Social Paleontology on Twitter: A Case Study of Topic Archetypes, Network Composition, and Structure

Lisa Lundgren1, Kent J. Crippen2, Jennifer E. Bauer3, and Richard T. Bex II2

Abstract
Social paleontology is a burgeoning field of research that seeks to understand the natural world through the collection, preparation, curation, and study of fossils via online communities. Such a community represents an ideal case for examining scientific practice as the expression of conversation topics in relation to the people who participate. Using Communities of Practice as a theoretical framework, we consider interactions within an egocentric Twitter network over a 397-day period to identify topic archetypes within the community, examine how such topic archetypes act as expressions of behavior that are indicative of community processes, and provide empirical evidence for detecting and indicating the health of an online community. Data were collected continuously and analyzed with a combination of topic modeling and social network analysis. Four unique archetypes were characterized based on the level of activity and longevity of interest. Participants for each were diverse, but not different. Structural differences in each network were noted with high levels of inter-group information flow within certain archetypes. Archetypes were interpreted using the life cycle states for Communities of Practice; sustained conversations and piques of interest indicate healthy online communities. These findings can inform efforts to design, implement, and research online, scientific communities.

Keywords
social network analysis, diversity indices, topic modeling, InfoMaps, Netlytic, NodeXL

Introduction
Approximately 72% of Americans use social networking sites such as Facebook, Twitter, and Instagram (Pew Research Center, 2019a), where diverse users can create original content and reply to one another, information can be aggregated via hashtags, or users can point to additional information through the use of URLs. Sites like Twitter have been shown to encourage diverse communities who are interested in science, regardless of their level of expertise, to communicate about topics of interest within disciplines such as paleontology or ornithology (Bex et al., 2019; Lundgren et al., 2022; Côté & Darling, 2018; Liberatore et al., 2018). In addition to the richness of content that is offered on the platform, Twitter-specific research is fruitful for studying online exchanges around different topics because most of the interactions involve publicly accessible data and users whose profiles are public (Ahmed & Lugovic, 2018).

This study builds on previous studies investigating community response to message design elements and features across Facebook and Twitter based on expressed identity within a paleontological community (Bex et al., 2019; Lundgren et al., 2022). We take the approach that this is a model community as paleontology is a charismatic field of scientific study that draws members with varied interests and who make scientific contributions through the collection, preparation, curation, and study of fossils (Catalani, 2014). Thus, similar to ornithology and other biological fields of study, the online paleontological community represents a wide range of interests and contributions (Tancoigne, 2019). Here, we expand the aims and scope of our original explorations, taking an inductive data science approach to understand...
this community. We examined data patterns that emerged over time (i.e., 1 year), included relatively robust numbers of social interactions (i.e., ~8,000), and used a form of natural language processing to orient our investigation. We use the Communities of Practice (CoPs) (Wenger, 1998, 2000) theoretical framework to understand the results and to make and assess predictions. Machine learning techniques, such as topic modeling, afford identification of underlying structures and patterns within networks (Nikolenko et al., 2017); however these structures and patterns need to be contextualized to understand and make interpretations about the network.

The context of this work was an egocentric social network on Twitter. We take the stance that online, social communities such as Twitter function as ecosystems that include organisms (i.e., individuals, organizations, and other entities) that interact with one another and their environment (i.e., the platform of Twitter). Therefore, there is merit in characterizing online ecosystems and applying traditional ecological concepts like diversity indices (Shannon & Weaver, 1962; Simpson, 1949) to online communities, which, to our knowledge, has yet to be done. Our purpose is three-fold: to identify topic archetypes within the communication of an interest-based science community, to examine how such topic archetypes act as expressions of behavior that are indicative of community processes, and to provide empirical evidence for detecting and indicating the health of an online community. Next, we delineate our theoretical perspective, outline relevant literature, and describe the research methodology that guided this study.

Theoretical Framework

The construct of CoP emerged from work done by Lave and Wenger (1991) on legitimate peripheral participation in which learners grow more involved in a community through participating more fully. Within the contexts Lave and Wenger examined, participation was contextually dependent on the learning environment, which was a radical shift that decoupled learning from formal school environments. As a theoretical framework, CoP expands upon and emphasizes the ways that people learn collaboratively (Wenger, 2000). The theory emerged from studies of people who were engaged in work-based learning, particularly focusing on how people were learning collaboratively as they contributed to their own personal knowledge and to the knowledge development of a company (Wenger, 2000). Multiple interpretations of the CoP framework exist (Kimble et al., 2008a, 2008b; Wenger et al., 2011; Wenger-Trayner et al., 2015); our work emerges from the branch of CoP theory that emphasizes three interrelated elements: community, domain, and practice (Wenger, 1998, 2000) We explore how these elements interact within online spaces (Gunawardena et al., 2009; Lundgren et al., 2022).

The three elements of CoPs allow for particular behaviors to be identified and defined as benchmarks that can then be used analytically to evaluate and understand the unique properties and processes inherent to CoPs. Within this study, we emphasize the interconnected nature of the elements (Wenger, 2000; Wenger et al., 2002): the community is the individuals, groups, and organizations who contributed to Twitter activity; the domain is the field of social paleontology (Crippen et al., 2016); and varied social paleontological topics comprise the practices.

The domain of a CoP encompasses the complex, long-standing, and shared interests of the community at hand. Limited studies have specifically addressed the concept of domain, those that do explicate that “developing new knowledge, exchanging relevant information, and/or personal growth” are of import (Britt et al., 2020, p. 2) and that focusing on domains can lead to effective domain-specific interventions (Watkins et al., 2018). Such development occurs through the enactment of practice (Handley et al., 2006), in which members engage fully in “a task, job, or profession” (Brown & Duguid, 2001, p. 203). Practice is qualified as the development of shared elements, both explicit and tacit, with which participation and contribution are identified, including: stories, tools, language, documents, shared worldviews, and ways of addressing problems (Wenger et al., 2002). Practice is a well-scrutinized aspect of the CoP theoretical framework; however, most descriptions of practice fall short as it is ill-defined or given a cursory examination (Smith et al., 2017).

The element of community can be described as those who interact, learn, create relationships, and develop a sense of belonging and commitment to a domain (Wenger et al., 2002). Studies focused on community building in higher education emphasize that community relationships are key to CoPs devoted to teaching and learning at a university level, yet no empirical measurements support such claims (Bondy et al., 2017; Carroll, 2005; De Cindio, 2012; Nistor et al., 2015). Qualitatively describing the ways people commune as well as their dispositions is valuable but lacking robust analytical frames and empirical descriptions, these narratives lose credence.

CoPs are recognized as having a life cycle of development, change, and alteration. Of particular interest is the conceptualization from Wenger and colleagues (2002) who originally proposed five stages, which have been subsequently documented empirically (Knaus & Callcott, 2017; Marques et al., 2016; Pohjola & Puusa, 2016). These stages further explicate CoP theory while also providing evaluation benchmarks that can be used by designers and project staff for decision-making. In developmental order, the stages include: potential, in which shared interest or activity among a core, loosely affiliated group causes them to identify “common knowledge needs” (Wenger et al., 2002, p. 71); coalescing, the phase of most rapid growth in the size of the group, where members “establish the value of sharing knowledge of their domain” (Wenger et al., 2002, p. 82); maturing, a period of less rapid growth, in which the community develops a
comprehensive body of knowledge; stewardship, a phase of change in which the community works to maintain their “intellectual focus” (Wenger et al., 2002, p. 104); and transformation, the final stage of alteration, in which a community ends, becomes something else, or is institutionalized. We make use of these five stages of development to orient our interpretation of the community processes that we predicted as likely to have been expressed as topic archetypes within the Twitter network.

Review of Relevant Literature

Borgatti and Ofe (2010) and de Laat and colleagues (2007) have shown that Social Network Analysis (SNA) is well-suited for investigating online virtual spaces and communities since it allows for relationships to be measured across time and space. Digital niches that have been explored with SNA include blogs, forums, wikis, eLearning (Saq et al., 2018), emails, and social networking sites, which provide researchers easily accessible, digitally documented, and recorded social interaction (Cela et al., 2016; Himelboim et al., 2017). To date, most SNA research has focused on classifying users based on their centrality and the ties between them to understand how information flows (Himelboim et al., 2017). The focus on SNA has been on topics that are not educative in nature; for instance, ties and information flow has been studied in regards to political topics (Wojcieszak & Mutz, 2009), within branding and marketing literature (Habibi et al., 2014) and understanding the connections between people with similar medical ailments (Gabarron et al., 2019). While this research has helped to build our understanding of online social communication, there has been limited research that employs SNA methods to examine how people, groups, and entities interact and learn within science communities (Bex et al., 2019). Researchers that have focused on this employ bibliometric analysis (e.g., Raban & Gordon, 2020) which excludes people who are interested in science but are not a part of the formal academic community. In addition, the literature on the use of online community spaces tends to focus on snapshots in time rather than on extended periods of time (Kimmons & Veletsianos, 2016).

