Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023

Social Media and Digital Collaboration

Dec 11th, 12:00 AM

Understanding Behavioral Drivers in Twitter Social Media Networks

Wingyan Chung
The University of Texas at Tyler, wingyan@acm.org

Understanding Behavioral Drivers in Twitter Social Media Networks

Short Paper

Wingyan Chung
Department of Computer Science
Soules College of Business, The University of Texas at Tyler
3900 University Blvd., Tyler, Texas 75799, U.S.A.
wchung@uttyler.edu

Abstract

As social media platforms facilitate user interactions, organizations increasingly use social media networks (SMNs) to build network ties. Studying user behavior on SMNs can help to uncover strategic information and improve situation awareness. However, there is a lack of understanding of behavioral drivers of SMN participants. This research developed a theoretically-based IS development framework for modeling user behavior in large evolving SMNs. To demonstrate the feasibility of our framework, we developed a proof-of-concept system for simulating user activities in the SMNs of Twitter social communities. Our system models the complex behavioral features in the SMNs by using a wide range of theoretically-driven features and machine-discovered features, and predicts user activities by using a pipeline of statistical and machine-learning techniques. Preliminary results of a simulation study provide insights of the importance of comprehensive network features to model SMN group behavior accurately and quality of commitment features to model SMN user behavior.

Keywords: social media networks, design science, computational modeling, simulation, machine learning, behavioral drivers.

Introduction

Social media platforms are widely used to support digital collaboration among project workers, distributed teammates, campaign participants, and software developers. These platforms help to understand user behavior to support strategic decision making on brand management, talent recruitment, supply chain management, and innovation forecasting (Cui et al. 2018; Debolina 2018; Hu et al. 2019). A report finds that 97 percent of Fortune 500 companies rely on social media platforms to promote their initiatives and communicate with stakeholders (Porteous 2021). Customers routinely share their experience on social media platforms through posting messages, replying to other customers' messages, and communicate with companies directly. Companies collect and analyze these messages and customers' behavioral data to predict sales, to identify trends, and to increase situation awareness.

As social media platforms facilitate user interactions, organizations increasingly use social media networks (SMNs) to build relationship and form network ties (Quinton and Wilson 2016). An SMN is a digitized network of entities that represent corresponding entities connected socially in the physical world (Chung et al. 2019). For example, a node in an SMN can represent a user and a link can represent a message sent from a user to other users. Using SMNs, companies can examine users' complaints on product issues discussed in Twitter SMNs to facilitate timely resolution. Managers may follow tweets and retweets to understand brand perception and discover new topics of interest to their organizations (MacMillan and Torbati 2021).

Studying user behavior on SMNs can help to uncover strategic information to guide product offering and improve situation awareness.

Despite the widespread adoption of social media platforms by companies and customers, there is a lack of understanding of behavioral drivers of SMN participants. Prior research in social media analytics has examined specific facets of user behavior, such as sentiment prediction, interpersonal interaction, and rating attribution (Kane et al. 2014; Xie et al. 2022). By contrast, prediction of comprehensive behavioral features covering multiple dimensions (e.g., time, action, type, entity) is scarce. This type of high-fidelity prediction can possibly help managers to extract useful intelligence from an online environment to support timely and accurate decision making.

In this research, we proposed an information systems (IS) development framework for modeling user behavior in large evolving SMNs. Our goal is to support new design and development of IS to enhance comprehensive understanding and accurate prediction of user behavior in SMNs. Our framework is built upon social, psychological, and network theories that explain human collective behavior and social interaction in groups undergoing changes (Reicher 2001). To demonstrate the feasibility of our framework, we developed a proof-of-concept system for simulating and predicting user activities in the SMNs of Twitter social communities. Our system models the complex behavioral features in the SMNs by using a wide range of hand-designed features and machine-discovered features, and predicts user activities by using a pipeline of statistical and machine-learning techniques. We conducted a preliminary study to compare different approaches to modeling the SMN behavior, with a view to understand the way our models performed with respect to ground-truth data. Our research seeks to answer several questions pertinent to digital collaboration and IS design: (1) What are representative features that characterize user behavior in evolving SMNs? (2) How can a framework for modeling user behavior be developed? (3) What is the performance of applying the models to predicting and simulating activities in Twitter social communities?

