

Studying Social Network Sites via Computational Methods

Tasos Spiliotopoulos

Madeira-ITI / Dept. of Mathematics and Engineering
University of Madeira
Funchal, Portugal

Abstract— Social Network Sites (SNSs) are fundamentally changing the way humans connect, communicate, and relate to one another and have attracted a considerable amount of research attention. In general, two distinct research approaches have been followed in the pursuit of results in this research area. Firstly, established traditional social science methods, such as surveys and interviews, have been extensively used for inquiry-based research on SNSs. More recently, however, the advent of Application Programming Interfaces (APIs) has enabled data-centric approaches that have culminated in theory-free “big data” studies. Both of these approaches have advantages, disadvantages and limitations that need to be considered in SNS studies. This PhD work proposes that a suitable combination of these two approaches can help understand the limitations and address the shortcomings that researchers are faced with when following each approach separately. In order to illustrate the practicability and value of this combination of approaches, I propose to employ usage and network data collected via an API to complement survey metrics in two SNS studies. The first study examines the motivations for using Facebook and Twitter, in order to enhance our understanding of how and why people use these services, while the second study focuses on aspects of the interpersonal relationships on Facebook, such as tie strength, trust and information disclosure.

Keywords— *computational social science; social network sites; computer-mediated communication; Facebook; Twitter; uses and gratifications; tie strength; network structure*

I. INTRODUCTION

Social Network Sites (SNSs), such as Facebook, Twitter, LinkedIn and Pinterest, are fundamentally changing the way humans connect, communicate and relate to one another. Gaining in popularity the past few years, SNSs also exhibit high diffusion and an increasing number of features. For instance, Facebook, which currently holds a prime position among SNSs, has a continuously evolving feature set and as of December 2014 has 890 million daily active users on average, approximately 82.4% of whom are from outside the U.S. and Canada, and 745 million of whom access the site via mobile devices [1]. Similarly, Twitter has reached 288 million monthly active users creating 500 million Tweets each day in just eight years of service [2]. First defined as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view

and traverse their list of connections and those made by others within the system” [3], SNSs have evolved and expanded to include complementary features and services, such as messaging among users, organization into groups and events, and more nuanced ways of sharing information (e.g., by creating lists of recipients or by more sophisticated privacy mechanisms). In some cases, bottom-up user proposals have become conventions in SNSs, such as hashtags in Twitter.

Understanding the dynamics of interaction on SNSs is a complex and challenging research topic with the ultimate goal of elucidating the impact of the social web on society. In the case of Facebook, for example, studies have examined social capital [4, 5], personality [6], tie strength [7, 8], user motivations [9, 10], site adoption [11], participation [12], information seeking [13], user privacy [9, 14], network structure [15, 16], and the use of specific Facebook features [10]. Similarly, studies on Twitter have examined topics such as tie formation [17], retweeting [18], influence [19, 20], hashtags [21, 22], communities [23, 24], events [22, 25], and user motivations [26]. At the same time, researchers have found striking similarities between the social structures in offline social networks and SNSs [27] and have suggested that online social networks can represent effective proxies for hard-to-establish real world friendship networks [16].

Although this research is diverse in its focus, it shares strong commonalities in the research methods that are used and the factors that impact their validity. In general, two distinct research approaches have been followed in the pursuit of results in this line of research. On one hand, established inquiry-based research methods, such as surveys and interviews, have graduated from traditional social science to the study of SNSs. On the other hand, the advent of Application Programming Interfaces (APIs) has enabled data-centric approaches that have culminated in theory-free “big data” studies. However, both approaches have advantages, disadvantages, and limitations.

This PhD work proposes that a suitable combination of these two approaches can help understand the limitations and address the shortcomings that researchers are faced with when following each approach separately. In order to illustrate the practicability, potential and value of this combination of approaches, we need to conduct studies that utilize both kinds of methods and data and show how these provide useful

insights for SNS researchers and designers. To this end, we propose to employ usage and network data collected via an API to complement survey metrics in two two-part SNS studies as described in later sections.

