Factors Impacting Social Media Users' Information Behavior:

The Concept of Social Noise

Tara Zimmerman, Sahar, Behpour, and Suliman Hawamdeh

University of North Texas, Denton, TX 76203, USA

tarazimmerman@my.unt.edu

Abstract. Social media communication involves the discussion and sharing of information in an environment subject to the influence of online relationships and perceived expectations of those in the social network. The ability to filter the resulting noise depends largely on our understanding of Social Noise and its underlying constructs. We introduce the concept of Social Noise and investigate methods of identifying it using a quantitative, data analytics approach. Understanding this phenomenon has taken on increasing importance as it can influence attitudes and behavior surrounding social issues, political campaigns, and other core areas of society. Results from the topic modeling and data clustering techniques represent part of ongoing research into Social Noise and general keywords and combinations of keywords related to its underlying constructs.

Keywords: Social Media, Information Behavior, Social Noise

1 Introduction

Use of social media is changing the everyday communication and information behavior of billions of people [1]. In mobile and fast-paced societies, individuals can struggle to develop and maintain meaningful personal relationships, so social media is used for interacting more frequently with an array of connections outside the limits of time and distance. However, the structure of social media that allows individuals' communication and interaction to be observed by members of their social network introduces additional levels of complexity into information behavior. For example, on Facebook, social activity among family, friends, professional connections, and strangers are all under constant observation by the social networks of everyone involved [2]. Blurring the line between interpersonal and mass communication in this way [3], social media communication is subject to the influence of the user's perceived expectations of those in their social network. This results in Social Noise, the influence of personal and relational factors on information received via social media which can confuse, distort, or even change the intended message. The presence of Social Noise is demonstrated when people interact differently with information on social media than if it were encountered privately due to the awareness of being observed by members of their social network. Under the influence of Social Noise, a user modifies communication based on external cues regarding what behavior is acceptable in an attempt to increase their social capital.

Lymperopoulos and Lekakos [4] note that social media interactions have taken on a global dimension causing “a host of complex social dynamics such as opinion formation, spread of ideas, influence, epidemics, and communication formation among others,” (p. 125). Seemingly innocuous individual behaviors on social media may be contributing to larger shifts in politics, cultural conflict, and even decisions about who we trust to deliver truthful information. Therefore, it is imperative that information science study social media information behavior in general and the effect of Social Noise specifically.

Social Noise has at least four key constructs: Relationship Management, Image Curation, Cultural Agency, and Conflict Engagement. Relationship Management refers to a user’s desire to build community with individuals or groups with high social value. This can be driven by the desire to be accepted within a particular group (formal or informal) or to connect with others and maintain good relationships [5]. Image Curation is the effort by a social media user to craft their online identities. This term denotes the intentional filtering of artifacts, such as posts, photos, comments, and shared links, performed by the user to create a personal exhibition that satisfies them [6]. Cultural Agency refers to a user’s understanding of their roles and responsibilities within social institutions and their willingness to speak out, believing in their power to shape culture through civic participation and active involvement in social issues [7]. Conflict Engagement is the level of social conflict with which a user is comfortable, and Barnidge [8] found that social media may increase perceptions of disagreement among users, particularly on Facebook which has become a hotbed of social conflict. Social Noise is not limited to these four constructs; however, these four are being initially proposed and tested. This poster reports the use of data analytics techniques and text mining to analyze data sets with the goal of identifying patterns and terminology indicative of Social Noise. The research question is: What words are seen in Facebook comments and posts that indicate elements of Relationship Management, Image Curation, Cultural Agency, or Conflict Engagement?

2 Related Work

Social Noise has emerged with the advent of social media and its widespread use. Claude Shannon’s [9] Mathematical Model of Communication notes the interference of noise between sender and receiver, identifying physical and semantic noise as entropy in the communication flow. Signals are encoded and decoded in the presence of personal and environmental noise that may distort or alter the message. Physical and semantic noise alone are inadequate when applied to social media communication because they do not account for potential distortions caused by the influence of social observation.