Research pertaining to the use of Twitter spans nearly 15 years and has been approached from multiple perspectives (Gruzd et al., 2011). Much Twitter-specific research describes the use of this social media platform for interest-based activities such as professional development in education and natural sciences (e.g., Bex et al., 2019; Xie & Luo, 2018); such work is often based on snapshots in time that describe users’ interactions with hashtags (e.g., Bert et al., 2016; Bombaci et al., 2016; Britt & Paulus, 2016; Lundgren et al., 2022; Smith et al., 2009). Studies that are snapshots of Twitter users’ activity are useful, but contextualizing these snapshots requires a longer term perspective. Even when longer term quantitative measures are employed to provide meaning within a Twitter network, these studies do not account for the potential diversity of users involved (Greenhalgh et al., 2020).

We take the perspective that an egocentric network on Twitter acts as a community that can foster interest-based conversations among diverse participants which, in turn, can be analyzed to determine the CoP life cycle of development, change, and alteration. While social media niches afford interaction in different ways through their design and use of conventions, we take the approach that individual messages (i.e., tweets) serve as the origin of conversation and as starting points for constructive discourse (Lundgren et al., 2022; Michaels et al., 2008). On Twitter, individuals, groups, and organizations contribute to a rich social world via such conversation. Thus, a conversation on Twitter can be defined as a collection of messages that has the capacity for generating subsequent interest-based interaction. While human coders can tackle the task of analyzing Twitter conversations, when large amounts of data are collected, researchers have turned to machine learning techniques such as topic modeling to aid with analysis. Topic modeling is a statistical technique by which latent patterns in large datasets of texts are identified (Hong & Davison, 2010; Vayansky & Kumar, 2020). Such analysis techniques provide ways to make sense of patterns that may be difficult for human coders to identify as well as allow for researchers to interpret large datasets efficiently (Quinn et al., 2010).

An ecosystem view of community members interacting with each other, groups, and other entities is useful when considering the diversity of communities who engage in conversation. Ecologists query ecosystems, specifically, the interconnected nature of organisms and their associated environments to understand the past, present, and future natural world. Other fields, including education, have appropriated the ecosystem concept (Corin et al., 2015; Hecht & Crowley, 2020) to account for the interconnected nature between people, places, and concepts relevant to their particular field of study. A central concept to the study of ecosystems in the traditional sense is understanding the diversity and abundance of species, which allows for ecologists to make inferences about the past, present, and future health of an environment. SNA draws on these concepts and can be applied to understand how the work of scientists with expertise in ecology and evolution is correlated, how it has developed over time, and how authorship networks form (Borgatti & Ofe, 2010; Borrett et al., 2014; Wasserman & Faust, 1994).

Methodology

This study expands on research into the classification of Twitter topic networks as exemplified by Himelboim et al. (2017). Specifically, Himelboim et al. established a methodology for conceptualizing and classifying Twitter networks
based on structures that act as indicators of information flow (i.e., density, modularity, centralization, and isolates); essentially, the flow of information drove network structures. We parallel this methodology but focus on different indicators that relate to the CoP theoretical framework. Establishing such indicators can account for the nature and health of an online, scientific community which in turn drives practice-based archetypes. This single case investigation of an annual cycle (397 days; July 2017–August 2018) of Twitter activity centered on the FOSSIL project (@projectfossil), an effort focused on building connections and shared practice among a diverse community of paleontologists (Crippen et al., 2016). The case was bounded by our intent to understand how people in the network identify with paleontology, the social structures supporting their communication, the resulting patterns of conversation, and how such people, structures, and patterns serve as indicators of the health of an online community. We calculated the established growth trajectory (average 1.95 new followers per day) and pattern of activity (average 6.98 tweets per day) during the annual cycle of activity and determined that such a period of time would provide a sufficient diversity of people, interactions, and topics for answering our research questions. The detailed and consistent social media messaging practice used by @projectfossil, which involved a stated plan and included original practice-based messages that were distributed with a goal of building community to an established group of diverse followers (Lundgren et al., 2022), made it a unique case study as a model system. The limited research involving conversation structures in social network-based communities dictated the context-dependent method and combination of data analysis methods. We sought to address the following research questions: (1) What topic archetypes exist in this social network? (2) Which participants are involved with different topic archetypes? (3) What relationships exist among the topic archetypes, the composition of participants, and community network structures?

Data Sources and Preparation

Using Netlytic, a browser-based text and SNA service (https://netlytic.org/), we scheduled a sampling of the Twitter public search API every 15 min from July 2017 to August 2018. This representative sample (Rafail, 2017), which used the @projectFOSSIL account as a proxy, included 7,753 messages that originated from the account or included it as a mention. Records included the text of each message, the author’s account, and for re-tweets, the account name who passed the message along, as well as any additional accounts who were mentioned. Attributes such as the biography and number of followers were added to all named accounts. In addition, we prepared the data by classifying each account based on a content analysis (Krippendorff, 2012) of each accounts’ biography using the Paleontological Identity Taxonomy (PIT)—a hierarchical taxonomy based on self-identity with the domain

![Figure 1. Maximum likelihood estimation of number of topics.](image-url)
included similar naming conventions remained separate for analysis. Following the descriptive naming of topics, we applied the paleontological practice-based post type (P3T) framework, which consisted of five unique types of content: Information, News, Opportunity, Research, and Off-Topic (Lundgren & Crippen, 2017) to further distill the nature of each topic.

Each day in the study period was recognized as having the potential for unique tweet activity related to each topic, which allowed for the construction of topic-over-time graphs (Figure 2). Initial review of these graphs identified the volume and regularity of tweets as two distinctive features, which when combined could be used to describe the collection. These features were used to construct two unique metrics. First, the total number of tweets affiliated with a topic was operationalized as the Magnitude of Activity or Magnitude which for these topics varied from 0 to 173 tweets. A base-10 logarithmic transformation (i.e., logMagnitude) was used to better account for the range of values and to support transferability to other studies at larger scales. Topics were binned into categories based on their location on the graph (Figure 2). Next, an Occurrence was defined as a day in the study period that had tweet activity. Occurrences are also synonymous with individual peaks in the topic-over-time graphs. Percent Occurrence was calculated as the total Occurrences divided by the duration of the study. For example, topic activity on the dates of 21 October 2017 (15 tweets), 22 October 2017 (21 tweets), 24 October 2017 (18 tweets) and again on 28 October 2017 (12 tweets) would result in four Occurrences (1% Occurrence) with a corresponding logMagnitude value of 1.75. Percent occurrences were binned into categories based on their location. We interpret the variable of percent Occurrence as a measure of the longevity or duration of interest (high values indicate greater longevity), whereas the Magnitude was interpreted as the strength of interest.

Topic archetypes were delineated based on a visual content analysis of the topic-over-time graphs for each type (Rose, 2013) as well as examining a graph of logMagnitude versus percent Occurrence where each axis was bifurcated to create quadrants and the potential for four archetypes. Potential archetypes thus represented as a combined expression of logMagnitude and percent Occurrence. The potential archetypes consisted of high Magnitude–high percent Occurrence (HH), high Magnitude–low percent Occurrence (HL), low Magnitude–high percent Occurrence (LH), and low Magnitude–low percent Occurrence (LL).

