SCANNING THE ISSUE

The Digital Age and the Future of Social Network Science and Engineering

By JESSICA C. FLACK
Guest Editor

RAISSA M. D'SOUZA
Guest Editor

This special issue of the Proceedings of the IEEE focuses on how digital technology is changing the structure and dynamics of social networks and the tools we have for studying and designing them. As a testament to the importance of this topic consider that as of 2014, 72% of Internet users in the United States and 64% of users worldwide use social media [1], [2]. Facebook alone has 1.15 billion users, up from 1 million in 2004, just ten years ago [2]. In the United States, the average user spends more than one fourth of every online hour on social media, and almost 50% of Americans say that Facebook is their #1 influencer of purchases [2]. Google+ has been around for only three years and already there are 1 billion Google+ enabled accounts [2].

This special issue focuses on how digital technology is changing the structure and dynamics of social networks and the tools we have for studying and designing them.

This rapid growth in connectivity stands in stark contrast to the vast majority of human history in which social networks were small and geographically localized, institutions changed slowly, and power and influence were concentrated in a small subset of the population [3]. The rise of the world wide web—and particularly the invention of powerful search technologies, social media platforms, and novel file sharing technologies—has led to an explosive growth in social network size and connectivity as well as the development of new kinds of reputation-based barter systems (e.g., reviewed in [4] and [5]) and underground economies (e.g., [6]). Standard geographic definitions of population compete for causal relevance with definitions that group people together based on behavioral criteria, including hyperlinking on the web [7]. It is now possible to track events as they unfold in real time, to coordinate relatively rapidly over long space scales (for a potential example, see [8]), and to access semiglobal and global information to make decisions. This means that individuals and subgroups without much power in the traditional sense potentially can serve as instruments of large-scale social change. The growth of social networks on digital media also means it is possible to collect big, reliable data sets on human behavior and associated events such as earthquakes [9], as human behavior on digital media leaves a relatively easily harvestable data trail [10].

These technological changes, the new social structures they seem to be producing, and the data being generated permit reconstructing the behavior of individuals from their digital footprints (e.g., [11]; see also the paper by Coviello *et al.* of this special issue [12]). This will enable empirical study of *how* different kinds of social structures arise from interactions among individuals at the microscopic level and *what* the implications are of alternative social structures for social stability and contagion processes. As progress is made on this front, large-scale intervention into human

Digital Object Identifier: 10.1109/JPROC.2014.2368790

behavior and demographics will no longer be restricted to coarse manipulation of environmental variables or interventions based on qualitative understanding of how social systems work (as an example, consider the so-called causal loop diagram produced to describe the counter insurgency dynamics in Afghanistan [13]).

Precision, quantitatively justified interventions into behavioral dynamics will become increasingly feasible. Already such interventions are being employed; the Facebook "61-million-person experiment" in which users' affective states were manipulated [14] serves as one example. These interventions will allow the interveners to influence individual decision-making and permit modulation and possibly control of macroscopic properties of social systems; currently, for example, large, digital data sets on online social behavior are being harvested to inform the design of social media platforms (e.g., [15]). Although these large-scale experiments on human behavior will initially occur in the digital world, they will eventually inform the mechanism design of real-world infrastructure to optimize communication and better meet basic human needs such as access to food, sanitation, healthcare, and schooling (e.g., [16]).

Hence social media, search, and data extraction technologies are not only changing the structure and dynamics of social networks but are also potentially changing how *controllable* these systems are. The purpose of this special issue is to review what is currently known about how these technologies are changing social networks and what the consequences will be for human social dynamics. Understanding the impact of changing technology on social networks is relevant to the readers of the PROCEEDINGS OF THE IEEE for several reasons. One reason is that it is engineers, mathematicians, and computer scientists who are largely responsible for developing the social media technology, file sharing protocols, and pattern extraction algorithms that are producing changes to social networks. An understanding of cognitive principles and behavioral interaction rules underlying social network formation can facilitate the development of social media applications that individuals are more likely to use.