II. CONTEXT AND MOTIVATION

A. Challenges in Traditional Social Science Research

Until recently, research on SNSs has borrowed methods from social network research studying offline social networks. This traditional line of research on the human interactions that constitute social networks has relied mainly on one-time, self-reported data on relationships, something that has restricted the size, scope, and ultimately, the value of the insights from SNS studies [28, 29, 30]. These major challenges in the study of SNSs are explained in detail below.

First, the majority of social network research, and SNS research in particular, so far has been obliged, for practical reasons, to limit the overall quantity of ties that each respondent can report or to limit the number of respondents [31, 32]. More particularly, traditional network research generally involves gathering a number of participants and querying them about their ties (e.g., via the use of a name generator). Primarily due to time constraints, the participants can only report on a limited number of ties or interactions. Furthermore, the collection of data from large, meaningful groups of people (e.g., citizens of a particular city, fans of a particular person, proponents of a particular idea, or just members of any large arbitrary group) has been time-consuming and laborious merely due to the size of the potential sample. However, even the absence of a small number of key ties is reported to have the potential to substantially alter the profile of a network [32, 33, 34]. In turn, this has typically confined network analysis to either small, but complete, groups such as a work group or sports team, or samples of incomplete personal networks derived from an individual's report on *perceived* ties with friends and family [31]. For example, Wellman and colleagues [35] report on a series of seminal studies that took place in the 1960s and '70s, the East York studies, that could either afford to survey a large random sample of 845 people and identify six ties at most (in the first study), or perform in-depth interviews to get a more comprehensive picture of the personal social networks of 33 participants (in the second study).

Second, much past research on SNSs has drawn upon a very small portion of the wealth of available data [30] – SNSs are complex services capable of supporting a range of different interactions and relationships. While exceptions exist where more than one aspect of an SNS have been studied together via different modes, such as using survey methods and the Facebook API to study the association between socioeconomic status and network structure [15] or using longitudinal surveys matched to Facebook server logs to study social capital [4], the overall picture shows that the majority of SNS studies typically fail to harness information that could be relevant and useful. For example, many Facebook studies have relied exclusively on survey methods and avoided the site altogether [e.g., 14], others studied solely the network structure of Facebook [e.g., 19, 27], and others focused only on profile data, ignoring the network ties between users [e.g., 9, 36].

Third, scholars in these studies have struggled with issues such as recall bias [37], interviewer effects [38], and other sources of measurement error that may accompany survey research (see [30] for discussion). Interestingly, even the most extensive and detailed survey research, such as the U.S. General Social Surveys (GSS) whose results are freely available and widely used in sociological research, have been found to suffer from systematic biases linked to interviewers [38].

B. Challenges in Data-Centric Research

Researchers have highlighted the potential to alleviate these sampling and data collection problems by capturing and analyzing large quantities of online SNS data [29]. Compared to the methods and data available to traditional social scientists, online information can be accessed and analyzed computationally in ways that are both efficient and accurate [16, 29]. For example, in the case of Facebook, a rich, robust Application Programming Interface (API) allows researchers to collect large volumes of data relating to issues such as site feature use and personal network structure with unprecedented accuracy, granularity, and reliability. Similarly, the Twitter API allows the mass collection of information about tweets, their posters and the relations among them. In addition, new tools for efficiently handling these large volumes of data are constantly being developed and updated [e.g., 16, 39], focusing on central aspects of SNSs, such as social network analysis, text analysis, and visualization.

Data-centric approaches may use data generated specifically from a service, but in the case of SNSs we are mainly interested in “found data”, the digital exhaust of people's use of the service. They are easy to collect via an API, and are free from the types of biases found in traditional social science approaches. These “found data”, however, are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis [40] and, as such, are subject to their own problems and limitations. Overall, they are good for providing a more general view of what happens, but they cannot answer to “why”. This type of research relies in exposing statistical patterns in the data and is useful for finding correlation rather than causation. Often there are so many variables in the data, that even if controlled experiments were possible they could still not determine causality because of possible confounding factors. In fact, in many cases causality may not be possible to determine without the inquiry methods or the theory behind traditional statistics.

