel?

2. THEORETICAL FRAMEWORK AND LITERATURE

2.1 Connectivism as an Overarching Framework

The current research draws primarily on connectivist theories of learning and human development (e.g., Siemens, 2005), which recognize the connections between learners as a necessary part of the individual's knowledge creation process. Siemens (2005), in a seminal article on connectivism, describes personal knowledge as a network, "which feeds into organizations and institutions, which in turn feed back into the network, and then continue to provide learning to the individual." We also draw on the ideas from networked learning on professional learning networks (PLNs); such networks have the potential to sustain learning beyond a single workshop (e.g. Britt and Paulus, 2016).

Both theories recognize the importance of a network in lifelong professional learning, such that this “cycle of knowledge allows learners to remain current in the field through the connections they have formed” (Siemens, 2005). This emphasis on connections between learners has become more evident with the increasing use of digital and social media such as Facebook and Twitter as platforms to share personal and professional information, and the digital records these platforms create (Greenwood et al., 2016).

In an increasingly connected world, Siemens (2005) argues that, “The capacity to form connections between sources of information, and thereby create useful information patterns, is required to learn in our knowledge economy.” Li and Greenhow (2015) concur, explaining that “from this [connectivist] perspective, being knowledgeable can be seen as the ability to nurture, maintain, and traverse these connections; to access and use specialized information sources just-in-time.”

3. TWITTER AS PROFESSIONAL ONLINE DIALOGUE

3.1 Twitter as a Networked Discourse Community

Gillen and Merchant (2013) suggest that Twitter, as a New Literacy and Web 2.0 practice, provides a meaningful space for online dialogue. The researchers, through an autoethnographic study of their own Twitter practices, suggest that use of Twitter can serve many functions and purposes including, but not limited to such practices as: a backchannel for a conference, crowdsourcing information, and social networking. In addition to the varied purposes of Twitter usage, as a form of computer-mediated-communication, it can function as a network of data and communication. Such networked computer-mediated-communication can be studied through an examination of measures such as the *centrality*, which measures how well-connected a given participant is to all other participants, and typically calculated through social network analysis (Enriquez, 2010; Ryymin et al., 2008).

3.2 Twitter as Connected Learning for Educators

Recent scholarship has noted the ways that Twitter has served learning purposes for educators (Britt and Paulus, 2016; Davis, 2015; Rehm and Notten, 2016; Visser et al., 2014). Carpenter and Krutka (2014) note that educators use Twitter for their own professional learning for its affordances such as personalized learning and ability to overcome isolation in the profession. Overall, the idea of educators using Twitter can be seen as a form of “DIY professional development” in times of increasingly connected and mobile learning (e.g., as discussed by Biddolph and Curwood, 2016). Importantly, Biddolph and Curwood suggest that more research is needed in areas related to online professional development and teacher communities of practice.

3.3 The Nature of “Influencing” on Twitter

The Twitter social network allows up to 140 characters in a “tweet,” and although these may be original content, tweets can also include links to other electronic material, forwarded material from others' tweets (retweets–RT), acknowledgment of other Twitter users (include @Name in text) or a connection to a specific community (hashtag #CommunityName). Twitter users can also opt to follow other users and the tweets from those being followed will appear in their timelines.

Twitter use by participants in a conference appears to follow general use for social media, where a few users in a social network platform post a large number of the total posts (Chen, 2011; Nielson, 2006). Chen's 2011 study of seven academic conferences between 2009 and 2011 found that this skewed distribution of postings roughly follows the 80/20 rule, where 20 percent or less of the “tweeters” post 80 percent or more of the tweets.

In a conference Twitter space marked by a dedicated hashtag, influence can be measured by the number of followers for a particular user, the number of each user's tweets, the number of re-tweets for each user, and each user's overall connections compared to others' posts including the hashtag.

4. SOCIAL NETWORK ANALYSIS AND NATURAL LANGUAGE PROCESSING

The authors use social network analysis as a primary analytical framework (for a brief overview of SNA in the social sciences, see Borgatti et al., 2009), under which the data set was conceptualized: consisting of participants, or *nodes*, whose interactions are represented as lines, or *edges*, connecting them. The overall pattern of participants and their interactions (*nodes* and *edges*) forms a *network*. The network can be analyzed quantitatively, using automated analysis, as well as qualitatively, via manual and visual inspection. SNA has been applied to a wide range of problems in social media in general and on Twitter specifically, including detecting latent and highly influential networks (Huberman et al., 2008) and mapping the evolution of conversations over time (Bruns, 2011).

The present authors also draw on techniques from natural language processing to analyze the NCTE data set. As with SNA, there has been extensive work with NLP toolsets on both Twitter and other social media data; selected examples include the automated extraction of information from Twitter (Verma et al., 2011), identification of belief structures (Fast and Horovitz, 2016), and analysis of differing language use by ethnically diverse users (Blodgett et al., 2016).

Despite the rich applications of SNA and NLP tools to social media data, the researchers have identified a relative lack of research that incorporates and integrates approaches from both fields simultaneously. While there has been some work that does incorporate both, it often prioritizes methodologies from one of the two fields, and uses those from the other as secondary, augmenting materials (Ebrahimi et al., 2016). The authors have an interest in applying both domains in concert, rather than using one to augment the other.

5. BACKGROUND AND CONTEXT OF THE LITERACY ORGANIZATION

This section of the paper provides a contextual framework of the literacy organization in order to better understand the current study.

6. BACKGROUND OF THE LITERACY ORGANIZATION

The National Council of Teachers of English, established in 1911 (<http://www.ncte.org/centennial>), is an international organization which provides members with a way to connect and grow as literacy-focused professionals. Their mission statement is as follows:

“The Council promotes the development of literacy, the use of language to construct personal and public worlds and to achieve full participation in society, through the learning and teaching of English and the related arts and sciences of language.”

(<http://www.ncte.org/mission>)

In addition to an annual conference, NCTE provides open-access digital resources, including social media channels (such as Twitter and Facebook), a primarily member-written blog (<http://blogs.ncte.org/>), online discussion forums, and a website (<http://www.ncte.org/Default.aspx>), in addition to traditional platforms such as print and online journals (which are member-only with limited free access to publications). These can be explored at their community page: <http://www.ncte.org/community>. Peggy Semingson, the first author of this paper, is a member of the National Council of Teachers of English and a contributor to the blog. Additionally, Semingson's affiliation with the literacy organization and the literacy field of study provided important contextual information for the analyses carried out in this study.