Alfred Bandura's [10] Social Cognitive Theory asserts that personal agency and social structure operate in tandem as causal structures for an individual's behavior. In 2001, Bandura [11] expanded his theory to include mass and social media, stating that when people encounter information online, they combine personal values with perceived values of their social network to determine their behavior toward information. Overlaying Shannon's model with Bandura's Social Cognitive Theory gives insight into the psychosocial mechanisms that influence social media information behavior. These theories have been used together previously by Benjamin Nye [12,13] in his Systems Model for Meme Transmission in which he used Shannon's model to illustrate the outward process of communication alongside Bandura's theory to illustrate the internal motivations of the individual. Nye [13] noted that, "These theories provide complementary processes for examining the flow of information between and within individuals" (p. 308) and that this combination could be used to understand and model other observable behaviors.

Valenzuela, Park, and Kee [14] noted that Facebook users align themselves with valued individuals, organizations, and belief systems. Ranzini and Hoek [15] found the perception of being watched by others can cause users to modify behavior to improve the impression they make when communicating on Facebook. Researchers have found that social media users identify with desirable people and ideas by strategically directing their interactions toward them, and conversely, they distance themselves from those with whom they do not want to be identified [16, 17].

Similarly, Su and Chan [2] concluded that Facebook behavior is influenced by social expectations as users respond to information differently depending on who posts it. Positive responses to others' posts indicate support, affirm connections, and build social status [18, 19, 20, 21]. Users build social capital by constructing responses that align with the group's belief system [14], sometimes even adjusting their reaction to affect social perceptions [22, 23]. Managing relationships is a prime motivation for Facebook use [24], and self-presentation is one of the main reasons people give for actively posting on social media [25].

3 Methodology

Content analysis and text mining are often used when studying social media communication to process user-generated text [26]. Three different methods of data analytics were selected for this study: Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and Kmeans clustering. Using three techniques simultaneously helped validate the data and view it from multiple perspectives, providing more information and broader understanding. Each process generated commonly used words, while clustering also helped show the relationships between some of the words. While all three techniques have unique approaches to the data and their results cannot be directly compared, each data processing technique provided its own set of information. These were then viewed collectively in order to generate the best possible interpretation of the data overall. Prior to the core data being processed, additional datasets focusing on Facebook discussions of political candidates and fake news were used as training datasets.

LDA is a topic modelling technique which describes the distribution of words within a corpus of documents, revealing a hidden layer of meaning [27,28]. Similarly, LSA is a text analytics algorithm for understanding large volumes of text that reveals word correlations [29]. The number of topics to be modelled in both LDA and LSA is not rigidly set and must be determined by the researcher. This is done by considering the most important variables in the data and the goal of the study: the extract keywords that provide better understanding of the data. Using Python programming language, it was simple to choose multiple, iterative numbers of topics to model and view the different results generated (Fig. 1). Fine tuning the parameters, we attempted to select the optimum number of topics but found that this outcome did not necessarily generate most valuable keywords and thus continued modifying the number of topics. For the LDA and LSA techniques in this particular research, optimizing the number of topics was not as important as grouping the data to provide insight into Social Noise. We reviewed keywords for each of the topics generated using the lens of Social Noise and found groupings of words that seemed to represent the influence of one or more of the four constructs of Social Noise.

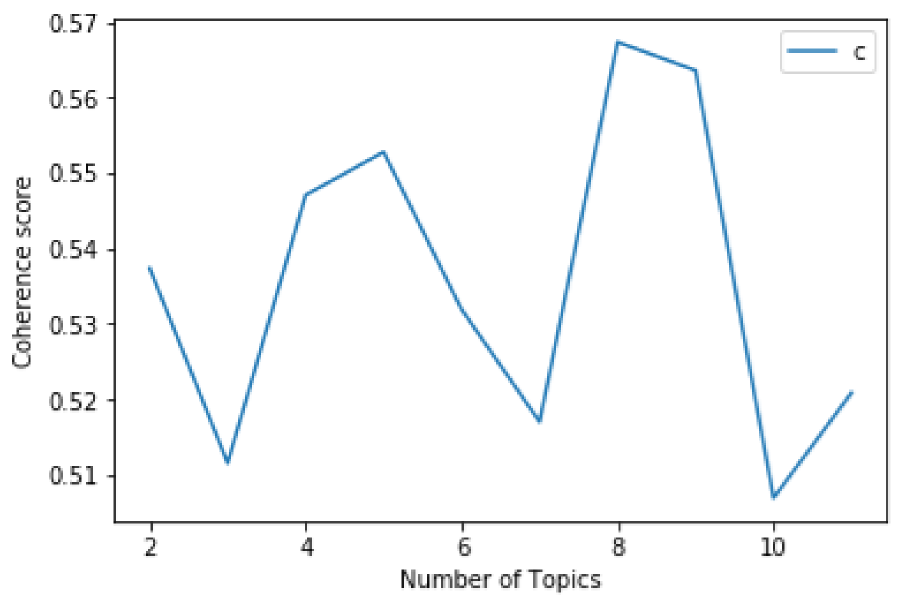

Fig. 1. Topic optimization.