A second reason is that an understanding of cognitive and behavioral constraints could improve the performance of machine learning and other pattern detection and data extraction algorithms by providing principled means to restrict the search space.

Finally, with the rise of online social networks information becomes a concept as central as energy to informing design of infrastructure. Societies are composed of multiple, overlapping social networks. This is true for networks developing in "real" space as well as on digital media. The local connections of a node in these multiple, overlapping social networks can be thought of as that node's social niche [17], with the edges in each of the networks representing different kinds of relationships, for example, friends, co-workers, mentors, and students. Whereas the ecological niche [18] (the term "niche construction" was invented in ecology; see [19]) is composed of resource vectors (availability of wood for building dams, prey items, and so on), an individual's social niche is composed of its vector of behavioral relationships in the set of overlapping social networks in which it participates [17]. These behavioral relationships provide critical information that facilitates resource extraction, and the construction of these networks by users can be thought of as a form of social niche construction [17], [20].

To the extent this conceptualization is correct, social networks—perhaps more than other types of networks—are about the flow of information. This means that higher order features of network structure (that is, minimally, connections beyond a node's direct connections) may be of critical impotence to understanding the functional consequences of a given network for its nodes, both individually and collectively. This observation is increasingly valid as social networks move online and become further divorced from direct resource extraction. In addition, issues of sociotechnical congruence—the match or mismatch between social networks and the technological networks they build and rely upon—have implications for information flow that need to be considered when assessing how information impacts function and energy extraction (e.g., [21]). As Weaver and Shannon [22] noted more than 60 years ago, although we have a good theory of information defined as uncertainty, we have no theory relating information to function. The migration of social interactions into the digital domain means that the absence of semantic and functional theories of information is no longer a curious omission but a central challenge for theory and analysis.

In the first paper, Bettencourt suggests that a property of social, informational networks is superlinear scaling, which translates into increasing returns to scale. Bettencourt proposes that network dynamics in these kinds of complex social systems need to be understood in relation to evolutionary learning and inference principles. He develops a conceptual framework to explain the transition from relatively static, homogenous, information-poor networks to information-rich, diverse, and highly interconnected ones using current understanding of social scaling in urban environments to develop his argument, and then extending it to the digital social domain using the Internet and Wikipedia as case studies.

Networks typically have organization at multiple levels of scale, from correlations between individual nodes to larger scale modular structures and groupings into communities of nodes with similar characteristics. Understanding how to reliably detect these structures remains a central activity in the study of networks, especially as the goal is to connect structural features with the function of a network. This task is increasingly challenging as we comprehend the extent to which multiple communities can overlap in nontrivial ways. In the second paper, Yang and Leskovec tackle this issue by considering each community as a "tile" which can overlap extensively with other tiles. Their methodology decomposes the network into a combination of overlapping, nonoverlapping, and hierarchically organized communities. In contrast to previous work, they show that nodes residing in the overlap between communities are more densely connected than those in nonoverlapping regions. Moreover, they show that overlap between multiple communities identifies dense network cores, revealing the core–periphery structure in networks and thus offering a method to unify the study of network modularity and core–periphery structure.

Singh *et al.* point out, in the third paper, that the line between real or physical social networks and those on digital media is not as strict as many assume. The authors observe that fine-grained behavioral data on individuals collected from online sources and smartphones are now being merged to give an increasingly comprehensive picture of an individual's location, social ties, actions, and context in both settings. These merged data can be used to influence the actions and beliefs of individuals in both settings, largely because smartphones function as a bridge between worlds. This bridge allows individuals to behave simultaneously in both the digital and physical domains, createing real-time couplings between their social behavior in both worlds.

The question of how coupling occurs between social behavior in the online and physical worlds can be reframed as a question of synchrony and contagion: How does behavior spread over social networks through contagion and other mechanisms and when do contagion dynamics lead to correlation in social preferences and decision making and even behavioral synchrony? The large quantity of quantitative data on behavior and contagion available from social media sites means that this question can be addressed at a variety of scales, and it means that, in principle, the causal mechanisms and factors that influence spread and synchrony can be identified.