7. DIGITAL MEDIA OUTREACH OF NATIONAL COUNCIL OF TEACHERS OF ENGLISH

NCTE fosters digital connection of members and others using a variety of digital and print platforms and mediums. Many of these opportunities are social media-driven. For instance, the organization social media channels include Facebook, Twitter, regularly scheduled Twitter chats, a Pinterest page, Instagram, YouTube, LinkedIn, and scheduled “#NCTEonAir” live video conferencing and webinar sessions. Notably, recorded live

sessions are posted on the website and Twitter chats are curated and archived for later review. All of these archived materials are freely accessible via the NCTE website.

The focus of this study is on the online community and connective structure of the Twitter discourse surrounding and within a literacy organization's annual conference. To contextualize the Twitter discourse, it is helpful to discuss NCTE's activity on the platform. At the time of writing this, NCTE's Twitter handle (@ncte) has over 51,000 followers and 47,000 tweets, indicating a large following and an active twitter presence. The account has been active since its creation in May 2009. The organization's tweets regularly contain content relevant to literacy professionals, including links and retweets of other literacy organizations and NCTE affiliates. The organization promotes the #NCTEchat hashtag, which is used for recurring monthly discussions (organized by NCTE) and general year-round discussion. NCTE's consistent and high volume of online activity makes the organization a natural choice for these investigations.

8. SUMMARY OF RECENT TWITTER CHATS AS REGULAR CONNECTED LEARNING

Between September 2013 and April 2017, NCTE has hosted regular Twitter chats nearly every month. Beginning in 2015, NCTE has maintained a blog to cross-promote and market upcoming chats, as well as disseminate recommended discussion questions. Past chats have included questions such as “What are your core beliefs about the teaching of writing?” (in November 2016) and “Share an idea for celebrating the National Day on Writing this Thursday, Oct. 20” (in October 2016). Following each Twitter chat, participants' tweets are compiled using Storify (<https://storify.com/>) and linked to the NCTE chat page (<http://www.ncte.org/community/nctechat>).

9. OVERARCHING 2016 CONFERENCE THEMES AND TRENDS

To contextualize the setting, focus, and overall trends within the 2016 NCTE conference, these researchers examined the conference program. The conference theme was “Faces of Advocacy,” embracing an ongoing theme of advocacy that is visible and integrated presently and historically across NCTE's website, media, and policy positions statements (<http://www.ncte.org/annual/past/2016program>). The conference took place over four days, from November 16–19, 2016. Events included interactive workshops, panel presentations, film screenings, author presentations, awards, keynote speakers, and breakout sessions. The overarching themes were focused on topics of practical relevance to literacy educators at a wide range of levels.

10. METHODS

10.1 Data Collection

In this section, we describe an overview and procedure for collecting data for this study. We describe what constituted the data set and what were the parameters of the data set.

10.2 Overview

The primary data set consists of a total of 45,198 tweets from 7,824 unique participants, totaling just over one million words. When retweets are excluded, as they were for the linguistic analysis (they were still included in social network analysis), this total drops to 21,243 tweets, 444,000 words, and 2,577 unique participants (the remaining participants only posted retweets within the data set). Omitting hashtags, @-mentions, and URLs further decreases the word count to just over 318,000 (these terms are omitted from the linguistic analysis due to incompatibility with LIWC).

The authors collected the tweets through Twitter's "Firehose" streaming API, which provides access to all tweets posted to the site in real time. The stream was filtered to return only tweets using either of the conference hashtags, #NCTE16 and #NCTE2016 (and all capitalization variants, e.g. #ncte16). The tweets were collected continuously from November 13th through January 29th.

10.3 Processing of Tweets

After collection of the tweets, Henry Anderson, the second author of this paper, wrote a Python script to read the resulting data and extract the following information for each tweet: (1) who posted it; (2) what other participants, if any, were mentioned or retweeted in it; (3) its text.

For each participant in the data set, the script also collected: (1) their numerical Twitter ID; (2) their screen name (beginning with "@", e.g. @ncte); (3) their display name (e.g., NCTE). These data were collected through the Twitter Search API after tweet collection. The authors utilized these data to construct a network graph for the conference, with each participant as a node, and each retweet, @-mention, or reply as an edge between nodes. The resulting network was analyzed in Gephi (Bastian et al., 2009), while the text of the participants' Tweets was analyzed using LIWC.

Gephi is a free and open-source tool for interactively visualizing and analyzing networks and graphs (Bastian et al., 2009), which was selected for its ease of use, quality of visualizations, and suite of analytical tools. Gephi offers a range of social network analysis algorithms, including measures of degree, which counts the total number of connections a node has with other nodes, and modularity (Watts and Strogatz, 1998). Modularity is the

process of finding nodes in the main network that are more tightly connected to each other than to the rest of the graph. This allows the discovery of clusters, or “cliques,” of tightly-connected participants within the conference's Twitter backchannel.

10.4 LIWC Analysis

Before analyzing with LIWC, all tweets had @-mentions, hashtags, and URLs removed. LIWC does not recognize these words when assigning scores to the texts, but will include them in the total word count. All retweets were also excluded, since they contain full copies of the original tweet, which would cause duplicates to appear in the data set.

Prior to analysis, tweets were grouped by the following criteria: (1) all tweets from the conference; (2) grouped by participant; (3) grouped by modularity class, as assigned by Gephi (modularity class is described in the Data Analysis section of this paper, below). All tweets within each grouping were concatenated together before being analyzed in LIWC.

To add context to the assembled tweet materials, the researchers will reference results below to a randomly sampled baseline of English tweets, collected through the Twitter “Firehose” API. A baseline sample of one million random tweets was collected this way within 60 days of the NCTE conference. Additionally, the maximum number of retrievable tweets for 18,046 users in this set were collected.

10.5 Data Analysis

The primary goal in analyzing the Twitter conference data is understanding the linguistic patterns and network structure within the data set in comparison to a baseline sample. The authors performed social network analysis in Gephi (Bastian et al., 2009), and linguistic analysis in LIWC.