Clustering was also used to discover underlying patterns in the dataset by aggregating data points, or keywords, based on similarities [30]. The algorithm randomly chooses the optimum number of topics as it groups the data. Similar to our work with LDA and LSA, we ran the program multiple times with different parameters in order to find the optimum number of clusters. For this research, five clusters provided the most informative results.

Data was collected from an open dataset representing Facebook posts and comments from five public Facebook groups serving Cheltenham Township, Pennsylvania, USA [31]. Approximately 10,000 unique comments and 10,000 unique posts from the dataset were used for analysis. The raw data was cleaned by tokenizing the text articles, removing stop words, and performing word stemming, similar to previous studies using topic modelling and clustering [32]. These pages are used by members of the community to discuss diverse subjects, from local issues such as traffic and lost pets, to politics and national news. These interactions are centered around neighborhood concerns but are also affected by personal and environmental factors that influence how users communicate with one another.

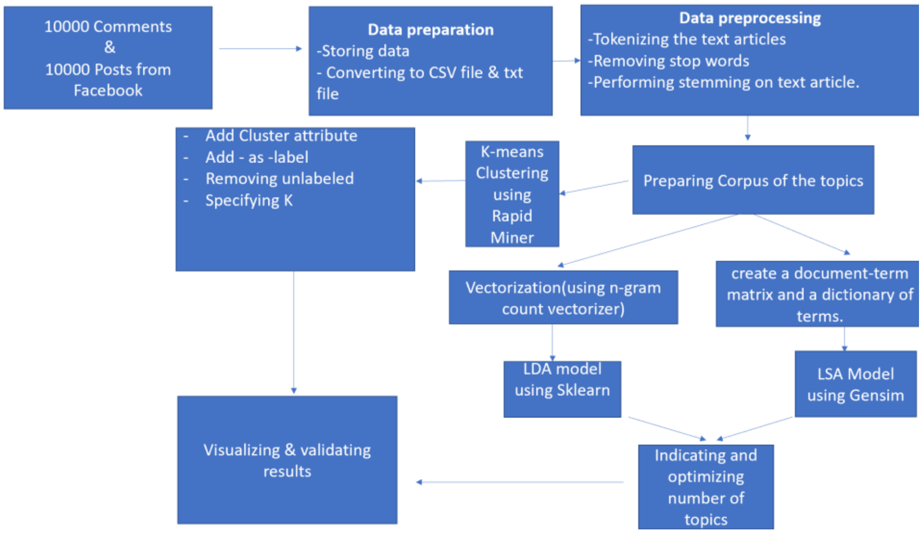

Fig. 2. Data collection, preparation, and processing.

Posts include words that may indicate the presence of one or more of the four constructs of Social Noise. An example post that may indicate the influence of Relationship Management says, “Hi Sandra, happy to help. Please give me a call at [phone number removed]”. This user is attempting to build a relationship with Sandra by using friendly words, offering help, and saying to please call him. It could be argued that this post might be an indicator of Image Curation as well, with the poster trying to present a friendly and helpful persona to the group. An example of a post seeming to indicate Cultural Agency appears in a discussion regarding the rash-causing plant poison ivy growing in the neighborhood. The poster says,” No that’s an old misconception, the Urushiol is not in the serous fluid. If you still have the oil on your hand, it’s another story. I’ve been doing research on it. [link to website].” This poster found relevant information about poison ivy and clearly wanted to share it with the group. This may be an effort in Image Curation as well, positioning oneself as knowledgeable and authoritative on a topic. Further in that same conversation, Conflict Engagement is almost certainly an issue in a post which says, “My goal with posting this is not to start a debate about global warming or climate change [link to website removed].” The poster shared something he thought might be controversial and purposely stated ahead of time that he did want to spark conflict in the group. This post could also represent Image Curation, indicating the poster does not want to be seen as someone who stirs up conflict.