Literature Review

Digitization of many aspects of human life and the rise of social media have accelerated the study of complex human collective behavior. SMNs formed in digital collaboration have been examined in such various disciplines as information systems (Kane et al. 2014), computer science (Wei and Carley 2015; Yasami and Safaei 2017), physics (Holme and Saramäki 2012), sociology (Wasserman and Faust 1994), statistics (Snijders et al. 2010), economics (Jackson 2008), and public policy informatics (Chung and Zeng 2016; Zeng 2015), among others. The prevalence of social media promotes further interest in the modeling and detection of human behavior in SMNs, prompting the development of new methods and approaches. The following provides a review of relevant theories and computational methods to model behavior in SMNs.

User Behavior in Social Media Networks

Theories can be used to explain and predict user behavior (Wheeler 1966), thereby providing clues to identify behavioral features in online social movement and digital collaboration (Stage 2013) (e.g., in SMNs). Existing theories can be categorized into three streams. *First*, traditional social theories emphasize on collective behavior observed from large groups. Social Contagion Theory (Reicher 2001) asserts that people behave based on the information available to them individually (e.g., rational thought, experience). In a large crowd, information is often contradictory because of a lack of agreement (Le Bon 1895), forcing individuals to look for additional cue. Emergent Norm Theory posits further that new norms occur when leaders and their members agree on a new normative status or purpose for the group (Turner and Killian 1957). Generally, mass contagion is more spontaneous and rapid than the emergence of new social norms (Quan-Haase 2015). Individuals need time to observe the emerging norms and dynamics of a mass group, and gradually identify with the group mentality (Gino et al. 2009) and purpose (Reicher 2001). This stream of theories point to importance of group activity features (activity level, masses, crowds) in representing SMN behavior.

Second, network theories emphasizes social positions of individuals and characterizes individuals as nodes and their relationships as links in the network. Social Interaction Theory states that people make decisions based on their social neighbors' decisions (Becker 1974). Cognitive Balance Theory states that two persons who have separate relationships to a third person tend to build a connection between them (Heider 1958).

These theories point to importance of user features (neighbors, social actors) and their network positions (link degree, closeness, clustering) in representing SMN behavior.

Third, social and psychological theories emphasize interaction among individuals and their evaluation in a social relationship. Social Comparison Theory (Festinger 1954) states that individuals have a drive to compare themselves to other people, possibly through online rating in SMNs. Social Exchange Theory states that people engage in social interaction, expecting such rewards as respect, approval, or recognition (e.g., positive rating) (Emerson 1976). Relational cohesion theory posits that frequent exchanges among individuals create relational cohesion, which triggers commitment behaviors such as token gifts, positive sentiment, and subjective comments. This stream of theories point to importance of quality of user commitment in representing SMN behavior.

Computational Modeling of Social Media Networks

Computational modeling has been examined to model SMN user behavior, detect community, and study social influence (Du et al. 2015; Leskovec et al. 2009; Wang et al. 2015). Kong et al. developed an evolutionary user score to measure herding behavior in temporal social networks and predict the future growth of networked items in Twitter, Weibo, Delicious, and Amazon (Kong et al. 2023). However, their prediction calculates only the statistical correlation between the users' behavior and does not use any feature information about the items or users. Bacaksizlar and Galesic examined commenter networks of U.S. political news websites and found that collectives adapt to the real or imagined threat from an outgroup by changing their network structure in favor of fewer influential members (Bacaksizlar Turbic and Galesic 2023). They used four measures (2 individual-based and 2 network-based indices) to characterize inequality of attention in SMNs. Other approaches include supervised machine learning for predicting links in SMNs (Scellato et al. 2011), deep learning approach to community detection using feature extraction and proximity information (Wu et al. 2020), and a framework for measuring information spread in online social platforms by network topology and user type (bot or not) in cybersecurity discussion forums (Shrestha et al. 2020). These approaches make use of some selected features relevant to the tasks specified in the respective research. However, approaches that use comprehensive features of overall network and individual users to study collective behavior in SMN are not available. These features could have implication for not only specific tasks or applications, but may benefit understanding of human behavior in large evolving SMNs.