Community ecology diversity indices were utilized to consider the diversity of participants by PIT category in each identified topic. These metrics have been long employed to assess species diversity in a given location, consider topics of biodiversity scales, identify geographic biodiversity hotspots, and consider how biodiversity is accumulated through space and time (Patzkowsky & Holland, 2007; Stigall et al., 2017). These same metrics were used to consider the diversity of individuals contributing to topical Twitter conversations and the diversity of individuals in defined clusters. Shannon (Shannon & Weaver, 1962) and Simpson (Simpson, 1949) diversity indices were calculated in the R package vegan (Oksanen et al., 2019). Indices were plotted against one another to compare results as the Simpson index preferentially weights abundance of individuals and Shannon works to correct for this bias. Assessment of indices visualizations indicated that both tracked one another closely, thus, in the results, we report our findings solely using the Simpson diversity index (SDI). Diversity of topics was analyzed in the context of Magnitude and percent Occurrence as described above. The SDI ranges from zero to one, with one being highest diversity (Simpson, 1949; Table 1). Topics were subsequently binned via statistically significant ($p < .05$) natural gaps to indicate low (0.00–0.50), medium (0.51–0.63), or high diversity (0.64–1.00).

The network was visualized and further characterized using NodeXL (Hansen et al., 2011), where participants were nodes in the network with tweets, re-tweets, and mentions serving as the links among the nodes. We chose to use NodeXL for its powerful visualization and customization capabilities, whereas Netlytic was only used to collect tweets continuously for the study period. Groups within each archetype network were identified with the Clauset-Newman-Moore (Clauset et al.,

---

**Figure 2.** Two example topic-over-time graphs (logMagnitude vs. Date); (a) Paleoart (1.5% Occurrence), and (b) Fossils from Various States (4.5% Occurrence).
Finally, we used a community detection algorithm called InfoMap to visualize the directional flow of information among the groups for each topic archetype (Edler et al., 2016). The networks for each topic archetype were exported to Pajek files and loaded into InfoMap to calculate cod-length of the cluster data as well as to visualize the flow of

Table 1. List of Topics, Descriptions, Number of Associated Tweets, Paleontology Practice-Based Post Types, Magnitude, Percent Occurrence, and Archetype.

| Topic Number | Topic Description                                      | Tweets | Paleontology Practice-Based Post Type (P3T) | Log Magnitude | Percent Occurrence | Diversity Index | Archetype |
|--------------|--------------------------------------------------------|--------|--------------------------------------------|---------------|--------------------|----------------|-----------|
| 1            | Inclusion of Amateur Paleontologists                   | 91     | Opportunity                                | 1.96          | 12.59              | 0.56           | HH        |
| 2            | Citizen Science                                        | 144    | Opportunity                                | 2.16          | 23.68              | 0.39           | HH        |
| 3            | Exhibits at NHM London                                 | 35     | Opportunity                                | 1.54          | 5.00               | 0.45           | LL        |
| 4            | Events at Florida Museum                               | 54     | Opportunity                                | 1.73          | 5.54               | 0.67           | LH        |
| 5            | Fossils from Various States                           | 40     | Information                                | 1.60          | 4.53               | 0.62           | LL        |
| 6            | General Discussion of Paleontology                    | 69     | Information                                | 1.84          | 10.58              | 0.56           | HH        |
| 7            | Project Webinars                                       | 123    | Opportunity                                | 2.09          | 12.09              | 0.47           | HH        |
| 8            | Paleontology Education & News                         | 61     | News                                      | 1.79          | 9.57               | 0.26           | LH        |
| 9            | Undefined                                              | 63     | Information                                | 1.80          | 8.56               | 0.55           | HH        |
| 10           | Compliment Specimens & Prep Help                      | 153    | Information                                | 2.18          | 3.27               | 0.46           | HL        |
| 11           | Project Outreach Events                                | 61     | Opportunity                                | 1.79          | 4.28               | 0.53           | LL        |
| 12           | Women in Paleontology                                 | 244    | Information                                | 2.39          | 3.78               | 0.06           | HL        |
| 13           | In-Person Events                                       | 52     | Opportunity                                | 1.72          | 7.81               | 0.53           | LH        |
| 14           | Paleontology Resources                                | 102    | Information                                | 2.01          | 9.82               | 0.65           | HH        |
| 15           | News Stories About Paleontology                       | 40     | News                                      | 1.60          | 8.56               | 0.40           | LH        |
| 16           | Lists about Paleontology Topics                       | 83     | Information                                | 1.92          | 4.79               | 0.66           | HL        |
| 17           | Microfossils                                           | 79     | Research                                  | 1.90          | 1.51               | 0.60           | HL        |
| 18           | Historical Paleontology                               | 44     | Information                                | 1.64          | 4.53               | 0.66           | LL        |
| 19           | Project Newsletters and Conferences                   | 63     | News                                      | 1.80          | 5.79               | 0.52           | HH        |
| 20           | Leads to Resources                                    | 48     | Information                                | 1.68          | 5.29               | 0.60           | LH        |
| 21           | Multimodal Sharing                                    | 67     | Information                                | 1.83          | 11.84              | 0.31           | HH        |
| 22           | 3D Scanning and Printing                              | 122    | Information                                | 2.09          | 3.27               | 0.66           | HL        |
| 23           | Individual Network Discussion                         | 89     | Information                                | 1.95          | 3.02               | 0.51           | HL        |
| 24           | Resources for Ind. Fossil Collectors                  | 4      | Information                                | 0.60          | 0.76               | 0.63           | LL        |
| 25           | Paleoart                                               | 69     | Information                                | 1.84          | 1.51               | 0.06           | HL        |
| 26           | News About Paleontology                               | 33     | News                                      | 1.52          | 7.30               | 0.22           | LH        |
| 27           | Time Scavengers                                        | 56     | Information                                | 1.75          | 5.54               | 0.60           | LH        |
| 28           | Paleontology in K12                                   | 72     | Information                                | 1.86          | 5.04               | 0.23           | HL        |
| 29           | Reposting of Recent Research Papers                   | 59     | Research                                  | 1.77          | 1.26               | 0.52           | LL        |
| 30           | Natural History, Forams & Ancient Things              | 43     | Information                                | 1.63          | 7.81               | 0.35           | LH        |
| 31           | Project In-person Events                              | 69     | Opportunity                                | 1.84          | 6.30               | 0.66           | HH        |
| 32           | Reports or Discoveries                                | 52     | Research                                  | 1.72          | 10.08              | 0.18           | LH        |
| 33           | Fieldwork                                             | 24     | Research                                  | 1.38          | 1.76               | 0.66           | LL        |
| 34           | Technology and Fossils                                | 38     | Information                                | 1.58          | 4.03               | 0.54           | LL        |
| 35           | 3D Printing, Community and Databases                 | 79     | Information                                | 1.90          | 4.28               | 0.62           | HL        |
| 36           | Ways to Share Paleontology                           | 100    | Information                                | 2.00          | 18.39              | 0.42           | LH        |
| 37           | Opportunities                                         | 65     | Opportunity                                | 1.81          | 7.81               | 0.57           | LH        |
| 38           | Sabertooth Cats                                       | 51     | Research                                  | 1.71          | 0.50               | 0.29           | LL        |

HH = high Magnitude–high percent Occurrence; LL = low Magnitude–low percent Occurrence; LH = low Magnitude–high percent Occurrence; HL = high Magnitude–low percent Occurrence.

2004) clustering algorithm. Graph metrics included number of vertices (i.e., connections), number of edges (i.e., nodes or entities), density, and geodesic distance. Graph density indicates how interconnected vertices in a network are and is measured on a scale of 0–1. Geodesic distance indicates the length of the shortest path between two people within a network.
information. InfoMap involves a code-based equation or compression algorithm that is applied to a directional, weighted network to identify probability random walks across the regularities that were defined as groups (Rosvall & Bergstrom, 2008).

Findings

Topic modeling resulted in 38 distinct topics, which were subsequently described and characterized (Table 1). The magnitude among the topics was relatively low and positively skewed, ranging from 4 to 244 with a median of 63 tweets. The saturation of tweets during the time frame was relatively low and positively skewed with the percent occurrence ranging from 0.5 to 23.7 with a median of 5.4%. Diversity was negatively skewed, ranging from 0.06 to 0.67, with a median of 0.53. More than half of the topics were coded as Information (52.6%), followed by Opportunity (23.7%), Research (13.2%), News (7.9%), and zero topics coded as Off-Topic.

Topic Archetypes

Four archetypes were identified from the graph of Magnitude versus percent Occurrence. This graph was then overlaid with the additional variables to further characterize the archetypes (Figure 3).

The number of topics within each archetype was consistent with each archetype containing between seven and nine topics. The topic-over-time graphs for each archetype were distinctive (Figure 4).