4 Results

Results were visualized to aid in identifying patterns in the data. For the LDA results, each topic was represented as a bar graph. For example, Topic 66 (Fig. 3) appears to be a discussion of a lost dog in the neighborhood and includes words such as “help”, “please”, and “share.” These words are potential indicators of Relationship Management and Image Curation as users reach out for help from the group, ask them to share information, and present their request in polite terms. These words, along with “report,” may be indicators of Cultural Agency as users relay information to the group, presenting themselves as knowledgeable, helpful community members. Topic 48 (Fig. 4) regards an unknown person in the neighborhood and discussion of his whereabouts. Words in the topic, such as “parking,” “lot,” “missing,” “found,” “seen,” and “hurt,” are all possible indicators of Social Noise in the form of Cultural Agency or Conflict Engagement.

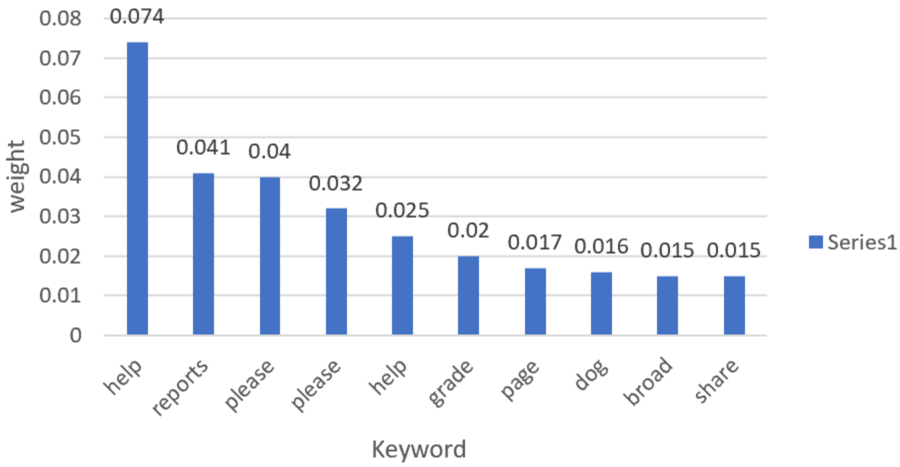

Fig. 3. Topic 66 of LDA results.

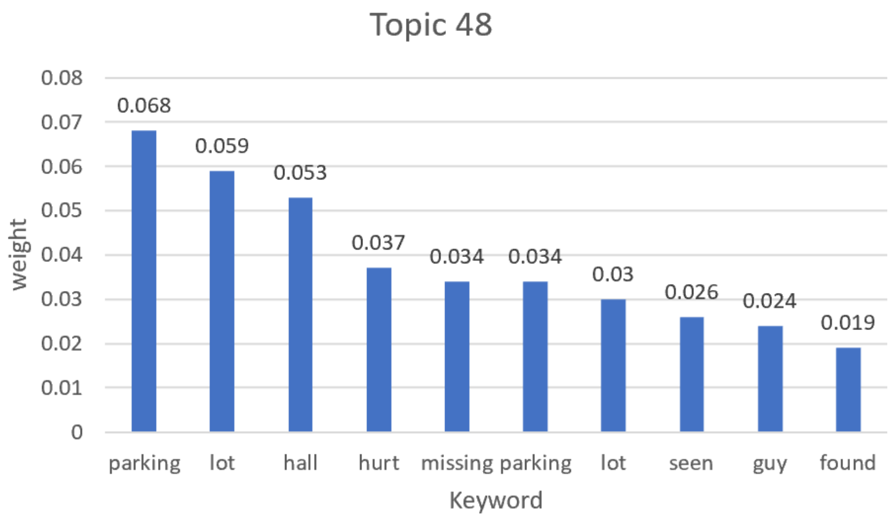

Fig. 4. Topic 48 of LDA results.

The results for each of the LSA topics were visualized in bar graphs as well, providing a basic view of subjects being discussed and their associated words. For example, Topic 0 seems to be discussing the community park and an upcoming event, with words like "want," "like," and "thank," possibly indicating the presence of Relationship Management or Image Curation (Fig. 5).