There are two broad classes of approaches that can be used to quantify causal relationships. One is experimental intervention. This can include perturbing the social media platform itself (e.g., how individuals are allowed to make posts, how long those posts can be, etc.), perturbing the social media environment (e.g., changing the aesthetics of the site, site branding, types of advertising to which users are exposed, etc.), and injecting information or behavior into the system (e.g., creating users and posts), and tracking its effects using time-series analysis.

The second approach for quantifying casual relationships is observational with statistical "interventions" (e.g., Pearl's "do" operator [23]). Although it is relatively easy to perform large-scale experiments using social media data, there are, as the Facebook controversy discussed above illustrates, many ethical issues that need to be carefully considered before proceeding. Developing methods for performing "statistical" interventions on observational data is consequently of utmost importance. The fourth paper by Coviello *et al.* develops a method for quantifying influence of users on one another. The method combines geographic aggregation and instrumental variables regression to measure the effect of an exogenous variable on an individual's expression and the influence of this change on the expression of others to whom that individual is socially connected. The authors demonstrate the power of the approach in the context of emotional contagion of semantic expression but also show that the approach is quite general and can be applied to many kinds of data collected from a variety of social networking platforms.

These ideas are reviewed in a broader, technical context by Holme, in the fifth paper. Holme highlights that the data available from social media sites and online social networks typically have time stamps. These time stamps allow analysis of *how* the timescales on which contacts and interactions occur can influence contagion dynamics and network construction. Holme reviews methods for analyzing temporal networks. He addresses challenges in representing temporal interactions in networks and in the construction of appropriate null models of temporally evolving network structure. He also discusses principled means for simplifying temporal networks to make them amenable to rigorous analysis.

As discussed above, a growing proportion of human activity is leaving behind digital footprints. In the sixth paper, Lamboitte and Kosinski show how pervasive details left in these footprints can be used to infer an individual's personality, where personality refers to the major psychological framework identifying individual differences among people in behavior patterns, cognition, and emotion. In addition to reviewing the factors of personality, they review a range of recent works focused on predicting personality from digital traces. They conclude with discussion of the increasingly important implications for privacy, security, and data ownership as well as future directions for research.

In the final paper, Walker and Muchnik delve into how our traditional methods of experimental design, which divide subjects into control and treatment groups, need to be radically redesigned for our digital world. Traditional methods neglect the natural connections that exist between individuals, and, more so, that the impact of treatment can propagate through such connections. In analogy to big data, they survey the burgeoning movement of big experiments (such as the 61-million-person experiment on Facebook [14]). They discuss the broad range of aspects that need to be considered for appropriate design of network-randomized trials, including the experimental setting, the process under study, and the impact of connectivity. Equally important is their discussion on emerging methods to draw statistically meaningful conclusions from such experiments as well as developing novel treatment schemes that leverage connections in social networks. They repeatedly highlight the important and subtle distinctions between offline and online experimental settings.

In conclusion, this is an exciting moment in human history as the digital world and the physical world become increasingly intertwined in a seamless manner. We offer three take-home messages.

- Social media, search, and data extraction technologies are not only changing the structure and dynamics of social networks, but are also changing how controllable these systems are.
- Precision, quantitatively justified interventions into behavioral dynamics are increasingly feasible within the digital domain, permitting large-scale experiments on human behavior and social systems. This is useful and presents challenges.
- We understand the relationship between energy and information—how bits get converted to watts—for electrical circuits, but not for social networks. In biology, computational social science, and the science of social engineering, the development of a functional theory of information is a central theoretical challenge that needs to be addressed if these disciplines are to have strong foundations. ■

Acknowledgment

The guest editors would like to thank D. Krakauer and G. Spencer for comments and ideas for the content and structure of this special issue. The work of J. C. Flack was supported by a grant from the John Templeton Foundation for the study of complexity. In addition, the work of J. C. Flack and R. M. D'Souza was supported by the U.S. Army Research Office Multidisciplinary University Research Initiative (MURI) Award under Contract W911NF-13-1-0340.