The LL archetype is characterized as consisting of topics that did not merit sustained interest or interaction. Many of the topics clustered in the LL quadrant were represented in the P3T framework as Research (Figure 3). Research topics included tweets in which members of the social world reposted or discussed Reposting of Recent Research Paper, Fieldwork, and about particular groups of animals (i.e., Sabertooth Cats). Within LL topics, especially those that were Research-specific, members of the social world with specific knowledge provided additional information or a reaction; however, those without the specific knowledge did not participate. Viewed over time, LL-affiliated topics followed a pattern of having few, if any, instances of people tweeting about the topic for any appreciable amount of time; these topics generated limited interest over short periods of time (Figure 4). For example, topics such as Fieldwork or Project Outreach Events generated a singular spike of activity on one day with minimal subsequent activity. These topics, while certainly engaging over a very short duration of

![Figure 3. Topics by quadrant illustrating the archetypes with topic descriptions from the P3T framework, represented by different shapes: Information (circles), News (squares), Opportunity (triangles), and Research (diamonds). Topics are colorized by archetype: LL (bright orange), LH (yellow), HL (light orange), and HH (dark orange/brown), those examined in depth include their text description.](image-url)
time, were likely to thwart community development as they did not entice members of the social world to engage.

The LH archetype is characterized as consisting of topics with frequent tweet activity, but of a low volume, suggesting a low-level conversation. There was a cluster of topics within the LH quadrant that were coded into the P3T Type of News, which included media outlet stories about paleontology (Figure 3). When viewed over time, these LH-affiliated topics exhibited evidence of many instances of members engaging with the topic; however, community member diversity was low. Topics rarely exhibited days of high magnitude during the study period, never showing more than 14 unique instances of activity (Figure 4). We interpret these topics to function in the same way that idle chatter does within an office setting: sustaining some level of basic interest over time creating a sense of community cohesion, but

**Figure 4.** View of the four quadrants from the Magnitude versus Percent Occurrence graph with exemplar topic-over-time graphs representing the archetypes. The internal y-axis indicates the logMagnitude associated with each topic per day. Note that topics of the HL archetype (light orange) exhibit extreme, singular spikes of activity, whereas topics of the LH archetype (yellow) follow a pattern of more sustained activity.
not fomenting new ideas or exhibiting patterns consistent with knowledge generation (Probst & Borzillo, 2008).

The HL archetype is characterized as consisting of topics where people were highly interested, but their interaction was not sustained over time. Topics within the HL quadrant were mostly coded as Information—general resources for paleontology, reports of recent project activity, personal connections to paleontology, or links to blogs (Figure 3). HL topics that were Information-based included the topics of Women in Paleontology, an Individual Network Discussion, and Paleoart. Examining and visualizing HL-affiliated topics over time revealed activity that was nearly identical in nature: activity that peaked with a singular large spike that was brief in nature, usually lasting 1 day (Figure 4). For example, the topic of Microfossils generated 68 tweets and replies on one day in April 2018, but not before or after this date. This topic archetype is an important component of Twitter’s social world as it indicates areas of interest that might encourage participation. However, the ephemeral nature indicative of this archetype reveals that conversations are not sustainable, which in turn, might mean that new participants might not have enough exposure to continue to participate and contribute.

The HH archetype is characterized as consisting of topics that drew both high and sustained activity over a period of time. These topics most often fell into the P3T Type of Opportunity, which are defined as ways for members of the social world to participate in or contribute to paleontology (Figure 3). We interpret these topics, which included Inclusion of Amateur Paleontologists, Citizen Science, and Project Webinars, to be highly important, as they were capable of involving members of the social world in sustained interactions. We interpret this to mean that HH topics need to include ways for people from diverse backgrounds to express their interest in paleontology as well as contribute to or participate in it.

When viewed over time, HH-affiliated topics displayed multiple instances of tweets and replies over consecutive days with occasional spikes of activity. An example of this is the topic Paleontology Resources. Throughout the study period, this topic generated activity consistently, which was visualized as small incremental activity interspersed with four distinct spikes of activity (Figure 4). This suggests that the topic was of sufficient interest to the community that it was capable of sustaining a general level of conversation, but is also such that it could produce instances of more intense interest that resulted in peaks of activity. Such a pattern can be viewed as productive for a community because the topic produces a need for conversation, maintaining involvement over time, but it also provides instances of invigoration.

Diversity Within Archetypes

Our second research question concerned participant involvement within the different topic archetypes, which was answered by applying the PIT to participants and examining which participants were interacting within each archetype and their included topics. The most diverse topics in the network were Events at the Florida Museum (0.67; LH archetype), Project In-Person Events (0.66; HH archetype), and Fieldwork (0.66; LL archetype). The least diverse topics were Reports or Discoveries (0.18; LH archetype), Women in Paleontology (0.06; HL archetype) and Paleoart (0.06; HH archetype). When subjected to a one-way analysis of variance (ANOVA), no significant difference in diversity among the archetypes were found, $F(3, 34)=1.604$ ($p=.21$), $r^2=.12$; however, we report our findings concerning raw numbers, percentages (Figure 5), and the range of diversity indices within archetypes (Figure 6) to highlight the potential of our method.

Within the LL archetype, the breakdown of participants included 178 Scientists (41%), 156 Education and Outreach (35.9%), and 100 Public participants (23%). The SDI average was 0.55, indicating medium diversity, but also the most diverse of the four archetypes (Figure 5). The two topics with the lowest diversity in the cluster were Sabertooth Cats (0.29) and Exhibits at the NHM London (0.45). The two topics with the highest diversity were Historical Paleontology (0.66) and Fieldwork (0.66) (Figure 6).

Within the LH archetype, the breakdown of participants included 62 Scientists (15.1%), 275 Education and Outreach (70.2%), and 54 Public participants (13.8%). The LH archetype had the highest percentage of Education and Outreach participants. The SDI average was 0.40, indicating low diversity (Figure 5). The two topics in the cluster with the lowest diversity were Reports or Discoveries (0.18) and News About Paleontology (0.22). The two topics with the highest diversity were Time Scavengers (0.60) and Events at the Florida Museum (0.66) (Figure 6).

Within the HL archetype, the breakdown of participants included 470 Scientists (47.4%), 199 Education and Outreach (20.1%), and 321 Public participants (32.4%). The HL archetype had the highest percentage of Scientists and Public participants. The SDI average was 0.43, indicating low diversity; however, individual topic diversity ranged from 0.06 to 0.66, and HL was the only archetype containing topics with diversities below 0.1 (Figure 5). The two topics with the lowest diversity in the cluster were Paleoart (0.06) and Women in Paleontology (0.06). The two topics with the highest diversity were Lists about Paleontology Topics (0.66) and 3D Scanning and Printing (0.66) (Figure 6).

Within the HH archetype, the breakdown of participants included 932 Scientists (24.8%), 1,220 Education and Outreach (60.1%), and 617 Public participants (14.9%). The SDI average was 0.52, indicating medium diversity (Figure 5). The two topics with the lowest diversity were Project Webinars (0.47) and Project Newsletters and Conferences (0.52). The two topics with the highest diversity were Paleontology Resources (0.65) and Project In-Person Events (0.66) (Figure 6).
Figure 5. Simpson’s diversity indices (SDI) and percentages of participant types per topic archetype. No significant differences in diversity were found between archetypes.

Figure 6. Topics colored by archetype, sized by diversity category, and shapes correspond to associated P3T topic descriptions. LL (bright orange), LH (yellow), HL (light orange), and HH (rust).
Community Network Structures

A comparison of network characteristics indicates important differences among the archetypes (Table 2). For three of the four archetypes, participants were mostly connected through a centralized hub, with little connection to one another. The HH archetype included the most vertices \( (n = 267) \), while the LL archetype included the least \( (n = 146) \). In addition, the HH archetype had the highest number of vertices in a connected component. This suggests that within the HH archetype, there were many participants that were connected; however, the graph density was sparse, at 0.01, suggesting that the vertices were only loosely connected to one another. In comparison, the most dense archetype was HL with a graph density of 0.02. Metrics for graph density are calculated on a scale of 0–1, thus no archetype was densely connected. We interpret this to mean that within an egocentric network that was centered on the science of paleontology, connections were mainly facilitated by the egocentric node and few external conversations occurred outside of those created by the node.