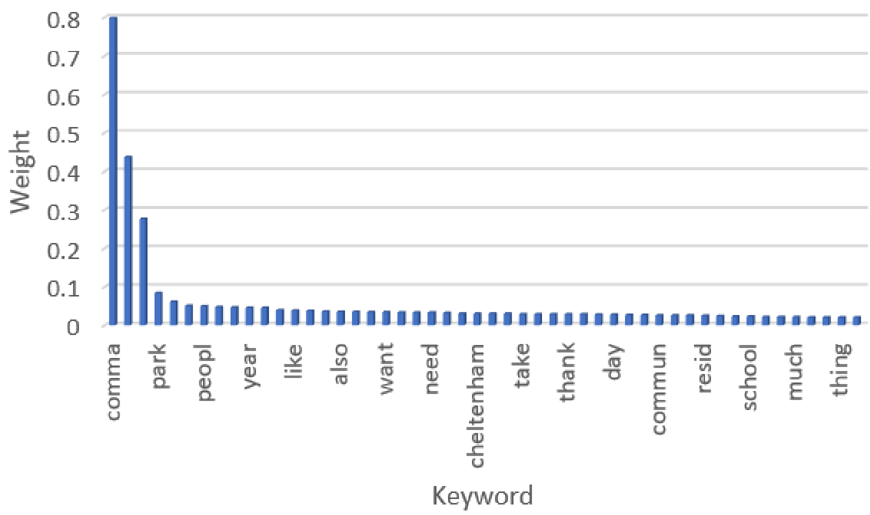

Fig. 5. Topic 0 of LSA results.

The clustering results grouped posts and comments from the dataset by inherent similarities, showing relative cohesion between words within the text and indicating broad subjects being discussed. Figure 6 shows the organizations of the data into five clusters, indicated by different colors. Distance and overlap of the data points show that Cluster 1 is the most consistently cohesive cluster, indicating a similar subject or topic being discussed.

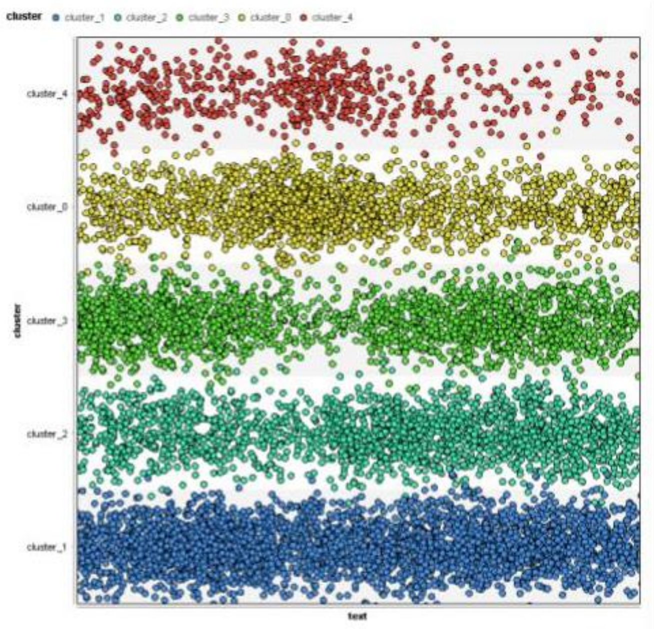

Fig. 6. Clustering results.

5 Discussion and Conclusion

For this initial study, we chose a text mining approach and the techniques of LDA, LSA, and clustering to search for latent patterns hidden in a large dataset of Facebook posts and comments. These topic modelling techniques did provide information about the content of the dataset and the subjects being discussed by the users; however, they did not give insight into the motivation for that communication, which is key to identifying Social Noise. The authors used their own analysis to interpret the keywords and groups of keywords that emerged from the data and understand how they could be reflections of Social Noise (Fig. 7). Although the results did not provide much information about semantic context of the individual words, these frequently used words from both posts and comments can be used in future studies to help identify possible indicators of the four constructs of Social Noise.