10.6 Social Network Analysis (Gephi)

These researchers are interested in the discovery of clusters and the identification of highly connected participants, to investigate the degree to which Gephi and LIWC can identify meaningful clusters of participants. To investigate the structure and patterns of connections between Twitter users, the present researchers analyzed the Twitter data in Gephi.

For the collected tweets, the researchers conceptualize each participant as a node in a network, and each retweet, @-mention, and reply as an edge connecting the poster of the tweet to the participant(s) retweeted, mentioned, or replied to. The pattern of resulting nodes and edges form the network which is the present subject of investigation. The degree of participants, the modularity of the network, and various connectivity measures such as centrality were computed on the resulting network.

10.7 Linguistic Analysis (via LIWC)

To investigate the thematic, affective, and stylistic patterns within the tweets, preliminary data analysis of the groupings described above was performed via LIWC, a computer software research tool designed for the analysis of written and transcribed verbal texts. LIWC analysis is based on the belief that the words people use “provide important clues to their thought processes, emotional states, intentions, and motivations” (Tausczik and Pennebaker, 2009). Additional analysis can be made of a speaker's or author's attentional focus, thinking styles, and idiosyncratic use of language. Although it is important to recall that language use is highly contextual and findings may not generalize to differing groups of people or across contexts, such analysis none the less offer a systematic look into these important areas (Tausczik and Pennebaker, 2009). A variety of researchers have utilized LIWC for Twitter analysis (Boyd, 2017), and the LIWC manual provides results from analyzing a set of tweets, so there is precedent for applying it to the data set.

Analysis of natural language provides important information about how individuals process the environment around them and make sense of their situation. Thinking can vary in complexity and depth, and this is frequently reflected in the words people use to express and to connect their thoughts (Tausczik and Pennebaker, 2009). The researchers were particularly focused on several new analytical features present in the 2015 LIWC release, which allow researchers to examine more complex elements of discourse including the four aspects of authenticity, tone, clout, and analytic approach (Pennebaker et al., 2015b).

10.8 Coding Primary Influencers on Twitter Backchannel

Research topic 2 focused on identifying the primary roles and professional identities of the key influencers in the NCTE 2016 conference Twitter backchannel. To determine the affiliations of the top 50 participants with the highest degree value (as reported by Gephi), participants were manually assigned to 11 different categories by Semingson and Anderson. For a description of the codes used, see Table 1. The top 50 participants were selected as this represents the highest degree participants, who are assumed to be heavily influential, while still being a feasible number to hand-code.

TABLE 1: Numeric Codes for Affiliation of the top 50 tweeters at the National Council of Teachers of English 2016 annual conference. The codes identify the affiliations and identities of the primary tweeters—those with the top degree of connectivity.

Affiliation	Code	Description
Other	0	
Organization	1	National Council of Teachers of English organization itself
Publisher	2	Publishing company
(Full-time) Practitioner	3	Teacher, PK-12th grade affiliated educator
Author-Professional Development books	4	Author of practitioner-friendly book (such as Heinemann)
Author-Children's/Young Adult	5	Primarily a children's/YA author
Part-Time or Full-Time Consultant	6	Private consultant or consultant affiliated with Teachers College
Affiliated with literacy organization (full-time employee)	7	NCTE employee or administrative affiliate
Affiliated (employee) of book publisher and/or education company	8	Book publisher employee or education-focused company
(Full-time) Academic	9	Professor/academic/researcher
Other education-related organization (non-profits)	10	Example: 90-second Newbery or foundation

10.8.1 Coding Process

In the process of coding, Semingson brought knowledge of the professional field of literacy to the determination of the coding categories for professional affiliation. The categories were determined by a preliminary qualitative inspection of the 50 participants with the highest degree, and Semingson's prior knowledge about NCTE's practitioner-focused audience and membership. Participants were assigned to as many categories as was appropriate. Semingson and Anderson met to compare independently categorized these participants, starting with a manual and qualitative investigation of their Twitter profiles, with particular focus on their 160-character “biography” sections (where Twitter users can briefly describe themselves, visible on their main profile page). Nearly all of the Twitter

profiles linked to a primary professional website which contained further information about the participant's professional identity. If a link was provided to an external personal or professional webpage, that page was investigated and information was used to further determine the appropriate manual coding of the participant's affiliation. A general web search using Google was also conducted to discover publicly accessible information on participants, such as news stories, other websites, or press releases.

10.8.2 Reaching Coding Consensus

Following the preliminary manual coding of affiliations, the two researchers met to compare results, reach a consensus on codes, and author qualitative analytical memos describing preliminary themes and trends (Miles et al., 2014). Further online searching following the above procedures helped to clarify any discrepancies between the researchers. Following coding, key themes, trends, and methodological ideas and issues were discussed until intersubjective agreement was achieved.

10.8.3 Coding Affiliations in Gephi

To further investigate the nature of the manually coded affiliations, the authors examined them in Gephi, adding each affiliation (e.g., “Practitioner”) as a node, and an edge to connect each coded participant to their respective affiliations. We then included these nodes in the modularity analysis of the network.

The “Consulting” designation proved the most difficult affiliation to define. Participants with this affiliation had undertaken a wide variety of consulting activities, such as public speaking, outreach, and staff development. Participants also varied from independent consultants to those who were affiliated with larger organizations, such as Teachers College.

11. RESULTS

11.1 Social Network Analysis: General Remarks and Observations

A visual representation of the conference network can be seen in Fig. 1, generated by Gephi. Several global statistics for the network are shown in Table 2. The size of each node directly corresponds to its degree value. The color of each node (most clearly visible in the digital, full-color issue) indicates its modularity class. The layout was generated using the Force Atlas 2 algorithm (Jacomy et al., 2014). For readability, only the nine clusters comprising of at least 4% of the total participant population are shown (the general structure of the network is still accurately represented, and 66% of participants are still visible. The hidden nodes are peripheral to the network, and make it “noisy” and difficult to

interpret, except as very large image sizes). A final graph comprised 7,273 total nodes and 105,812 total edges.