REFERENCES

[1] Statista, "Number of social network users worldwide from 2010 to 2017 (in billions)." [Online]. Available: http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

[2] J. Bullas, "22 social media facts and statistics you should know in 2014." [Online]. Available: http://www.jeffbullas.com/2014/01/17/20-social-media-facts-and-statistics-you-should-know-in-2014/

[3] J. M. Diamond, *Guns, Germs, Steel: The Fates of Human Society*. London, U.K.: Norton, 1999.

[4] A. Josang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online provision," *Decision Support Syst.*, vol. 43, pp. 618–644, 2007.

[5] P. Resnick, K. Kuwabara, R. Zeckhauer, and E. Friedman, "Reputation systems," *Commun. ACM*, vol. 43, pp. 45–48, 2000.

[6] A. Greenberg, "Drug market 'agora' replaces the silk road as king of the dark net," *Wired*, Sep. 2, 2014. [Online]. Available: http://www.wired.com/2014/09/agora-bigger-than-silk-road/

[7] D. Gibson, J. Kleinberg, and P. Raghavan, "Inferring web communities from link topology," in *Proc. 9th ACM Conf. Hypertext Hypermedia: Links, Objects, Time and Space—Structure in Hypermedia Systems*, 1998, pp. 225–234.

[8] W. Ghonim, *Revolution 2.0: The Power of the People Is Greater Than the People in Power*. New York, NY, USA: Houghton Mifflin Harcourt, 2012.

[9] L. Burks, M. Miller, and R. Zadeh, "Rapid estimate of ground shaking intensity by combining simple earthquake characteristics with tweets," presented at the 10th U.S. Nat. Conf. Earthquake Eng., Front. Earthquake Eng., Anchorage, AK, USA, Jul. 21–25, 2014, 10NCEE.

[10] Edge, "Reinventing society in the wake of big data: A conversation with Alex (Sandy) Pentland," Aug. 30, 2012. [Online]. Available: http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data

[11] N. Eagle, A. Pentland, and D. Lazer, "Inferring social network structure using mobile phone data," *Proc. Nat. Acad. Sci.*, vol. 106, no. 36, pp. 15 274–15 278, 2009.

[12] L. Coviello, J. H. Fowler, and M. Franceschetti, "Words on the web: Noninvasive detection of emotional contagion in online social networks," *Proc. IEEE*, vol. 102, no. 12, Dec. 2014, DOI: 10.1109/JPROC.2014.2366052.

[13] PA Consulting Group, "Dynamic planning for counterinsurgency in Afghanistan, PowerPoint presentation," 2009. [Online]. Available: http://msnbcmedia.msn.com/i/MSNBC/Components/Photo/_new/Afghani-stan_Dynamic_Planning.pdf

[14] R. M. Bond*et al.* "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, pp. 295–298, 2014.

[15] V. K. Singh, R. Jain, and M. Kankanhalli, Mechanism design for incentivizing social media contributions," in *Social Media and Modeling*, Hoi*et al.*, Ed. London, U.K.: Springer-Verlag, 2011.

[16] A. Wesolowski*et al.* "Quantifying the impact of human mobility on malaria," *Science*, vol. 338, pp. 267–2702014.

[17] J. C. Flack, M. Girvan, F. B. M. de Waal, and D. C. Krakauer, "Policing stabilizes construction of social niches in primates," *Nature*, vol. 439, pp. 426–429, 2006, see supplementary information for social niche definition.

[18] G. E. Hutchinson, "Concluding remarks," *Cold Spring Harbor Symp. Quantitative Biol.*, vol. 22, pp. 415–427, 1957.

[19] J. F. Odling-Smee, K. N. Laland, and M. W. Feldman, *Niche Construction: The Neglected Process in Evolution. Monographs in Population Biology 37*. Princeton, NJ, USA: Princeton Univ. Press, 2003.

[20] J. C. Flack, "Multiple timescales and the developmental dynamics of social systems," *Phil. Trans. Roy. Soc. B 5*, vol. 367, pp. 1802–1810, 2012.