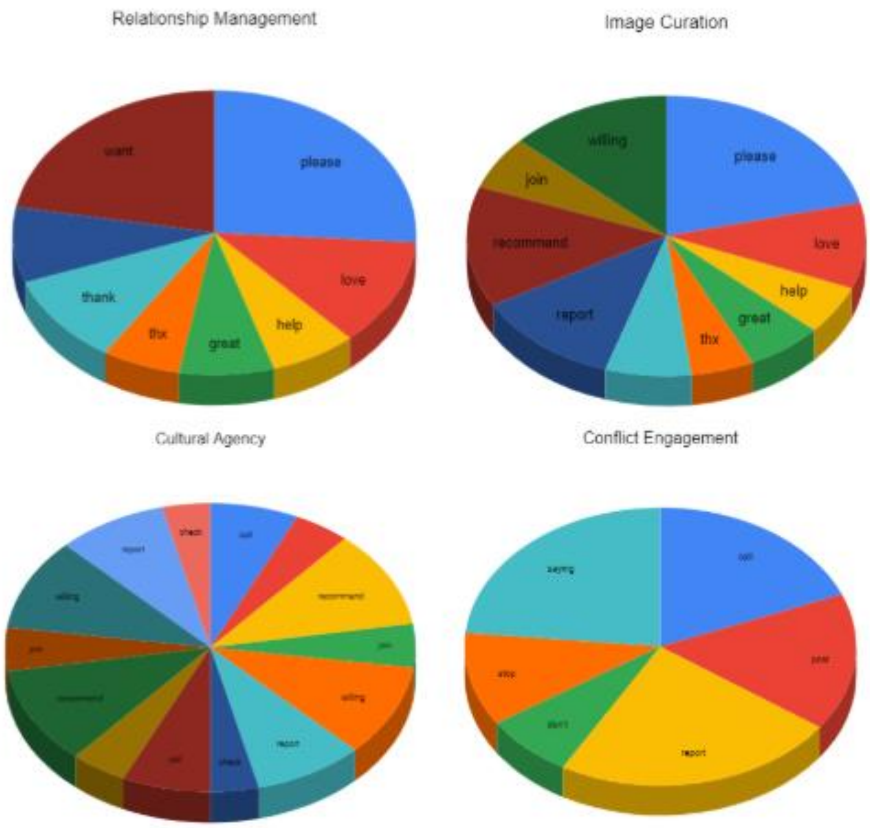

Fig. 7. Keywords possibly indicating Social Noise.

Future studies of Social Noise will benefit from this work in two ways. First, this study has revealed that using natural language processing techniques might result in greater insight into the semantic meaning of posts and comments in the dataset. Second, topic modelling and clustering techniques used in this early investigation of Social Noise provide some insight into commonly used words that may indicate the presence of Social Noise via Relationship Management, Image Curation, Cultural Agency, or Conflict Engagement. However, these quantitative techniques do not provide enough information to confirm either the presence or absence of Social Noise. This exploratory research will inform ongoing approaches and techniques in studying Social Noise with additional results to be shared in the future.