For all archetypes, the shortest geodesic distance ranged from 2.32 (HH archetype) to 2.59 (LL archetype). This suggests that for all archetypes, nodes within the network were directly connected to one another or connected through a mutually affiliated entity. As this was a study of an egocentric network, @projectfossil is the most likely to be the mutually affiliated entity.

The network shape and directional flow of the information within the LL archetype indicates a broadcast network (Figure 7), according to Himelboim and colleagues (2017), a distinguishing feature of a broadcast network is one large group acting as the source of information for receiving groups. The LL archetype involved one main group that accounted for most of the information flow to 17 smaller groups. A large amount of information flowed into and out of the main group, with little connection between the other smaller groups, another common feature of broadcast networks (Himelboim et al., 2017). The low density (0.01) is a result of the lack of interconnectedness within the network.

The HH and LH archetypes resulted in networks with the highest number of groups \( (n = 19 \text{ and } n = 18, \text{ respectively}) \). Based on the shape and the directional flow of information, we interpreted both to also be broadcast networks. Like the LL archetype, HH and LH had the characteristic shape of a broadcast network, with much of the information flowing in and out of one main group. Although HH and LH were broadcast networks, they were both more interconnected than the LL archetype. The HH archetype network had the highest number of connections between groups \( (n = 39) \) and LH had the second most interactions between groups \( (n = 30) \) of all the archetype networks. While the HH and LH broadcast networks were more interconnected than LL, the high number of groups resulted in a similarly low density (0.01).

The HL archetype network involved far fewer groups \( (n = 5) \) than any other. Each of these groups were medium sized, and four of the five had information flowing among them. The size of the groups and the flow of information between them created a structure that is unique to Tight Crowd networks, a situation where participants in groups are tightly connected to each other for ideas, information, and opinions (Himelboim et al., 2017). Since there were so few groups and interconnectedness between these groups, the density of the network (0.02) was higher than the other archetype networks.

Table 2. Network Graph Characteristics by Topic Archetype.

| Graph metric                        | High Mag. Low Occur. | High Mag. High Occur. | Low Mag. Low Occur. | Low Mag. High Occur. |
|-------------------------------------|----------------------|-----------------------|---------------------|----------------------|
| Graph type                          | Directed             | Directed              | Directed            | Directed            |
| Vertices                            | 156                  | 267                   | 146                 | 221                 |
| Total edges                         | 574                  | 408                   | 270                 | 313                 |
| Self-loops                          | 1                    | 1                     | 1                   | 4                   |
| Reciprocated vertex pair ratio      | 0.21                 | 0.11                  | 0.07                | 0.12                |
| Reciprocated edge ratio             | 0.34                 | 0.20                  | 0.13                | 0.21                |
| Connected components                | 1                    | 2                     | 1                   | 1                   |
| Single-vertex connected components  | 0                    | 0                     | 0                   | 0                   |
| Maximum vertices in a connected component | 156                  | 264                   | 146                 | 221                 |
| Maximum edges in a connected component | 574                  | 406                   | 270                 | 313                 |
| Maximum geodesic distance (diameter) | 2.41                | 2.32                  | 2.59                | 2.27                |
| Average geodesic distance           | 0.02                 | 0.01                  | 0.01                | 0.01                |
| Graph density                       | 0.49                 | 0.44                  | 0.57                | 0.40                |
| Groups                              | 5                    | 19                    | 17                  | 18                  |
Discussion

While we are certain our study makes substantial contributions to the understanding of topic archetypes as community processes and provides evidence for indicating the health of an online community, this study is not without its limitations. Indeed, the study of interactions on Twitter is a limiting factor, as only 22% of adults in the United States use Twitter, and those that use it do not use it frequently, meaning that a large segment of people who may be interested in and contributing to paleontology are excluded from our analysis (Pew Research Center, 2019b). In addition, social media’s fluid and dynamic nature can make replication studies challenging as methods for collecting data change as algorithms, application programming interfaces (APIs), and scraping methods change. More research that takes into account multiple scientific communities as well as multiple platforms could provide a better understanding of scientific online communities. With these limitations in mind, we situate our results within the larger corpus of literature and describe how our research can illuminate new directions for designing, developing, and evaluating sustainable scientific online communities.

Topic Archetypes as Expressions of Community

Drawing from Himelboim and colleagues’ (2017) typologies of Twitter topic-network structures, we interpret the results of our study and propose archetypes that are related to the states of CoP development from Wenger and colleagues (2002). Specifically, we offer an interpretation and contextual renaming of the four distinct topic archetypes as expressions of behavior that are indicative of community processes occurring within the Twitter network (Table 3). With both piques of interest and sustained conversation, the HH archetype exemplifies Sustainable Stewardship—sustained momentum created by relevant, cutting-edge, domain-specific topics that were both “lively and engaging” (Wenger et al., 2002, p. 104). The HL archetype, characterized by little more than a singular spike of extreme activity, exemplifies Coalescing Community—having similar interests, focus, and knowledge, but scarce energy for assimilation. The lack of piqued activity but consistent conversation of the LH archetype exemplifies Mature Membership—members clarifying the community’s “focus, role, and boundaries” while developing “a comprehensive body of knowledge [that] expands the demands on community members” (Wenger et al., 2002, p. 97). Finally, the LL topic archetype with a lack of piqued activity and a conversation exemplifies Potential Practice—connecting on common grounds, an online community grappling with uniting their “heartfelt interests” (Wenger et al., 2002, p. 71) into something that aligns with that of the whole. The contribution of this study is to advance the CoP theoretical framework by examining how topic archetypes provide empirical evidence for detecting and indicating the life cycle of development,
change, and alteration of an online community. While these interpretations were appropriate for this online, social world, we foresee future research directions in determining if the CoP stages of development can be utilized to orient interpretations of community processes within other CoPs.

The Potential Practice (LL) archetype featured topics that engaged limited members, correlating to the CoP development stage of potential (Wenger et al., 2002). We postulate that topics within this archetype limit community development as they do not allow for newcomers to enter easily nor do they encourage widespread activity among other members (Kraut et al., 2012). Such archetypal aspects relate to how social worlds have been utilized to broadcast scientific research (Bex et al., 2019). Many scientists have taken to using Twitter to publicize new research findings (Côté & Darling, 2018; Didegah et al., 2018; Vainio & Holmberg, 2017); however, much of this research is laden with jargon and discussed with targeted groups, thus there is little opportunity for communication of this nature to sustain interest or interaction beyond a narrow few (Carlson & Harris, 2020; Shuai et al., 2012). Our research shows that within online scientific spaces, some topic archetypes are engaged with less frequently and by limited numbers of individuals. Engagement by a diverse community with the scientific, online, social world can be expanded via the employment of targeted topics based on anticipated responses such as those that include opportunities for members to share their interests. The topic archetypes identified here provide a means for understanding the impact of such an effort.

Archetypes that illustrated opportunities to share interests included Sustainable Stewardship (HH) and Mature Membership (LH). These archetypes are directly correlated with two mature stages of CoP development (Wenger et al., 2002), specifically maturing and stewardship. Within these CoP development stages, the focus of the community is shifted toward growth, change, and integration. Topics of the Sustainable Stewardship (HH) and Mature Membership (LH) archetypes highlighted community members interacting with comprehensive bodies of knowledge and engaging in “lively and engaging” sustained activity (Wenger et al., 2002, p. 104). In addition, the presence of such archetypes provides evidence for the health of online, scientific communities. Sustained activity is indicative of both identity- and bond-based attachment within the community (Kraut et al., 2012). These attachments imply commitment to the community’s purpose (identity-based attachment) and to members of the community (bond-based attachment). The description of these archetypes based on the CoP stages of development and their connection to different types of attachments is important as future researchers can employ such archetypal descriptions when exploring other scientific, online communities. If topic archetypes are indicative of topics engendering different patterns of communication within the community, then we

| Community Interpretation | Archetype | Characteristics | General Graph Shape | Example Topics |
|--------------------------|-----------|-----------------|---------------------|---------------|
| Sustainable Stewardship  | HH        | Piques of interest and sustained conversation | Sustainable Stewardship | Inclusion of Amateur Paleontologists, Citizen Science Project, In-person Events |
| Coalescing Community     | HL        | Piques of interest, but no conversation | Coalescing Community | Women in Paleontology, Microfossils, Paleoart |
| Mature Membership        | LH        | No piques of interest, but conversation | Mature Membership | Events at the Florida Museum, Paleontology Education & News, News about Paleontology |
| Potential Practice       | LL        | No piques of interest, minimal conversation | Potential Practice | Fossils from Various States, Project Outreach Events, Historical Paleontology |

HH = high Magnitude–high percent Occurrence; HL = high Magnitude–low percent Occurrence; LH = low Magnitude–high percent Occurrence; LL = low Magnitude–low percent Occurrence.
conjecture that targeting the production of certain archetypes through designed strategies offer the potential for growing and sustaining online communities.