FIG. 1: A visual representation of the Twitter network at the NCTE 2016 Conference. The size of nodes (participants) directly corresponds to their degree (total number of connections to other users). The color, most easily visible in the digital publication, corresponds to the participant's modularity class.

TABLE 2: Global metrics for the NCTE 2017 Twitter network, calculated in Gephi.

Metric	Value
Average Degree	14.549
Network Diameter	11
Total Modularity	0.422
Average Clustering Coefficient	0.229
Average Path Length	3.377
Graph Density	0.002

The majority of tweet connections in the dataset were retweets (42.94%) or @-mentions (53.56%). Only 3.4% of the tweets were replies, as classified by Twitter. We do not differentiate in the present study between the ranges of functions performed by each type of interaction at present. For example, an @-mention can be used to reply to a specific participant or a specific post, but this will be recorded in the data set as an @-mention rather than a reply, unless the participant clicked the “reply” button on a tweet, causing Twitter to explicitly flag the message as a reply.

Several interesting structures are evident in the graph. There are several instances where a group of participants is all connected to a single other participant, but not to each other. These interactions tended to be a mix of retweets and @-mentions, and were primarily directed at the single node (rather than the single participant retweeting, mentioning, or replying to participants in the group). It is not clear what factors in the data set may lead to these structures, and this presents an opportunity for future inquiry.

12. SOCIAL NETWORK ANALYSIS: CLUSTERING

Gephi utilizes the clustering approach introduced by Blondel et al., (2010) to identify *modularity classes*, or clusters of nodes. Those researchers define clusters as groups of nodes (conference attendees in the current data set) that share many connections among themselves, but fewer connections with nodes outside of the group. In the context of social groups such as the NCTE conference, these clusters can be interpreted as communities or “cliques” that form within the main Twitter discourse. This clustering is based purely on the number of connections between two participants; it does not account for the content of any messages that connect, and thus misses some nuances that may be detectable through deeper linguistic analysis.

Clustering was run using Gephi's default settings. Gephi discovered 1,130 total clusters; the majority of these were extremely small, consisting of only a very small number of participants (often just one) who were not connected to the main portion of the network.

The overall modularity value for the network was 0.422 (a value of 1 indicates that every node is connected to every other node; a value of 0 indicates that there are no edges at all). The average clustering coefficient for the participants in the network was 0.229. This is the average value of the clustering coefficient of each node in the network. A value of 0 for a given node indicates that there are no connections between other nodes that are directly connected to the current one, whereas a value of 1 indicates that all such possible connections are made. Together, these values indicate a reasonably well-connected network.

A chart showing the number of participants in the 15 largest clusters, making up 81.7% of the total participants, is shown in Fig. 2; exact counts for all clusters with at least 75 participants can be found in Table 3. The clustering revealed several major groups of interest. In the largest cluster (15.15% of participants), the official NCTE Twitter account was by far the largest node. In the second largest (13.23%), the largest node was related to Heinemann Publishing, a primary publishing house in the market space. Of note is that these two nodes are rather far from each other in the visualization of the network, which indicates that there are relatively few connections between them. The most prominent nodes in all other clusters were individual conference participants. Some of these participants were well-established authors who received an exceptionally large number of tweets from other participants, while some had been speakers at the conference. However, it was not clear for all participants why they were so prominent within their respective networks. What factors contribute to the centrality of participants in each cluster is an interesting question that lends itself to future investigation.