An IS Framework for Modeling User Behavior in SMNs

Prior research has provided theories and techniques to study human social interaction and behavior. However, prior studies have not employed theoretically-driven features to examine SMN behavior. Recent rapid proliferation of social media has created large evolving SMNs that are more complex and dynamic than those studied before. Existing computational techniques use selected measures such as item correlation and node degrees to measure nodal activities and link weights. More complex techniques like deep learning (DL), convolution neural networks, and auto-encoders (Wu et al. 2020) convert social networks into abstract representations that may not handle dynamically changing relationships, which are commonly found in SMNs. The abstract features as well as the predictive methods in DL approaches are often considered a "black box" and cannot be interpreted easily.

Design and Rationale

This research developed an information systems (IS) framework for modeling user behavior in large evolving SMNs. Developed based on a design science paradigm (Hevner et al. 2004), the framework incorporates all the core types of theoretically-driven features (group, user and network, user commitment) identified in the literature to build predictive models of SMN behavior and support the use of machine learning and time-series techniques for dynamic prediction of SMN activities. The framework emphasizes modeling SMNs as temporal graphs and representing temporal online social behavior using theoretically-based features. The design artifacts include the IS framework, a temporal network model built upon the framework, and instantiations of the model with application to simulating SMN activities in Twitter communities.

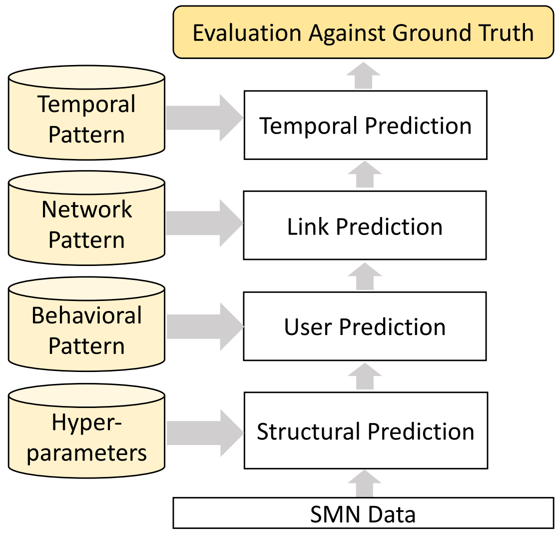

Figure 1. An IS Framework for Modeling User Behavior in SMNs

The framework aims to accurately predict occurrences of users (modeled as nodes) and their activities (modeled as links) in future SMNs, and to simulate these activities with a high fidelity. In the lowest part of Figure 1, SMN data (to be explained below) are fed into the structural prediction model, which uses hyper-parameters specified for learning the activities in the networks. The four components of the framework predicts the macroscopic structure of the SMN at the immediate next time step, the users, the network links, and timestamps of the links. These components use specific computational techniques in their predictions.

Computational Techniques

First, the structural prediction model predicts a macroscopic network structure using a weighted time series model that combines averaging and different exponential aggregations using transmission and time-window hyperparameters (Shumway and Stoffer 2017). The results include the predicted counts of users and links that are then fed to the next two models. *Second*, the user prediction model makes use of machine learning (ML) techniques and a specialized set of features (to be discussed below) to predict the users (whose total count was predicted by the structural prediction model) that will participate in the SMN at the immediate next time step. We selected recurrent neural network (RNN) as the ML technique for this prediction because of its flexibility and predictive accuracy in wide-ranging sequence prediction tasks (Yu et al. 2019). In addition, we used support vector machine (SVM) (Vapnik 1995) as a benchmark for comparing with RNN. RNN can allow highly efficient training and testing that are needed for this large-scale fine-grained prediction. SVM uses a different learning strategy that serves as a good comparison against RNN's learning approach. *Third*, the link prediction model uses the same ML techniques in the previous step to predict links (whose total count was predicted by the structural prediction model). *Fourth*, the temporal prediction model predicts fine-grained timestamps (hourly time steps) of the links by following temporal patterns learned from the training data. These four components of the framework constitute the Temporal Network Model (TNM).