The Potential of Using Diversity Indices for Research and Evaluation of Online Communities

The members of @projectfossil’s online social world were varied; through member classification using the PIT, we show that regardless of interest, diverse members of the social world participated in and contributed to various topic archetypes. Previous research has described network participation holistically with a secondary focus on centralized connectors (Brown et al., 2016; Grudz et al., 2016; Himelboim et al., 2017; Smith et al., 2018) or on the geographic location of where tweets originated (Pruss et al., 2019). Thus, when network participants are considered, they are broadly described as nodes, and or in terms of their connectivity to one another. Some exceptions exist in which community members are defined based on members’ narrative representations of self-identity with a domain (i.e., their Twitter biographies) (Bex et al., 2019; Kimmons & Veletisianos, 2016; Rosenberg et al., 2020; Vainio & Holmberg, 2017). However, this research is still emergent and little has been done regarding classification. Our method of classifying members via the PIT can be modified and used in future studies as training data for machine-assisted classifications.

Twitter is promoted as an effective science communication pathway where scientists can connect to the public (Bombaci et al., 2016; Van Noorden, 2014). In this online social world centered on the domain of paleontology, it is difficult to argue for Twitter’s effectiveness as a science communication medium as we did not find any topic archetypes that included a majority of Public participants. Within the online, social worlds of scientists, there is evidence for unsustainable activity and filter bubbles (Flaxman et al., 2016). The Potential Practice (LL) and Coalescing Community (HL) topic archetypes included high percentages of Scientists; these topic archetypes did not sustain activity. This finding expands work from Côté and Darling (2018) who found that a majority of scientists who used Twitter had mutual following relationships with other scientists unless their follower counts surpassed 1,000. To circumvent filter bubbles and circular conversation, our research indicates that diverse participants who can fulfill multiple roles need to be included to sustain interest-based activity (Wenger et al., 2002). Furthermore, such diversity of perspectives can help to remove filter bubbles (Min & Wohn, 2020) and increase knowledge generation (Didegah et al., 2018; Lei & Xin, 2011; Rosenberg et al., 2020).

Within the Mature Membership (LH) and Sustainable Stewardship (HH) topic archetypes, Education and Outreach members made up the majority of participants. We infer this to mean that some topics, like those in the Sustainable Stewardship (HH) archetype, allow for diverse participants to contribute to and participate in social paleontology for extended periods of time. This finding aligns with previous research on online paleontological social worlds (Bex et al., 2019) in which Education and Outreach members were able to connect across the network. In this study, increased numbers of Education and Outreach participants corresponded with archetypes in which sustained activity occurred.

Our method of applying diversity indices to a social network allows us to account for community membership. This study has shown that quantifying and qualifying diversity within a community is possible; such diversity affects the longevity and health of Twitter topic archetypes. While evidence for the importance of diversity indices exists in ecological studies across space and time (Patzkowsky & Holland, 2007; Stigall et al., 2017), such indices have, to the best of our knowledge, never been applied to an online social world. Pohjola and Puusa (2016) have suggested that community participation and group dynamics shape CoPs in that members’ interests can become dispersed and growth creates different roles. Additional studies that apply diversity indices to such online social worlds are needed to explore how low, medium, or high diversity of members can affect conversations and activity.

Exploring Topics as a Way to Grow and Sustain the Online Community Life cycle

The evolution and expression of a conversation over time within an online community is the essence of digital practice. Yet, our capacity to understand this phenomenon has been limited by the availability of tools and techniques for connecting the key characteristics of people with such activity in meaningful ways. Previous research has mainly considered network characteristics and structure as the means for analysis and inference (Himelboim et al., 2017), offering an important but limited perspective. By connecting the expression of conversation topics through topic modeling with the characteristics of participants, we were able to contextualize the network in a way that offers new insight about digital expressions of behavior that are indicative of community processes and provides empirical evidence for detecting and indicating the health of online communities.

In this study, community network structures showed sparse connectivity. Others have argued that sparse network structures facilitate diffusion of ideas among groups (Behfar et al., 2018) and that loosely connected networks benefit from entities that can act as central connectors (Ergün & Usuel, 2016). Networks emerge from basic principles of community, and thus can be explored via the CoP theoretical framework. From CoP perspective, the egocentric network
within this study could be acting as an effective CoP, as a loosely connected network with multiple groups can imply that more members are involved at varied positions of participation (Wenger et al., 2002). Our work has determined the archetypes, composition, and community network structures within an online social world within the domain of paleontology; future research should examine other scientific social networks to see if patterns vary depending on the domain.

Conclusion

Our findings demonstrate that distinct topics featuring a diverse assemblage of members have varied impacts on an online social world that was centered on paleontology. These impacts depended on topic composition—topics with greater magnitude and higher percent occurrence were associated with more diverse member composition and specific P3T post types (i.e., Opportunities and Information). While others have recognized how network structures (Himelboim et al., 2017), topics (Nikolenko et al., 2017), and member composition (Britt & Paulus, 2016; Xie & Luo, 2018) independently can be applied to online social worlds, our study shows that a time series approach, content analysis, and machine learning techniques such as topic modeling can be applied to understand and predict the ways that community members contribute to and participate in interest-based activities in online social worlds. In addition, this study provided a functional conceptual model for understanding and interpreting patterns of behavior as topic archetypes and a way to explore relationships between the nature of participants and the social network; we see the potential to further explore and replicate results within other scientific fields as well as within educational research.

Data Accessibility Statement

We have all supplementary data and code stored on Open Science Framework available at this link: https://osf.io/jvp6z/?view_only=d4e9bdeaf0d0411b96c788e0b803825e.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

All data collected complied with Twitter’s terms of service, as laid out by Twitter’s Developer Agreement and Policy, Section C entitled Respect Users’ Control and Privacy. This research included Institutional Ethics Review and approval (UF-IRB2020002652).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was funded in part by the National Science Foundation under Grant No. DRL-1322725. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ORCID iDs

Lisa Lundgren https://orcid.org/0000-0001-7358-4092
Jennifer E. Bauer https://orcid.org/0000-0002-6337-6270

References

Ahmed, W., & Lugovic, S. (2018). Social media analytics: Analysis and visualisation of news diffusion using NodeXL. *Online Information Review, 43*(1), 149–160. https://doi.org/10.1108/OIR-03-2018-0093

Behfar, S. K., Turkina, E., & Burger-Helmchen, T. (2018). Knowledge management in OSS communities: Relationship between dense and sparse network structures. *International Journal of Information Management, 38*(1), 167–174. https://doi.org/10.1016/j.ijinfomgt.2017.09.004

Bert, F., Zeegers Paget, D., & Scaioli, G. (2016). A social way to experience a scientific event: Twitter use at the 7th European Public Health Conference. *Scandinavian Journal of Public Health, 44*(2), 130–133. https://doi.org/10.1177/140349815612932

Bex, R. T., Lundgren, L., & Crippen, K. J. (2019). Scientific Twitter: The flow of paleontological communication across a topic network. *PLOS ONE, 14*(7), Article e0219688

Bombaci, S. P., Farr, C. M., Gallo, H. T., Mangan, A. M., Stinson, L. T., Kaushik, M., & Pejchar, L. (2016). Using Twitter to communicate conservation science from a professional conference. *Conservation Biology, 30*(1), 216–225. https://doi.org/10.1111/cobi.12570

Bondy, E., Beck, B., Curcio, R., & Schroeder, S. (2017). Dispositions for critical social justice teaching and learning. *Journal of Critical Thought and Praxis, 6*(3), Article 1.

Borgatti, S. P., & Ojem, B. (2010). Overview: Social network theory and analysis. In A. J. Daly (Ed.), *Social network theory and educational change* (pp. 17–29). Harvard Education Press.