For example, recent research identifies two big “traps” in big-data analysis [40]. First, *big data hubris*, described as the “often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis” [40], has led to studies that have possibly overlooked considerable information that could be extracted by traditional social science methods. Second, *algorithm dynamics*, referring to “changes made by engineers to improve a commercial service and by consumers in using that service”, can lead to problems in capturing the theoretical constructs of interest and to measurement errors and inconsistencies across cases and over time [40]. In the case of Facebook, for instance, it can be argued that the user interface affects the use of

specific features, that the newsfeed algorithm that determines what users see in their feeds and in what order affects their activity, or that the friend suggestion algorithm affects the process of establishing friendship links. Furthermore, users change their behavior over time; they gain better understanding of specific features, they adapt their behavior to stimuli from the designers, and certain social norms evolve. Changes in any of the above can produce inconsistencies in the activity and network data gathered from the API and hinder their interpretation.

In addition to these limitations, a computational approach to studying SNSs has to face novel challenges with regards to privacy, ethical concerns and sampling. Significant new questions arise, such as how representative the web-based participant sample is, what its limitations are and how it should be selected in order to generalize better, what is the best way to elicit informed consent from participants in an ethical way, what understanding do users have of their data, how the data will be used and analyzed and how this understanding can be enhanced.

III. RESEARCH APPROACH

Based on the above, this PhD work argues that the effective study of the social web, and SNSs in particular, requires a combination of traditional social science and computational methods. Combining the two approaches can help balance the limitations and weaknesses that they exhibit when used in isolation. On one hand, a computational approach provides a wider range of data that are objective, and more accurate and granular than what can be collected by traditional social science methods. On the other hand, the inquiry nature of traditional science methods, such as surveys or interviews, can provide meaningful interpretation to the statistically significant associations unearthed from the computational data [41].

Throughout this work, we are interested in providing empirical evidence for how traditional social science methods and data can be efficiently and harmoniously combined with computational methods and data in order to inform SNS researchers and designers. To this end, we designed two two-part studies that realize this combination of methods and data. The first part of each study has been completed, while the second part of each study is future/ongoing work.

A. Study 1

The first study employs the Uses and Gratifications (U&G) theory to study motivations for using Facebook and Twitter. At the heart of this theory is the fact that media are consumed for a wide range of purposes and individuals utilize different media channels to achieve very different ends [10, 42]. U&G is a theoretical framework for studying these motives and outcomes – fundamentally, the “how” and “why” of media use [9]. A key strength of the approach is its established and broadly applicable frame of analysis (covering media as diverse as tabloids, reality TV and the Internet) that combines *motives* for media use (such as entertainment or social connection) with social and psychological *antecedents* (such as demographics) and cognitive, attitudinal, or behavioral *outcomes* (such as usage patterns) [43]. While traditional U&G studies elicit the motivations for use by employing a survey

instrument, in our approach we expanded the methodological scope of U&G by combining a typical survey tool with data captured using the Facebook API.

In the first part of the study [44], 208 participants were recruited with a request to complete an online survey. They had to explicitly click a link to login with their Facebook credentials and access the survey, which is an equivalent action to installing a Facebook application. After logging in, they were directed to an online survey capturing demographics and presenting 28 questions regarding their gratifications (motivations for use) from Facebook, as done during a typical U&G study. However, in the background and with the users’ explicit consent, the Facebook API collected a range of numeric data about each participant’s usage (e.g., number of status updates posted, number of photographs uploaded - 11 metrics in total) and their personal Facebook networks (e.g., density of the personal network, number of connected components - 8 metrics in total). An exploratory factor analysis was conducted on the 28 items, yielding seven factors, corresponding to motives for Facebook use; *social connection*, *shared identities*, *photographs*, *content*, *social investigation*, *social network surfing* and *newsfeed*. Then we investigated the extent to which these motives can be predicted through the usage and network metrics collected. Results showed that all three variable types in this expanded U&G frame of analysis (covering social antecedents, usage metrics and personal network metrics) effectively predicted motives and highlighted interesting behaviors. For instance, our findings suggest that a particular group of users suffers from information overload in their feeds, while another group are essentially “lurkers”, seemingly inactive on the site, but potentially enthusiastic and regular users aiming to achieve specific, largely observational, goals. These types of users can now be identified from their usage and network data and have their needs catered for. The added value of this expanded analytic framework of U&G theory also manifested [45] when comparing the results to a prominent 2007 U&G study [9] on Facebook.