User Counts for 15 Largest Modularity Classes

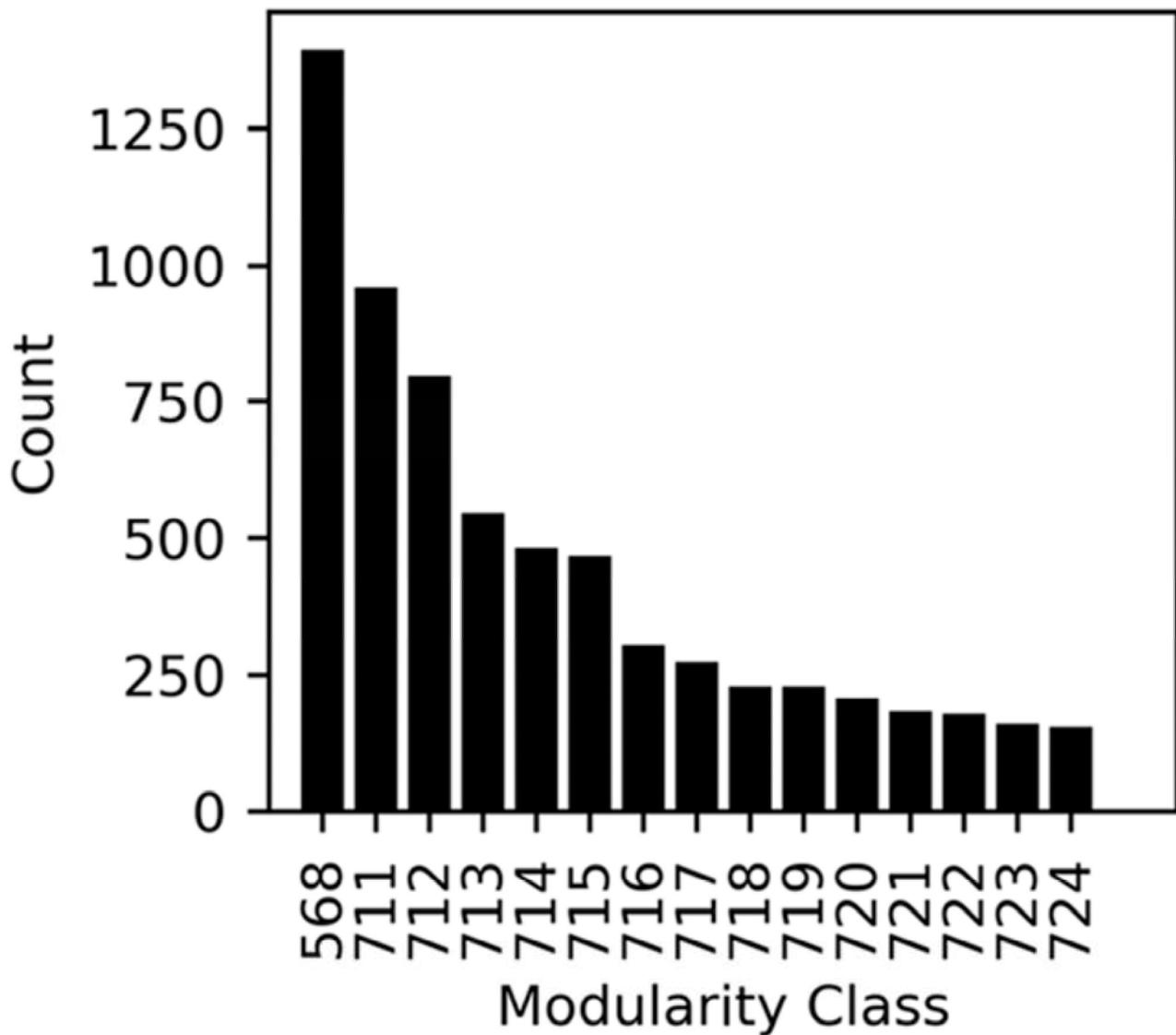


FIG. 2: The number of users in each of the top 15 modularity classes out of a total of 1,130, accounting for 81.7% of participants. The numeric class labels are arbitrarily assigned by Gephi and do not carry any inherent meaning. Omitted classes had very few users, and the count of users per class decreases very slowly among them.

TABLE 3: Counts of users per modularity class. This table shows the cluster numbers and the number of participants per cluster for the 19 clusters with at least 75 members. Note that the cluster labels themselves do not signify anything, and are assigned by Gephi when the algorithm is run.

Cluster	Size	Cluster	Size
4	1395	544	208
9	960	89	183
14	798	7	179
12	547	13	161
19	482	162	155
17	468	469	142
8	304	635	13
63	273	74	117
644	229	1034	75
10	229	All Others	1427

13. LINGUISTIC ANALYSIS: SENTIMENT OF THE TWITTER BACKCHANNEL

In a bag-of-words analysis of the approximately 318,000 total lexical elements in the full NCTE tweet data set, conference language was noted as markedly “analytic” (80.69 on a normalized scale of 0–100), representative of “clout,” a close proxy to expression of expertise (86.78), and expressed in a positive emotional “tone” (95.72). Only on the result scale for “authenticity” did the NCTE lexical data set register in a more neutral zone (31.86). To provide additional context, a comparative, randomly sampled baseline of 1,000,000 English tweets (9,357,084 words total), generated via the Twitter “Firehose” API as described above, resulted in baseline LIWC readings of analytics (71.11 on a normalized scale of 1–100), clout (63.58), tone (63.58), and authenticity (29.23). Per-user distributions for authentic language and clout between the conference data and the baseline is shown in Fig. 3.

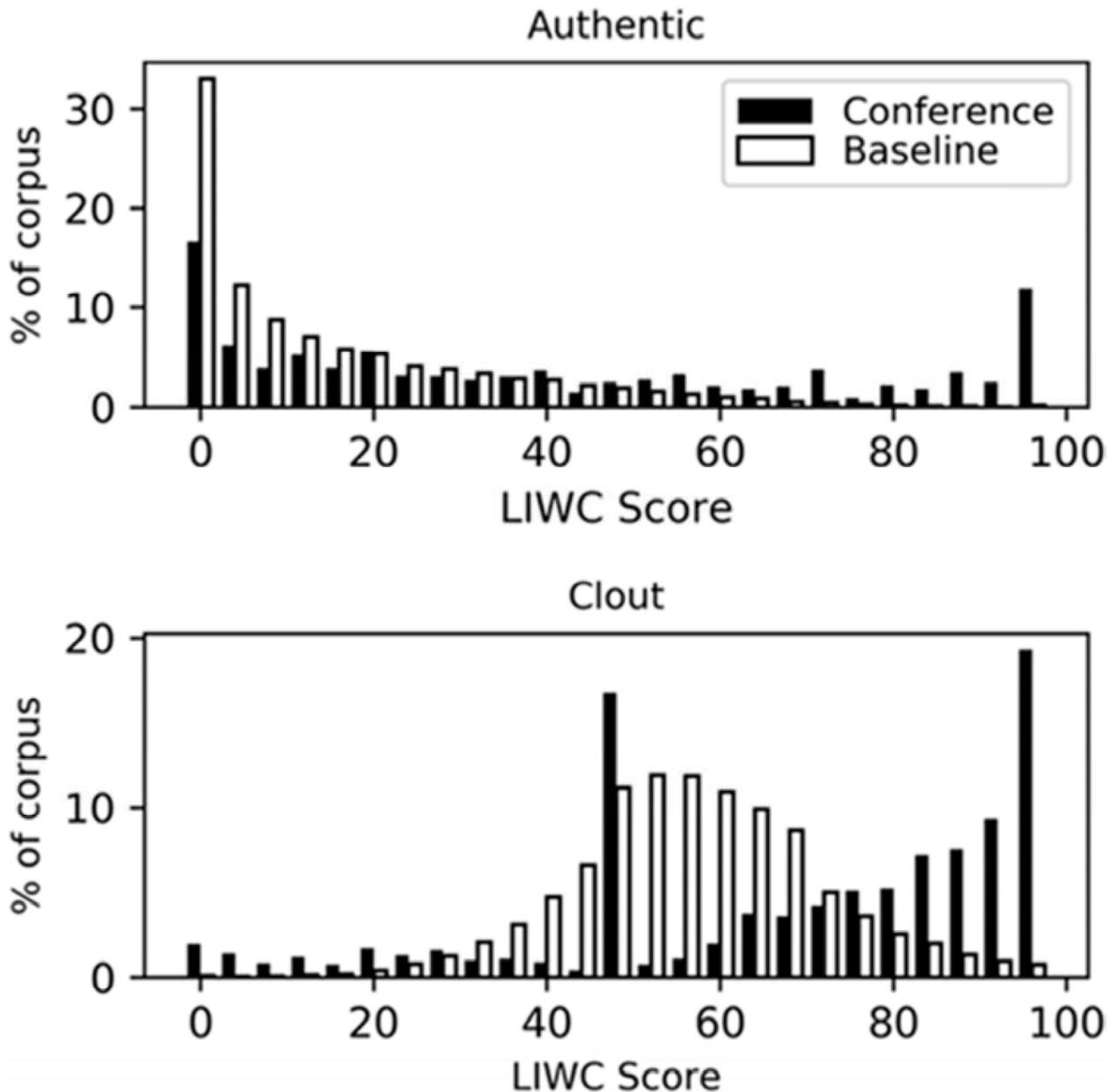


FIG. 3: Authentic language and clout distributions. A comparison of the by-user distributions of authentic language and clout, reported by LIWC, between the conference Tweets and the 18,046-user baseline.