Data Features

Existing computational techniques use some basic measures (such as correlation and node degree) to learn from activities in SMN. By contrast, TNM uses a comprehensive set of behavioral features to capture SMN activity patterns. These features are categorized into three aspects: user, group, and network. These aspects correspond to three levels of activities respectively – individual, community, and macroscopic environment. For each of the first two aspects (user and group), we identified three sets of features: social activity, network positioning, and quality of commitment. A user is modeled as a node in an SMN, whereas a group consists of a set of users, often having a common interest or having the same affiliation (e.g., commentators of the same discussion post in an SMN; discussants using the same hashtag on Twitter). *Social activity* (SA) features represent the level of activities of the users or groups being considered. *Network positioning* (NP) features represent the centrality and connectedness of the users or groups in the SMN. *Quality of commitment* (QC) features represent the scores or ratings provided by other users or groups (e.g., counts of up-vote or down-vote). For the network aspect, we computed statistics of each SMN snapshot captured over time. For example, an SMN snapshot of Twitter consists of the users (nodes), the messages (links that connect users who send and who receive messages, and message types), and timestamps of the messages.

Using the data extracted from SMNs, TNM simulates new predicted snapshots of SMN to represent all activities of users and messages together with message timestamps. The predictions are then compared with ground-truth (i.e., the actual SMNs) to evaluate accuracies of the simulation and predictions. This comparison helped to verify the prediction of TNM against actual SMN behavior.

The rationale of the design is to provide a high flexibility in modeling and in entity representation in online environments and to enable high fidelity in network simulation and activity prediction.

Simulation Study and Findings

We conducted a simulation study to understand the performance of the TNM in predicting and simulating activities on Twitter. We designed the study to compare performances among the use of different types of features to examine effectiveness of modeling users, groups, and networks.

Dataset

The Twitter dataset consists of 2,566,923 events (messages) related to cybersecurity (e.g., hackers, malware, ransomware), which was selected because of its rapidly growing importance as shown in recent research using various approaches, such as content analysis and event study (Nikkhah and Grover 2022), deep learning (Ebrahimi et al. 2022; Samtani et al. 2022), and sequential mixed methods (Cichy et al. 2021). The data cover the time period from June 30 to August 31, 2017. The dataset contains these fields: date of an event, time of an event, type of an event (quote, tweet, retweet or reply), ID of an event, ID of the user in an event, user name of the user in an event, number of followers of the user in an event, number of following of the user in an event, number of tweets this user had posted, number of likes this user gave to other tweets, number of tweets this user posted in lists (list is a group of people like a channel in YouTube), registered date of this user, registered time of this user, number of quotes this event has, number of replies this evant has, number of retweets this event has, number of likes this event has, event ID of that this event quoted or replied to, location of this user, hashtags in this event, and textual content of this event (non-English content is purged, encoded in UTF-8). Some sample tweets are shown below.

- "Exploring Peer to Peer Botnets" nice analysis on kehlios botnet https://t.co/uObR3Fcwk3 #infosec
- Understanding Distributed Denial of Service Attacks https://t.co/TJatDdZXyv #DDoS #BotNet #OpNewBlood #Anonymous
- Someone Hacked into Botnet Network & Replaces Malware Infection with an #Antivirus Installer https://t.co/HEodUYFT8a https://t.co/lIEKlB9ILy
- New #ransomware variant taunts victims with audio ransom note. https://t.co/cBzS373Sj4 #infosec
- OS X Malware With a Possible #HackingTeam Connection w @patrickwardle via @threatpost https://t.co/Ur31he337C https://t.co/Nmkqa2fNdC