[21] D. Posnett, R. M. D'Souza, P. Devanbu, and V. Filkov, "Dual ecological measures of focus in software development," in *Proc. 35th Int. Conf. Softw. Eng.*, 2013, pp. 452–461.

[22] W. Weaver and C. E. Shannon, *The Mathematical Theory of Communication*. Urbana, IL, USA: Univ. Illinois Press, 1949.

[23] J. Pearl, *Causality: Models, Reasoning, Inference*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

ABOUT THE GUEST EDITORS

Jessica C. Flack received the B.A. degree with honors from Cornell University, Ithaca, NY, USA, studying anthropology, evolutionary theory, and biology, and the Ph.D. degree from Emory University, Atlanta, GA, USA, studying animal behavior, cognitive science, and evolutionary theory.

She is a Codirector of the Center for Complexity and Collective Computation (C4), Wisconsin Institute for Discovery, University of Wisconsin at Madison, Madison, WI, USA and External Professor at the Santa Fe Institute, Santa Fe, NM, USA. After her Ph.D., for the next eight years, she was in residence at the Santa Fe Institute, first as a Postdoctoral Fellow and then as Professor. She moved to the University of Wisconsin at Madison in 2011 to build C4. Her research focuses on coarse graining and collective computation in nature and their role in the evolution and development of new levels of biological and social organization. The approach is highly empirical, beginning with rich microscopic data, including network time series, gained from model systems in neuroscience, animal behavior, and human social systems. The goal is to find common algorithmic principles underlying the emergence of novel, functionally significant spatial and temporal scales, and ultimately new kinds of collectives and individuality. This research has involved development of computational techniques (*Inductive Game Theory*) for extracting strategies (and potentially game structure and payoffs) from time-series data and constructing stochastic strategic circuits that map input to the target macroscopic output. Dimension reduction techniques are used to identify the most appropriate minimal model for representing the compressed circuits. One can think of this approach as a *strategic statistical mechanics*, embedding complex decision rules in formalisms for connecting microscopic configurations to macroscopic averages. An ultimate goal is to determine whether living systems—composed of noisy, adaptive, heterogeneous components often with multiple goal states—are governed by laws derived from simple microscopic processes or reflect contingent events leading to irreducible complexity. Her work has appeared in journals such as *Nature*, *Proceedings of the National Academy of Sciences of the United States of America*, *Royal Society Interface*, and *Chaos*, and has been covered by scientists and journalists in popular press outlets, including *Science*, *Nature*, *The Economist*, *Current Biology*, and the *BBC*.

Raissa M. D'Souza received the Ph.D. degree in statistical physics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1999.

She is a Professor of Computer Science and of Mechanical Engineering at the University of California Davis, Davis, CA, USA, as well as an External Professor at the Santa Fe Institute, Santa Fe, NM, USA. She was a Postdoctoral Fellow, first in Fundamental Mathematics and Theoretical Physics at Bell Laboratories, and then in the Theory Group at Microsoft Research. Her interdisciplinary work on network theory spans the fields of statistical physics, theoretical computer science, and applied math, and has appeared in journals such as *Science*, *Proceedings of the National Academy of Sciences of the United States of America*, and *Physical Review Letters*. Her research has been funded by the National Science Foundation (NSF), the National Academy of Sciences of the USA, the U.S. Department of Defense (DOD), the U.S. Army Research Laboratory (ARL), and the U.S. Air Force Office of Scientific Research (AFOSR). Her research focuses on mathematical models of complex systems and networks, in particular on self-organization, phase transitions, and the structure and function of networked systems. Interdependence between collections of networks and sociotechnical congruence are additional themes underlying her research.

Prof. D'Souza has been selected a Kavli Fellow of the National Academy of Sciences four times and is a member of the Science Steering Committee of the Santa Fe Institute. She has served on the program committees for numerous conferences, including serving as the General Chair of NetSci 2014, and currently serves on the editorial board of numerous international mathematics, physics, and engineering journals. She is also a member of the World Economic Forum's Global Agenda Council on Complex Systems.