Indeed, the emotionality—positive and negative emotions—expressed within the NCTE Twitter corpus showed a marked leaning toward the positive. Positive emotion words outnumbered negative emotion lexical items at an almost 5:1 ratio. Interestingly, Tausczik and Pennebaker (2009) note that use of emotional words have also been seen as a proxy for degree of immersion or group engagement.

As a final step, LIWC analyses were conducted on the tweet content of the top 50 influencers, as identified by SNA analysis. Interestingly, participants in the “Author-

Children/Young Adult” category outscored all of the other influencers in tone (99 on a normalized scale of 0–100) and authenticity (36.07), while registering far and away the most minimal use of clout or expertise language (77.69).

14. PRIMARY ROLES AND PROFESSIONAL IDENTITIES OF THE KEY INFLUENCERS

The present authors chose to more closely examine the top 50 participants by degree of connectivity to address research topic 2: *What are the primary roles and professional identities of the key influencers in the NCTE 2016 conference Twitter backchannel?* The total number of participants assigned to each code can be found in Table 4. Nearly all of the top 50 participants by degree of connectivity had multiple professional affiliations during the hand-coding stage, which is to be expected.

TABLE 4: The number of participants in the top 50 (by degree) manually assigned to each affiliation. Participants were assigned to multiple affiliations, so the values will not sum to 50.

Affiliation	Code	Number of Participants
Other	0	1
Organization	1	1
Publisher	2	4
(Full-time) Practitioner	3	20
Author-Professional Development books	4	23
Author-Children's/Young Adult	5	5
Part-Time or Full-Time Consultant	6	27
Affiliated with literacy organization (full-time employee)	7	1
Affiliated (employee) of book publisher and/or education company	8	1
(Full-time) Academic	9	7
Other education-related organization (non-profits)	10	6

Of the academics in the top 50, none were solely research focused; all were also assigned to at least one other affiliation. Most of the academics had practitioner connections, such as writing books or consulting in schools. Many of the top 50 participants also were bloggers. Blogs varied in content and frequency of updates. As can be seen in Table 4, over half of these participants were consultants, either part-time or full-time.

Not all of the top 50 participants had large followings on Twitter as measured by their total follower counts during the conference (which ranged from 151 to 156,149, with an average 16,696; ten participants had fewer than 1,000 followers, and thirty had less than 10,000), indicating that being well-connected at the conference does not always entail a large Twitter following in other contexts. Of the practitioner-focused authors, the majority were associated with Heinemann Publishing. This corresponds with the finding that the Heinemann Twitter account had the second highest degree of any in this data set, and appeared to be the center of the second-largest cluster discovered by Gephi.

Performing the modularity (clustering) analysis with the affiliations added as nodes yielded several interesting results. Practitioners (code 3) and publishers (code 2) were in the same modularity class (cluster) as Heinemann publishing, indicating that practitioners had a large amount of interaction with Heinemann and Heinemann-associated participants. Other education-related organizations (code 10) displayed in the same class as NCTE. Publisher employees (code 8) and literacy organization employees (code 7), interestingly, were not in either NCTE's or Heinemann's modularity class, as might be expected. Practitioner-oriented authors (code 4), children-oriented authors (code 5), and consultants (code 6) appeared in the same modularity class as each other, but this class accounted for only 2.34% of the conference participants and did not have a clear common theme among its members.

These findings indicate that many educational practitioners interacted primarily with Heinemann Publishing and with other participants closely connected to Heinemann Publishing (recall that "interaction" here refers to retweeting, @-mentioning, and replying to other users). The fact that practitioner-oriented authors, children-oriented authors, and consultants appear in the same class, separate from Heinemann, NCTE, or other easily-identifiable networks, lends itself to an interesting interpretation. It indicates that these individuals did not have consistent affiliations with the more identifiable networks. This would be in line with the observation that these participants had highly varied professional backgrounds. Not all authors were associated with Heinemann, for example, and a very large number of participants were consultants, spanning a wider range of prior affiliations.

This part of the study has noteworthy limitations. It is not necessarily conclusive due to the small sample sizes of many hand-coded affiliations and of the total proportion of participants coded. However, the results indicate that hand-coding a larger portion of the

participant base could substantially enrich the results from established SNA analysis techniques.

15. DISCUSSION

Discipline-specific organized Twitter chats by professional organizations provide valuable conversation and professional development for theoreticians and practitioners alike in the 21st century (Biddolph and Curwood, 2016). Large-scale, complex conversations are clearly possible in these online settings, even if the motivations of participants and their uses of the medium vary by style and interest. This paper attempted to explore the methodology to analyze the networked learning and Twitter interaction of a group of language arts focused Twitter participants affiliated with the National Council of Teachers of English (NCTE) official conference hashtag. By examining *both* the lexical and network structure of such conversations, researchers can hope to better understand the scalable professional collaboration and conversations that scaffold and develop us as educators. Researchers and practitioners alike have been searching for impactful social media approaches for professional support and development, yet few obvious uses have emerged.

The approach described in this paper indicates what form such an investigation may take: an integration of tools and methods from a range of disciplines and research fields, both quantitative and qualitative, rather than relying primarily on one set of methodologies and techniques and merely supplementing the results with other approaches. The current work has shown that such a method can offer very promising insights to complex, multifaceted data sets such as the NCTE Conference Twitter backchannel by applying tools from SNA to analyze the network aspect of the backchannel, NLP to analyze the content, and qualitative analysis to direct and enrich the results from both. It is unlikely that similar results could be easily obtained by relying solely on SNA or solely on NLP; rather, it is the syncretic combination of both that provides the greatest depth of insight.