Evaluation Measurements and Metrics

To compare the performance between SVM and RNN used in the link prediction model, three measures were used: precision, recall, and F-score, which are widely used in automatic classification and profiling (Kowsari et al. 2019). Each event specifies a user (sender), an event type (a post or a reply), a user (recipient, none in case of a new post), and a time stamp). The events occurred in the ground truth were used to compare against simulated events to measure performance of an approach using two categories of measurements: network and user. Network measurements evaluate the quality of the SMN structure, and include the number of edges (num_edges), the number of nodes (num_nodes), the distribution of node degrees in network (degree_dist), the assotativity coefficient (assortativity_coef), the average clustering (avg_clustering), the community modularity (comm_modularity), the density of the network (density), the maximum degree of all nodes (max_node_degree), the mean degree of all node (mean_node_degree), the mean shortest path of all the shortest path (mean_shortest_path), the number of connected component in the network (num_connected_component). User measurements evaluate the quality of how information is spread among users, and include the distribution of user activities for all users (activity_dist), the most active users with respect to the total number of actions per user (most_active), the most popular users with respect to the number of comments across posts posted by each user (user_popularity), disparity of gini coefficient of the quote events (disp_gini_quote), disparity of gini coefficient of the reply and comment events (disp_gini_reply/comment), gini coefficient of all the events (gini_coef), palma coefficient of all the events (palma_coef), and the unique content of all the events (unique_content).

Each of the above measurements is applied to evaluating the simulated sequence produced by TNM. Then, each measurement value was compared against the ground truth (the benchmark) to produce a metrics value. This evaluation procedure is less intrusive (than using human subjects in an experiment) and provides more objective results than subjective methods such as user ratings and expert interview (Endsley 2012). Absolute Percentage Error (APE) was used to measure absolute percentage error difference between simulation and ground truth. Root Mean Square Error (RMSE) measures pairwise residuals between simulation and ground truth. Jensen-Shannon Divergence (JSD) was used to compare distributions between simulation and ground truth. Rank-Biased Overlap (RBO) was used to compare weighted non-overlapping ranked-lists based on the number of overlapping elements as a function of list depth. A high value of RBO indicate good performance whereas low values of APE, JSD, KS, and RMSE indicate good performance.

Results and Discussion

Preliminary results of the simulation study shows diverse performances across different feature types. Table 1 provides the performance values of using features of user social activity (UserSA), user network position (UserNP), user quality of commitment (UserQC), and all feature groups to to predict and simulate SMNs. UserQC features outperform other feature types in Twitter's network measurements (better performance is observed in 7 of 11 network measurements). However, in Twitter's user measurements, user's social-activity and quality-of-commitment features perform similarly (both have better performance in 3 of 9 user measurements). Table 2 provides the performance values of using features of group social activity (GroupSA), group network position (GroupNP) and both feature groups (Both) to predict and simulate SMNs. GroupNP features outperform GroupSA features for both network and user measurements.

The results provide several important insights. *First*, quality-of-commitment features have promising performance in most network measurements (better performance is observed in 10 of 20 measurements). Social-activity features also have good performance in most user measurements (better performance is observed in 5 of 20 measurements). *Second*, network positioning features have promising performance for predicting group activities. This highlights the importance of using comprehensive network features to gauge user activities in SMNs. These features serve to represent key drivers of SMN behavior.