References

1. Abeele, M.V., de Wolf, & Ling, R.: Mobile media and social space: How anytime, anyplace connectivity structures everyday life. Media and Communication 6(2), 5-14 (2018).
2. Su, C.C., Chan, N.K.: Predicting social capital on Facebook: The implications of use intensity, perceived content desirability, and Facebook-enabled communication practices. Computers in Human Behavior, 72, 259-268 (2017).
3. Cappella, J.N.: Vectors into the future of mass and interpersonal communication research: Big data, social media, and computational social science. Human Communication Research, 43, 545-558 (2017).
4. Lymperopoulos, I., Lekakos, G.: Analysis of social network dynamics with models from the theory of complex adaptive systems. Paper presented at the *AICT-399*, pp. 124-140 (2013).
5. Lin, K. Y., Lu, H. P.: Why people use social networking sites: An empirical study integrating network externalities and motivation theory. Computers in Human Behavior, 27(3), 1152–1161 (2011).
6. Hogan, B.: The presentation of self in the age of social media: Distinguishing performances and exhibitions online. Bulletin of Science, Technology & Society, 30(6), 377-386 (2010).
7. Garry, S.: A study of cultural agency and an analysis of its application. http://doras.dcu.ie/20300/, accessed 09/12/2019.
8. Barnidge, M.: The role of news in promoting political disagreement on social media. Computers in Human Behavior, 52, 211-218 (2015).
9. Shannon, C. E.: A mathematical theory of communication. Bell System Technical Journal, 27(3), 379-423 (1948).
10. Bandura, A.: Social foundation of thought and action: A social cognitive theory. Prentice Hall, Englewood Cliffs, NJ (1986).
11. Bandura, A.: Social cognitive theory of mass communication. Mediapsychology, 3, 265-299 (2001).
12. Nye, B. (2011). Modeling memes: A memetic view of affordance learning (Doctoral dissertation). Retrieved from University of Pennsylvania Scholarly Commons http://repository.upenn.edu/edissertations/1.
13. Nye, B. (2014). Cognitive modeling of socially transmitted affordances: a computational model of behavioral adoption tested against archival data from the Stanford Prison Experiment. *Computational and Mathematical Organization Theory*, *20*(3), 302-337.
14. Valenzuela, S., Park, N., Kee, K.F.: Is there social capital in a social network site? Facebook use and college students' life satisfaction, trust, and participation. Journal of Computer-Mediated Communication, 14(4), 875-901 (2009).
15. Ranzini, G., Hoek, E.: To you who (I think) are listening: Imaginary audience and impression management Facebook. Computers in Human Behavior, 75(October), 228-235 (2017).
16. Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S.: Social media: Get Serious! Understanding the functional building blocks of social media. Business Horizons, 54(3), 241-251 (2011).
17. Scissors, L., Burke, M., Wengrovitz, S.: What's in a Like? Attitudes and behaviors around receiving Likes on Facebook. In: *Proceedings of the 19th ACM Conference on Computer Cooperative Work and Social Computing*, pp. 1501-1510 (2016).
18. Gan, C.: Understanding WeChat users' liking behavior: An empirical study in China. Computer Human Behavior, 30-39 (2017).

19. Hayes, R.A., Carr, C.T., Wohn, D.Y.: One click, many meanings: Interpreting paralinguistic digital affordances in social media. Journal of Broadcasting and Electronic Media, 60(1), 171-187 (2016).
20. Kim, J.W.: Scan and click: The uses and gratifications of social recommendation systems. Computer Human Behavior, 33, 184-191 (2014).
21. Sumner, E.M., Ruge-Jones, L., Alcorn, D.: A functional approach to the Facebook Like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. New Media & Society, 20(4), 1451-1469 (2018).
22. Lee, E., Kim, Y.J., Ahn, J.: How do people use Facebook features to manage social capital? Computer Human Behavior, 36, 440-445 (2014).
23. Hollenbaugh, E.E., Ferris, A.L.: Facebook self-disclosure: Examining the role of traits, social cohesion, and motives. Computers in Human Behavior, 30, 50-58 (2014).
24. Rui, J., Stefanone, M. A.: Strategic self-presentation online: A cross-cultural study. Computers in Human Behavior, 29(1), 110-118 (2013).
25. Miller, D., Costa, E., Haynes, N., McDonald, T., Nicolescu, R., Sinahan, J., Spyer, J., Venkatraman, S., Wang, X.: How the world changed social media. UCL Press, London (2016).
26. Abrahams, A., Fan, W., Wang, G., Zhang, Z., Jiao, J.: An integrated text analytic framework for product defect discovery. Production and Operations Management, 24(6), 975–990 (2015).
27. Topic Modelling in Python with NLTK and Gensim, https://towardsdatascience.com/topic-modelling-in-python-with-nltk-and-gensim-4ef03213cd21, last accessed 09/12/2019.
28. Intuitive Guide to Latent Dirichlet Allocation, https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-latent-dirichlet-allocation-437c81220158, last accessed 09/12/2019.
29. Latent Semantic Analysis using Python, https://www.datacamp.com/community/tutorials/discovering-hidden-topics-python, last accessed 09/14/2019.
30. Understanding K-means Clustering in Machine Learning, https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa, 09/14/2019.
31. Cheltenham's Facebook Groups, https://www.kaggle.com/mchirico/cheltenham-s-facebook-group, last accessed 2016/11/21.
32. Al-Daihani, S. M., & Abrahams, A.: A text mining analysis of academic libraries' tweets. The Journal of Academic Librarianship, 42(2), 135-143 (2016).
33. Case, D. O., Given, L. M. Looking for information: A survey of research on information seeking, needs, and behavior. 4th edn. Emerald, Bingley, UK (2016).