16. FUTURE WORK

16.1 Expanding Methodologies and Toolsets

Future scholarly work still remains in fully integrating and applying methodologies from SNA and NLP, as well as other disciplines, to social media-based networked professional development and learning. One such area of systematic inquiry is the incorporation of more sophisticated natural language processing tool sets to better investigate the linguistic dimension of the current data set. Of particular interest would be word and document embedding tools such as Word2Vec (Mikolov et al., 2013) and Doc2Vec (Le and Mikolov, 2014) to measure whether and how much influential participants shape the dialogue in

Twitter conversations at scale, and sentence parsers such as Google's Parsey McParseface (Andor et al., 2016) to analyze the grammatical structure of tweets in this collection.

Approaches that incorporate the *content* of participants' communications, rather than strictly the *structure* of communications, are also an area for significant further work. Tools related to topic modeling provide one promising way to expand the scope of the current work in this regard. For example, Latent Dirichlet allocation (Blei and Ng, 2003), a widely used topic modeling algorithm, allows users to extract topics of discussions from Twitter data that are not directly observable, e.g., classifying a tweet as being about “literacy education,” “publishing,” or “response to conference speakers.” Knowing these topics would provide an interesting layer of analysis, particularly with regards to whether topics of discussion align with LIWC, SNA, or hand-coding results.

16.2 Applications to Broader Data Sets

Additionally, an automated approach to participant classification, comparable to manual classification in the current study, is an area for further work. Automatic classification of many thousands of participants would allow comparison between the manually defined affiliations and automatically detected clusters to be scaled up to much larger data sets. This would allow for a more rigorous and comprehensive analysis of the typical interactions of participants based on their affiliations.

The current study's work should also be applied to a broader range of topics and contexts, including other professional organizations (literacy and non-literacy focused) and purely social contexts. NCTE's monthly #NCTEChat Twitter events would be particularly apt, since they take place in the same institutional context as the current data set and analysis. Unlike the NCTE conference, #NCTEChat is a repeating event, which would further allow study of how patterns observed here evolve over time. A further analysis of the interesting network structures, such as the many-to-one connections the authors observed, may also reveal greater detail about the structure of the network and the nature of participant interactions; this particular avenue of inquiry lends itself naturally to the combination of SNA, linguistic, and qualitative approaches developed in the current paper.

Future work will have implications for both practitioners and scholars, as well as the broader field of big data analytics as it applies to digital connectivity of participants within a professional organization or setting.

17. CONCLUSION

The authors have demonstrated that a SNA, NLP, and manual qualitative coding of the data resulted in differing, but complementary (rather than supplementary) information

about the Twitter network, pursuant to the first guiding question. The present study is restricted to a single Twitter discussion in a professional education conference setting, thus limiting generalizability of the results themselves and leaving much room for further applications and refinement of the integrative methodology developed by the authors. The diverse and complex nature of the resulting discourse however, shown in this analysis, indicates the promise of future research into Twitter-based communities and discourses using an integrative, multidisciplinary methodology like the one described in this paper.

ACKNOWLEDGMENT

The authors wish to express appreciation to Mr. Srećko Joksimović (University of Edinburgh) for consultation in social network analysis tools, methodologies, and literature.

REFERENCES

- Andor, D., Alberti, C., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S., and Collins, M. (2016), Globally Normalized Transition-Based Neural Networks. Retrieved May 5, 2017 from <https://arxiv.org/abs/1603.06042>.
- Bastian, M., Heymann, S., and Jacomy, M. (2009), Gephi: an Open Source Software for Exploring and Manipulating Networks. Proceedings of the Third International ICWSM Conference, pp. 361–362. Retrieved May 5, 2017 from <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009>.
- Biddolph, C. and Curwood, J. (2016), #PD: Examining the Intersection of Twitter and Professional Learning, in M. Knobel and J. Kalman, Eds., *New Literacies and Teacher Learning: Professional Development and the Digital Turn*, New York: Peter Lang Publishing, pp. 195–218.
- Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003), Latent Dirichlet Allocation, *J. Machine Learning Res.*, **3**(Jan), pp. 993–1022.
- Blodgett, S.L., Green, L., and O'Connor, B. (2016), Demographic Dialectal Variation in Social Media: A case study of African-American English. Retrieved May 5, 2017 from <https://arxiv.org/abs/1608.08868>.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008), Fast unfolding of Communities in Large Networks, *J. Stat. Mech.: Theory Exp.*, **2008**(10), P10008.
- Borgatti, S.P., Mehra, A., Brass, D.J., and Labianca, G. (2009), Network Analysis in the Social Sciences, *Science*, **323**(5916), pp. 892–5.
- Boyd, R. (2017), Personal communication.
- Britt, V.G. and Paulus, T. (2016), “Beyond the Four Walls of My Building”: A Case Study of #Edchat as a Community of Practice, *Am. J. Distance Educ.*, **30**(1), pp. 48–59, doi:10.1080/08923647.2016.1119609.

- Bruns, A. (2011), How Long Is a Tweet? Mapping Dynamic Conversation Networks using Gawk and Gephi. *Information, Commun. Soc.*, **15**(9), pp. 1323–51, doi: 10.1080/1369118X.2011.635214.
- Carpenter, J.P. and Krutka, D.G. (2014), How and Why Educators use Twitter: A Survey of the Field, *J. Res. Technol. Educ.*, **46**(4), pp. 414–34, doi:10.1080/15391523.2014.925701.
- Chen, B. (2011, April), Is the Backchannel Enabled? Using Twitter in Academic Conferences. Paper presented at the 2011 Annual Meeting of the American Educational Research Association (AERA), New Orleans, Louisiana. Retrieved May 5, 2017 from https://www.researchgate.net/publication/267856863_Is_the_backchannel_enabled_Using_Twitter_at_academic_conferences.
- Davis, K. (2015), Teachers' Perceptions of Twitter for Professional Development, *Disab. Rehab.*, **37**(17), pp. 1551–8, doi:10.3109/09638288.2015.1052576.
- Ebrahimi, J., Dou, D., and Lowd, D. (2016), Weakly Supervised Tweet Stance Classification by Relational Bootstrapping. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1012–1017. Retrieved May 5, 2017 from <http://aclweb.org/anthology/D/D16/D16-1105.pdf>.
- Enriquez, J.G. (2010), Fluid Centrality: A Social Network Analysis of Social-Technical Relations in Computer-Mediated Communication, *Int. J. Res. Method Educ.*, **33**(1), pp. 55–67, doi:10.1080/17437271003597915.
- Fast, E. and Horvitz, E. (2016), Identifying Dogmatism in Social Media: Signals and Models. Retrieved May 5, 2017 from <https://arxiv.org/abs/1609.00425>.
- Gillen, J. and Merchant, G. (2013), Contact Calls: Twitter as a Dialogic Social and Linguistic Practice, *Language Sci.*, **35**, pp. 47–58, doi:10.1016/j.langsci.2012.04.015.
- Greenwood, S., Perrin, A., and Duggan, M. (2016) Social Media Update, Pew Research Center. Retrieved May 5, 2017 from <http://pewinternet.org/2016/11/11/social-media-update-2016/>.
- Huberman, B.A., Romero, D.M., and Fang, W. (2008), Social Networks that Matter: Twitter under the Microscope (December 5, 2008). Retrieved May 5, 2017. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1313405 or doi:10.2139/ssrn.1313405.
- Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014), ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS One*, **9**(6), e98679.
- Le, Q. and Mikolov, T. (2014), Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 1188–1196. Retrieved May 5, 2017 from <http://proceedings.mlr.press/v32/le14.pdf>.