Categ.	Measurement	Metric	UserSA	UserNP	UserQC	All
Network	assortativity_coef	APE	92.2	91.8	**89.1**	92.8
	avg_clustering	APE	24.2	37.9	6.9	**6.8**
	comm_modularity	APE	75.6	76.0	**73.0**	74.0
	degree_dist	JSD	0.5	**0.4**	**0.4**	**0.4**
	density	APE	396.5	347.5	**195.7**	205.7
	max_node_degree	APE	**42.5**	61.3	46.5	42.3
	mean_node_degree	APE	142.2	130.3	**84.8**	87.7
	mean_shortest_path	APE	31.1	30.5	**24.6**	26.1
	num_connected_comp	APE	100.0	100.0	100.0	100.0
	num_edges	APE	**19.6**	**19.6**	20.8	20.7
	num_nodes	APE	50.6	48.1	**33.1**	33.8
User	activity_dist	JSD	**0.2**	0.4	0.4	0.3
	disp_gini_quote	APE	155.5	167.2	**153.2**	157.7
	disp_gini_reply/comment	APE	376.9	409.4	**300.0**	344.6
	disp_gini_retweet	APE	**0.7**	6.8	2.5	3.7
	gini_coef	APE	25.1	29.5	**18.5**	19.6
	most_active	RBO	**0.8**	0.4	0.3	0.5
	palma_coef	APE	49.1	61.2	33.3	**32.9**
	popularity	RBO	0.2	0.2	0.2	0.2
	unique_content	RMSE	54.5	70.2	66.3	**51.2**

Table 1. Predictive Performance Using Different Feature Types to Model User Behavior

Categ.	Measurement	Metric	Group SA	Group NP	Both
Network	assortativity_coef	APE	92.8	92.6	**91.7**
	avg_clustering	APE	8.0	**5.9**	16.4
	comm_modularity	APE	74.2	**73.6**	75.2
	degree_dist	JSD	0.4	0.4	0.4
	density	APE	213.0	**203.2**	261.8
	max_node_degree	APE	**43.8**	47.3	45.7
	mean_node_degree	APE	90.1	**86.9**	105.4
	mean_shortest_path	APE	26.4	**25.6**	27.7
	num_connected_comp	APE	100.0	100.0	100.0
	num_edges	APE	**20.0**	20.8	20.5
	num_nodes	APE	34.9	**33.4**	40.0
User	activity_dist	JSD	0.3	0.3	0.3
	disp_gini_quote	APE	156.2	**152.5**	157.7
	disp_gini_reply/comment	APE	**217.6**	261.9	344.6
	disp_gini_retweet	APE	**2.3**	2.6	3.7
	gini_coef	APE	20.8	**19.0**	19.6
	most_active	RBO	**0.6**	**0.6**	0.5
	palma_coef	APE	39.1	**31.6**	32.9
	popularity	RBO	0.2	0.2	0.2
	unique_content	RMSE	52.7	**50.3**	51.2

Table 2. Predictive Performance Using Different Feature Types for Groups

Conclusion

As digital collaboration and social media proliferate in recent years, organizations increasingly use social media networks (SMNs) to build relationship and form network ties. However, there is a lack of understanding of behavioral drivers of SMN participants. This research develops an IS development framework for modeling user behavior in large evolving SMNs. Based on the framework, a temporal network model (TNM) was developed and used to predict and simulate user activities in large evolving SMNs, with a goal to enhance understanding of user behavior to support various strategic decision making. A simulation study of Twitter SMN shows that the framework and TNM are useful in predicting and simulating user behavior. Comparison with ground-truth data demonstrates the usefulness of various types of features and predictive techniques in SMN simulation. Preliminary results provide insights of the importance of comprehensive network features to model SMN group behavior accurately and quality of commitment features to model SMN user behavior. The research contributes to building a theoretically-driven framework for modeling SMN behavior. We are testing the use of another social network platform (Reddit) that is expected to provide more robust and interesting findings. Future directions may include applying TNM and the feature sets to different SMNs to compare cross-platform performance, integrating other ML techniques in various learning and prediction tasks, using additional benchmark models in evaluation, and simulating cascades and information diffusion in online environment.

References

Bacaksizlar Turbic, N. G., and Galesic, M. 2023. "Group Threat, Political Extremity, and Collective Dynamics in Online Discussions," *Scientific Reports* (13:1), p. 2206.

Becker, G. S. 1974. "A Theory of Social Interactions," *Journal of Political Economy* (82:6), pp. 1063-1093.

Chung, W., Rao, B., and Wang, L. 2019. "Interaction Models for Detecting Nodal Activities in Temporal Social Media Networks," *ACM Transactions on Management Information Systems* (10:4).