Borrett, S. R., Moody, J., & Edelmann, A. (2014). The rise of Network Ecology: Maps of the topic diversity and scientific collaboration. *Ecological Modelling, 293*, 111–127. https://doi.org/10.1016/j.ecolmodel.2014.02.019

Britt, B. C., Britt, R. K., Hayes, J. L., & Oh, J. (2020). Continuing a community of practice beyond the death of its domain: Examining the Tales of Link subreddit. *Behaviour & Information Technology*. Advance online publication. https://doi.org/10.1080/0144929X.2020.1797173

Britt, V. G., & Paulus, T. (2016). “Beyond the four walls of my building”: A case study of redchat as a community of practice. *American Journal of Distance Education, 30*(1), 48–59. https://doi.org/10.1080/088923647.2016.1119609

Brown, J. S., & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science, 12*(2), 198–213. https://doi.org/10.1287/orsc.12.2.198.10116

Brown, M. E., Ihli, M., Hendrick, O., Delgado-Arias, S., Escobar, V. M., & Griffith, P. (2016). Social network and content analysis of the North American Carbon Program as a scientific community of practice. *Social Networks, 44*, 226–237. https://doi.org/10.1016/j.socnet.2015.10.002
Carlson, J., & Harris, K. (2020). Quantifying and contextualizing the impact of bioRxiv preprints through automated social media audience segmentation. *PLOS Biology*, 18(9), Article e3000860.

Carroll, D. (2005). Developing dispositions for teaching: Teacher education programs as moral communities of practice. *The New Educator, 1*(1), 81–100.

Catalani, J. (2014). Contributions by amateur paleontologists in 21st century paleontology. *Palaeontologia Electronica*, 17(2), Article 17.2.3E. https://doi.org/10.26879/143

Cela, K., Sicilia, M.-Á., & Sánchez-Alonso, S. (2016). Influence of learning styles on social structures in online learning environments. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, 47(6), 1065–1082. https://doi.org/10.1111/1467-8535.12267

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), Article 066111. https://doi.org/10.1103/PhysRevE.70.066111

Corin, E. N., Jones, M. G., Andre, T., Childers, G. M., & Stevens, V. (2015). Science hobbyists: Active users of the science-learning ecosystem. *International Journal of Science Education, Part B*, 7(2), 161–180. https://doi.org/10.1080/21548455.2015.1118664

Côté, I. M., & Darling, E. S. (2018). Scientists on Twitter: Preaching to the choir or singing from the rooftops?. *FACETS*, 3(1), 682–694. https://doi.org/10.1119/facets-2018-0002

Crippen, K. J., Ellis, S., Dunckel, B. A., Hendy, A. J. W., & Côté, I. M., & Darling, E. S. (2018). Scientists on Twitter: Preaching to the choir or singing from the rooftops?. *FACETS*, 3(1), 682–694. https://doi.org/10.1119/facets-2018-0002

Corin, E. N., Jones, M. G., Andre, T., Childers, G. M., & Stevens, V. (2015). Science hobbyists: Active users of the science-learning ecosystem. *International Journal of Science Education, Part B*, 7(2), 161–180. https://doi.org/10.1080/21548455.2015.1118664

De Cindio, F. (2012). Guidelines for designing deliberative digital habitats: Learning from e-participation for open data initiatives. *The Journal of Community Informatics, 8*(2), Article 3040.

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 87–103. https://doi.org/10.1007/s11412-007-9006-4

Didegah, F., Mejlgaard, N., & Sørensen, M. P. (2018). Investigating the quality of interactions and public engagement around scientific papers on Twitter. *Journal of Informetrics, 12*(3), 960–971. https://doi.org/10.1016/j.joi.2018.08.002

Edler, D., Eriksson, A., & Rosvall, M. (2016). The MapEquation software package (Version 1.1.3) [Computer software]. http://www.mapequation.org.

Ergün, E., & Usluel, Y. (2016). An analysis of density and degree-centrality according to the social networking structure formed in an online learning environment. *Educational Technology & Society*, 19(4), 34–46.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly, 80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Gabarron, E., Dorronzoro, E., Rivera-Romero, O., & Wynn, R. (2019). Diabetes on Twitter: A sentiment analysis. *Journal of Diabetes Science and Technology, 13*(3), 439–444. https://doi.org/10.1177/1932296818811679

Greenhalgh, S. P., Rosenberg, J. M., Staudt Willet, K. B., Koehler, M. J., & Akcaoglu, M. (2020). Identifying multiple learning spaces within a single teacher-focused Twitter hashtag. *Computers & Education, 148*, Article 103809. https://doi.org/10.1016/j.compedu.2020.103809

Gruzd, A., Paulin, D., & Haythomthwaite, C. (2016). Analyzing social media and learning through content and social network analysis: A faceted methodological approach. *Journal of Learning Analytics, 3*(3), 46–71. https://doi.org/10.18608/jla.2016.33.4

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

Gunawardena, C., Hermans, M. B., Sanchez, D., Richmond, C., Bohley, M., & Tuttle, R. (2009). A theoretical framework for building online communities of practice with social networking tools. *Educational Media International, 46*(1), 3–16. https://doi.org/10.1080/09523980802588626

Habibi, M. R., Laroche, M., & Richard, M.-O. (2014). Brand communities based in social media: How unique are they? Evidence from two exemplary brand communities. *International Journal of Information Management, 34*(2), 123–132. https://doi.org/10.1016/j.ijim.2013.11.010

Handley, K., Sturdy, A., Fincham, R., & Clark, T. (2006). Within and beyond communities of practice: Making sense of learning through participation, identity and practice. *Journal of Management Studies, 43*(3), 641–653. https://doi.org/10.1111/j.1467-6486.2006.00605.x

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). Analyzing social media networks with NodeXL: Insights from a connected world. Morgan Kaufmann.

Hecht, M., & Crowley, K. (2020). Unpacking the learning ecosystems framework: Lessons from the adaptive management of biological ecosystems. *Journal of the Learning Sciences, 29*(2), 264–284. https://doi.org/10.1080/10508406.2019.1693381

Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter topic-networks using social network analysis. *Social Media + Society, 3*(1), Article 769154. https://doi.org/10.1017/S2056305117691545

Hong, L., & Davison, B. D. (2010, July 25). Empirical study of topic modeling in Twitter [Proceedings]. First Workshop on Social Media Analytics SOMA ’10 (SOMA ’10). Association for Computing Machinery, New York, NY, United States.

Kimble, C., Hildreth, P. M., & Bourdon, I. (Eds.). (2008a). *Communities of practice: Creating learning environments for educators, Volume 1*. Information Age.

Kimble, C., Hildreth, P. M., & Bourdon, I. (Eds.). (2008b). *Communities of practice: Creating learning environments for educators, Volume 2*. Information Age.

Kimmorn, R., & Veletisianos, G. (2016). Education scholars’ evolving uses of Twitter as a conference backchannel and social commentary platform. *British Journal of Educational Technology: Journal of the Council for Educational Technology, 47*(3), 445–464. https://doi.org/10.1111/bjet.12428

Knaus, M., & Callcott, D. (2017). The lifecycle of a student-led community of practice in higher education. In J. McDonald & A. Cater-Steel (Eds.), *Implementing communities of practice in higher education* (pp. 423–446). Springer. https://doi.org/10.1007/978-981-10-2866-3_19

Kotu, V., & Deshpande, B. (2015). *Predictive analytics and data mining: Concepts and practice with RapidMiner*. Morgan Kaufmann.
Kraut, R. E., Resnick, P., & Kiesler, S. (2012). Building successful online communities. MIT Press.

Krippendorff, K. (2012). Content analysis: An introduction to its methodology (illustrated). SAGE.

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

Lei, G., & Xin, G. (2011). Social network analysis on knowledge sharing of scientific groups. *Journal of System and Management Sciences, 1*(3), 79–89.

Liberatore, A., Bowkett, E., MacLeod, C. J., Spurr, E., & Longnecker, N. (2018). Social media as a platform for a citizen science community of practice. *Citizen Science: Theory and Practice, 3*(1), 3. https://doi.org/10.5334/cstp.108

Lundgren, L., & Crippen, K. J. (2017). Developing social paleontology: A case study implementing innovative social media applications. In D. Remenyi (Ed.), *Social media excellence awards 2017: An anthology of case histories* (pp. 11–26). Academic Conferences and Publishing International.