- Li, J. and Greenhow, C. (2015), Scholars and Social Media: Tweeting in the Conference Backchannel for Professional Learning, *Educ. Media Int.*, **52**(1), pp. 1–14, doi:10.1080/09523987.2015.1005426.
- Megele, C. (2014), Theorizing Twitter Chat. *J. Perspectives Appl. Acad. Practice*, **2**(2), pp. 46–51. Retrieved May 5, 2017 from <http://jpaap.napier.ac.uk/index.php/JPAAP/article/view/106>.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013), Efficient Estimations of Word Representations in Vector Space. Retrieved May 5, 2017 from <https://arxiv.org/abs/1301.3781v3>.
- Miles, M.B., Huberman, A.M. and Saldana, J. (2014), *Qualitative Data Analysis: A Methods Sourcebook*. Thousand Oaks, CA: Sage.
- Nielson, J. (2006), The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities. Retrieved February 18, 2017 from <https://www.nngroup.com/articles/participation-inequality/>.
- Pennebaker, J.W., Booth, R.J., Boyd, R.L., and Francis, M.E. (2015a), *Linguistic Inquiry and Word Count: LIWC 2015*. Austin, TX: Pennebaker Conglomerates Inc. (<https://liwc.wpengine.com/>).
- Pennebaker, J.W., Boyd, R.L., Jordan, K., and Blackburn, K. (2015b), *The Development and Psychometric Properties of LIWC2015*, Austin, TX: University of Texas at Austin.
- Rehm, M. and Notten, A. (2016), Twitter as an Informal Learning Space for Teachers!? The Role of Social Capital in Twitter Conversations among Teachers, *Teaching Teacher Educ.*, 60215–23, doi:10.1016/j.tate.2016.08.015.
- Ryberg, T., Buus, L., and Georgsen, M. (2012), Differences in Understandings of Networked Learning Theory: Connectivity or Collaboration? in L. Dirckinck-Holmfeld, V. Hodgson, and D. McConnell, Eds., *Exploring the Theory, Pedagogy and Practice of Networked Learning*, New York: Springer Science+Business Media, pp. 43–58.
- Ryymän, E., Palonen, T., and Hakkarainen, K. (2008), Networking Relations of using ICT within a Teacher Community, *Comput. Educ.*, **51**(3), pp. 1264–82, doi:10.1016/j.compedu.2007.12.001.
- Scott, J. (2012), *Social Network Analysis*, Sage.
- Schreurs, B. and de Laat, M. (2014), The Network Awareness Tool: A Web 2.0 Tool to Visualize Informal Networked Learning in Organizations, *Comput. Human Behav.*, 37385–94, doi:10.1016/j.chb.2014.04.034.

- Siemens, G. (2005), Connectivism: a Learning Theory for the Digital Age, *Int. J. Instruct. Technol. Distance Learning*, **2**(1). Retrieved May 5, 2017 from http://www.itdl.org/journal/jan_05/article01.htm.
- Skrypnyk, O., Joksimović, S., Kovanović, V., Gašević, D., and Dawson, S. (2015), Roles of Course Facilitators, Learners, and Technology in the Flow of Information of a cMOOC. *Int. Rev. Res. Open Distrib. Learning*, **16**(3), pp. 188–217, doi:10.19173/irrodl.v16i3.2170.
- Stake, R.E. (2005), Qualitative Case Studies, in N.K. Denzin and Y.S. Lincoln, Eds., *The Sage Handbook of Qualitative Research*, 3rd Ed., Thousand Oaks, CA: Sage, pp. 443–66.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., and Thomas, S. (2006), Professional Learning Communities: A Review of the Literature, *J. Educ. Change*, **7**(4), pp. 221–58.
- Tausczik, Y. and Pennebaker, J. (2009), The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods, *J. Language Social Psych.*, **29**(1), pp. 24–54, doi:10.1177/0261927X09351676.
- Veletsianos, G. and Kimmons, R. (2012), Networked Participatory Scholarship: Emergent Techno-Cultural Pressures toward Open and Digital Scholarship in Online Networks, *Comput. Educ.*, **58**(2), pp. 766–74.
- Verma, S., Vieweg, S., Corvey, W.J., Palen, L., Martin, J.H., Palmer, M., Schram, A., and Anderson, K.M. (2011), Natural Language Processing to the Rescue? Extracting “Situational Awareness” Tweets during Mass Emergency, In *ICWSM Proceedings*, 2011.
- Visser, R.D., Evering, L.C., and Barrett, D.E. (2014), #TwitterforTeachers: The Implications of Twitter as a Self-Directed Professional Development Tool for K–12 Teachers, *J. Res. Technol. Educ.*, **46**(4), pp. 396–413, doi:10.1080/15391523.2014.925694.
- Watts, D.J. and Strogatz, S.H. (1998), Collective Dynamics of “Small-World” Networks, *Nature*, **393**, pp. 440–2, doi:10.1038/30918.