Chung, W., and Zeng, D. 2016. "Social-Media-Based Public Policy Informatics: Sentiment and Network Analyses of U.S. Immigration and Border Security," *Journal of the Association for Information Science and Technology* (67:7), pp. 1588-1606.

Cichy, P., Salge, T. O., and Kohli, R. 2021. "Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars," *MIS Quarterly* (45:4), pp. 1863-1891.

Cui, R., Gallino, S., Moreno, A., and Zhang, D. J. 2018. "The Operational Value of Social Media Information," *Production and Operations Management* (27:10), pp. 1749-1769.

Debolina, D. 2018. "Social Media and Technology Trends in Hrm: Cases in Recruitment and Talent Management," in *Social Media and Journalism,* V. Ján and R. Jana (eds.). Rijeka: IntechOpen, p. Ch. 6.

Du, N., Jia, X. W., Gao, J., Gopalakrishnan, V., and Zhang, A. D. 2015. "Tracking Temporal Community Strength in Dynamic Networks," *IEEE Transactions on Knowledge and Data Engineering* (27:11), pp. 3125-3137.

Ebrahimi, M., Chai, Y. D., Samtani, S., and Chen, H. 2022. "Cross-Lingual Cybersecurity Analytics in the International Dark Web with Adversarial Deep Representation Learning," *MIS Quarterly* (46:2), pp. 1209-1226.

Emerson, R. M. 1976. "Social Exchange Theory," *Annual Review of Sociology* (2:1), pp. 335-362.

Endsley, M. R. 2012. *Designing for Situation Awareness : An Approach to User-Centered Design*, (2nd ed.). Boca Raton, FL: CRC Press.

Festinger, L. 1954. "A Theory of Social Comparison Processes," *Human Relations* (7:2), pp. 117-140.

Gino, F., Ayal, S., and Ariely, D. 2009. "Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel," *Psychological Science* (20:3), pp. 393-398.

Heider, F. 1958. *The Psychology of Interpersonal Relations*. Hillside, New Jersey: Erlbaum.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *Management Information Systems Quarterly* (28:1), pp. 75-105.

Holme, P., and Saramäki, J. 2012. "Temporal Networks," *Physics Reports* (519:3), pp. 97-125.

Hu, Y. H., Xu, A. B., Hong, Y. L., Gal, D., Sinha, V., and Akkiraju, R. 2019. "Generating Business Intelligence through Social Media Analytics: Measuring Brand Personality with Consumer-, Employee-, and Firm-Generated Content," *Journal of Management Information Systems* (36:3), pp. 893-930.

Jackson, M. O. 2008. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Kane, G. C., Alavi, M., Labianca, G. J., and Borgatti, S. P. 2014. "What's Different About Social Media Networks? A Framework and Research Agenda," *MIS Quarterly* (38:1), pp. 275-304.

Kong, Y.-X., Wu, R.-J., Zhang, Y.-C., and Shi, G.-Y. 2023. "Utilizing Statistical Physics and Machine Learning to Discover Collective Behavior on Temporal Social Networks," *Information Processing & Management* (60:2), p. 103190.

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., and Brown, D. E. 2019. "Text Classification Algorithms: A Survey," *Information* (10:150).

Le Bon, G. 1895. *The Crowd: A Study of the Popular Mind*. New York, NY: The MacMillan Co.

Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. 2009. "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters," *Internet Mathematics* (6:1), pp. 29-123.

MacMillan, D., and Torbati, Y. 2021. "How the Rich Got Richer: Reddit Trading Frenzy Benefited Wall Street Elite," in: *The Washington Post*. https://www.washingtonpost.com/business/2021/02/08/gamestop-wallstreet-wealth/.

Nikkhah, H. R., and Grover, V. 2022. "An Empirical Investigation of Company Response to Data Breaches," *MIS Quarterly* (46:4), pp. 2163-2196.