Lundgren, L., Crippen, K. J., & Bex, R. T., II. (2018, October). Digging into the PIT: A new tool for characterizing the social paleontological community. [Proceedings]. E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2018. Association for the Advancement of Computing in Education (AACE) San Diego, CA United States.

Lundgren, L., Crippen, K. J., & Bex, R. T. (2022). Social media interaction as informal science learning: A comparison of message design in two niches. *Research in Science Education, 52*, 1–20. https://doi.org/10.1007/s11165-019-09911-y

Marques, M. M., Loureiro, M. J., & Marques, L. (2016). The dynamics of an online community of practice involving teachers and researchers. *Professional Development in Education, 42*(2), 235–257. https://doi.org/10.1080/19415257.2014.997396

Michaels, S., O’Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*(4), 283–297. https://doi.org/10.1007/s11217-007-9071-1

Min, S., & Wohn, D. (2020). Underneath the filter bubble: The role of weak ties and network cultural diversity in cross-cutting exposure to disagreements on social media. *The Journal of Social Media in Society, 9*(1), 22–38.

Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science, 43*(1), 88–102. https://doi.org/10.1177/0165551515617393

Nistor, N., Daxecker, I., Stanciu, D., & Diekamp, O. (2015). Sense of community in academic communities of practice: Predicators and effects. *Higher Education, 69*(2), 257–273. https://doi.org/10.1007/s10734-014-9773-6

Oksanen, J., Blanchet, F. G., Friendly, M.,Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O’Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). vegan: Community ecology package (Version 2.5-6) [Computer software]. R-Project.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). SAGE.

Patzkowsky, M. E., & Holland, S. M. (2007). Diversity partitioning of a Late Ordovician marine biotic invasion: Controls on diversity in regional ecosystems. *Paleobiology, 33*(2), 295–309. https://doi.org/10.1666/06078.1

Pew Research Center. (2019a). *Key takeaways from our new study of how Americans use Twitter*. https://www.pewresearch.org/fact-tank/2019/04/24/key-takeaways-from-our-new-study-of-how-americans-use-twitter/

Pew Research Center. (2019b). *Social media usage in the U.S. in 2019*. https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

Pohjola, I., & Puusa, A. (2016). Group dynamics and the role of ICT in the life cycle analysis of community of practice-based product development: A case study. *Journal of Knowledge Management, 20*(3), 465–483. https://doi.org/10.1108/JKM-06-2015-0227

Probst, G., & Borzillo, S. (2008). Why communities of practice succeed and why they fail. *European Management Journal, 26*(5), 335–347. https://doi.org/10.1016/j.emj.2008.05.003

Pruss, D., Fujinuma, Y., Daughton, A. R., Paul, M. J., Arnot, B., Albers Szafr, D., & Boyd-Graber, J. (2019). Zika discourse in the Americas: A multilingual topic analysis of Twitter. *PLOS ONE, 14*(5), Article e0216922. https://doi.org/10.1371/journal.pone.0216922

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science, 54*(1), 209–228. http://www.jstor.org/stable/20647980.

Raban, D., & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics, 122*(3), 1563–1581. https://doi.org/10.1007/s11192-020-03371-2

Rafail, P. (2017). Nonprobability sampling and Twitter. *Social Science Computer Review, 36*(2), Article 770943. https://doi.org/10.1177/0894439317709431

Rose, G. (2013). *Visual methodologies: An introduction to researching with visual materials*. SAGE.

Rosenberg, J. M., Reid, J. W., Dyer, E. B., Koehler, M., Fischer, C., & McKenna, T. J. (2020). Idle chatter or compelling conversation? The potential of the social media-based #NGSChat network for supporting science education reform efforts. *Journal of Research in Science Teaching*. https://doi.org/10.1002/tea.21660

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America, 105*(4), 1118–1123. https://doi.org/10.1073/pnas.0706851105

Saqr, M., Fors, U., Terde, M., & Nouri, J. (2018). How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PLOS ONE, 13*(3), Article e0194777. https://doi.org/10.1371/journal.pone.0194777

Shalchiero, S., & Eder, M. (2020). Topic modeling, long texts and the best number of topics. Some Problems and solutions. *Quality & Quantity, 54*, 1095–1108. https://doi.org/10.1007/s11135-020-03371-2

Shannon, C. E., & Weaver, W. (1962). *The mathematical theory of communication*. University of Illinois Press.

Shuax., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, Twitter mentions, and citations. *PLOS ONE, 7*(11), Article e47523. https://doi.org/10.1371/journal.pone.0047523
Simpson, E. H. (1949). Measurement of diversity. *Nature, 163*(4148), Article 688.

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A., & Gleave, E. (2009). Analyzing (social media) networks with NodeXL [Proceedings]. Fourth International Conference on Communities and Technologies—C&T ’09. https://doi.org/10.1145/1556460.1556497

Smith, P. S., Trygstad, P. J., & Hayes, M. L. (2018). Social network analysis: A simple but powerful tool for identifying teacher leaders. *International Journal of Leadership in Education, 21*(1), 95–103. https://doi.org/10.1080/13603124.2016.1195016

Smith, S. U., Hayes, S., & Shea, P. (2017). A critical review of the use of Wenger’s community of practice (CoP) theoretical framework in online and blended learning research, 2000–2014. *Online Learning, 21*(1). https://doi.org/10.24059/olj.v21i1.963

Stigall, A. L., Bauer, J. E., Lam, A. R., & Wright, D. F. (2017). Biotic immigration events, speciation, and the accumulation of biodiversity in the fossil record. *Global and Planetary Change, 148*, 242–257. https://doi.org/10.1016/j.gloplacha.2016.12.008

Tancoigne, E. (2019). Invisible brokers: “Citizen science” on Twitter. *Journal of Communication Management, 18*(6), Article A05. https://doi.org/10.22323/2.18060205

Vainio, J., & Holmberg, K. (2017). Highly tweeted science articles: Who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics, 112*(1), 345–366. https://doi.org/10.1007/s11192-017-2368-0

Van Noorden, R. (2014). Scientists and the social network. *Nature, 512*, 126–129.

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems, 94*, Article 101582. https://doi.org/10.1016/j.is.2020.101582

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

Watts, D. J., Zavaleta, J., Wilson, S., & Francisco, S. (2018). Developing an interdisciplinary and cross-sectoral community of practice in the domain of forests and livelihoods. *Conservation Biology, 32*(1), 60–71. https://doi.org/10.1111/cobi.12982

Wenger-Trayner, E., Fenton-O’Creevy, M., Hutchinson, S., Kubiak, C., & Wenger-Trayner, B. (2015). *Learning in landscapes of practice: Boundaries, identity, and knowledgeability in practice-based learning*. Routledge.

Wenger, E., Trayner, B., & de Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework*. The Netherlands: Ruud de Moor Centrum.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Wenger, E. (2000). Communities of practice and social learning systems. *Organization, 7*(2), 225–246. https://doi.org/10.1177/135050840072002

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice*. Harvard Business School Press.

Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? *Journal of Communication, 59*(1), 40–56. https://doi.org/10.1111/j.1460-2466.2008.01403.x

Xie, Q., & Luo, T. (2018). Examining user participation and network structure via an analysis of a Twitter-supported conference backchannel. *Journal of Educational Computing Research, 57*(5), 1160–1185. https://doi.org/10.1177/0735633118791262

**Author Biographies**

Lisa Lundgren (PhD, University of Florida) is an Assistant Professor in the Department of Instructional Technology and Learning Sciences at Utah State University. Using social network analysis and content analysis, Lundgren’s work investigates through whom, for whom, and under what conditions learning occurs.

Kent J. Crippen (PhD, University of Nebraska—Lincoln) is a Professor of STEM Education at the University of Florida. His work focuses on theoretically grounded design for the dual purpose of addressing contemporary, in situ learning problems while generating theoretical insight related to the process of learning and the relationships among people, tools, and context.

Jennifer E. Bauer (PhD, University of Tennessee) is a Research Museum Collection Manager at the University of Michigan, Museum of Paleontology. As a trained paleobiologist, Bauer applies ecological metrics and system-based thinking to explore social niches.

Richard T. Bex II (MEd, Lehigh University) is a PhD candidate in Science Education in the School of Teaching and Learning at the University of Florida. His research applies social network analysis and learning analytics to understand, evaluate, and improve environments for STEM learning.