Porteous, C. 2021. "97% of Fortune 500 Companies Rely on Social Media. Here's How You Should Use It for Maximum Impact," Entrepreneur.com.
Quan-Haase, A. 2015. *Technology & Society : Social Networks, Power, and Inequality*, (Second edition. ed.). Don Mills, Ontario, Canada: Oxford University Press.
Quinton, S., and Wilson, D. 2016. "Tensions and Ties in Social Media Networks: Towards a Model of Understanding Business Relationship Development and Business Performance Enhancement through the Use of Linkedin," *Industrial Marketing Management* (54), pp. 15-24.
Reicher, S. 2001. "The Psychology of Crowd Dynamics," in *Blackwell Handbook of Social Psychology: Group Processes,* M.A. Hogg and R.S. Tindale (eds.). Malden, MA: Blackwell Publishers, pp. 182-208.
Samtani, S., Chai, Y. D., and Chen, H. C. 2022. "Linking Exploits from the Dark Web to Known Vulnerabilities for Proactive Cyber Threat Intelligence: An Attention-Based Deep Structured Semantic Model," *MIS Quarterly* (46:2), pp. 911-946.
Scellato, S., Noulas, A., and Mascolo, C. 2011. "Exploiting Place Features in Link Prediction on Location-Based Social Networks," in: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. San Diego, California, USA: Association for Computing Machinery, pp. 1046–1054.
Shrestha, P., Sathanur, A., Maharjan, S., Saldanha, E., Arendt, D., and Volkova, S. 2020. "Multiple Social Platforms Reveal Actionable Signals for Software Vulnerability Awareness: A Study of Github, Twitter and Reddit," *PLOS ONE* (15:3), p. e0230250.
Shumway, R. H., and Stoffer, D. S. 2017. *Time Series Analysis and Its Applications : With R Examples*, (4th ed.). Cham: Springer International Publishing : Imprint: Springer,.
Snijders, T. A. B., van de Bunt, G. G., and Steglich, C. E. G. 2010. "Introduction to Stochastic Actor-Based Models for Network Dynamics," *Social Networks* (32:1), pp. 44-60.
Stage, C. 2013. "The Online Crowd: A Contradiction in Terms? On the Potentials of Gustave Le Bon's Crowd Psychology in an Analysis of Affective Blogging," *Distinktion: Journal of Social Theory* (14:2), pp. 211-226.
Turner, R. H., and Killian, L. M. 1957. *Collective Behavior*. Englewood Cliffs, N.J.,: Prentice-Hall.
Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. New York, NY: Springer-Verlag.
Wang, P., Xu, B. W., Wu, Y. R., and Zhou, X. Y. 2015. "Link Prediction in Social Networks: The State-of-the-Art," *Science China-Information Sciences* (58:1).
Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.
Wei, W., and Carley, K. M. 2015. "Measuring Temporal Patterns in Dynamic Social Networks," *ACM Transantions on Knowledge Discovery from Data* (10:1), pp. 1-27.
Wheeler, L. 1966. "Toward a Theory of Behavioral Contagion," *Psychological Review* (73:2), pp. 179-&.
Wu, L., Zhang, Q., Chen, C., Guo, K., and Wang, D. 2020. "Deep Learning Techniques for Community Detection in Social Networks," *IEEE Access* (8), pp. 96016-96026.
Xie, J., Liu, X., Zeng, D. D., and Fang, X. 2022. "Understanding Medication Nonadherence from Social Media: A Sentiment-Enriched Deep Learning Approach," *Management Information Systems Quarterly* (46:1), pp. 341-372.
Yasami, Y., and Safaei, F. 2017. "A Statistical Infinite Feature Cascade-Based Approach to Anomaly Detection for Dynamic Social Networks," *Computer Communications* (100), pp. 52-64.
Yu, Y., Si, X., Hu, C., and Zhang, J. 2019. "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Computation* (31:7), pp. 1235-1270.
Zeng, D. 2015. "Policy Informatics for Smart Policy-Making," *IEEE Intelligent Systems* (30:6), pp. 2-3.

Acknowledgements

We thank the conference co-chairs and reviewers for providing valuable comments and advices.