In the second part of the study, we intend to compare the motives for use across SNSs, in particularly Facebook and Twitter. A similar data collection approach for Twitter users will allow us to compare users’ motives across Facebook and Twitter and examine the role of the computational and network data in both datasets. This is a very important expected contribution, since there is a gap in the literature in examining motives across SNSs, especially SNSs that can be considered complementary to some extent or at least exhibit features that may complement each other.

B. Study 2

The second study employs a Facebook application to calculate several facets of Facebook users’ friendships. In particular, a number of variables that characterize user friendships are collected by the Facebook API and users are requested to rate these friendships in terms of tie strength, trust, and propensity to disclose information. The goal of this study is to model and predict these constructs from the data collected via the API and to examine any possible interplay that takes place among these constructs.

In the first part of the study [46], we employed a Facebook application to calculate the strength of Facebook users' friendships (i.e., tie strength). Participants used a horizontal slider (Fig. 1) to rate the strength of some of their friendships, while in the background the application collected a range of numeric data about each friendship. Specifically, 18 variables (e.g., number of comments exchanged, number of mutual friends) were collected via the Facebook API for 1728 friendships and used to predict tie strength reported by 90 participants. A regression model was created with the data gathered from the API for each rated pair as predictors and the answer to the tie-strength question as the outcome variable. Interestingly, ten of the variables used in the model emerged as significant predictors. The tie-strength calculation was performed in real time by the application, conferring the key advantage that the result can be instantly used by the live application. Thus, we believe that this functionality has the potential to enable many novel customization and recommendation scenarios. Furthermore, examining the effect of the use of different Facebook features and types of communication on the perceived tie strength is a step towards a more comprehensive understanding of the concept of tie strength in social media and may shed light on people's use of social network sites.

The second part of this study aims at extending the first part by examining different facets of interpersonal Facebook relationships in addition to tie strength. In particular, we plan to follow a similar approach to investigate interpersonal trust and three facets of interpersonal privacy (or information disclosure), namely status updates, photograph sharing, and location sharing. By performing a similar analysis to the first part of the study we hope to attain a deeper understanding of these constructs. In addition, we aim to explore relationships among the different constructs and get insights into how these concepts can help us understand the nature of interpersonal relationships on Facebook as a whole.

C. The Research Approach

These two areas were selected for study in this PhD work as recent research suggests that the diversity and the complexities of the socio-computational processes that characterize social systems, as well as the complex interactions among them, require an elaborate research approach [28]. Such a research approach requires a multidisciplinary integration across several levels; interaction levels (e.g., the individual, the dyad, small groups, organizations, communities, society), domains (e.g., different SNSs), and sociotechnical processes or theories (e.g., human motivation, group formation and leadership, incentive design, collective action) [28]. The two studies described in this thesis are designed in a way that attends to this plurality, while employing both traditional social science and computational methods (see Fig. 2). On the domain level, they address Facebook and Twitter, arguably the two most popular SNSs. On the interaction level, they study theories and phenomena about individuals and dyads. Finally,

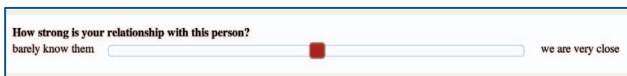


Fig. 1. Screenshot of the horizontal slider used for friendship rating.

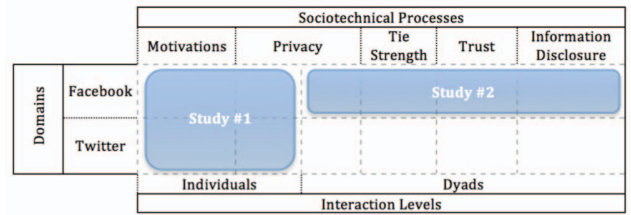


Fig. 2. The two studies provide a multi-perspective exploration of the research area.

on the theoretical level they examine user motivations, tie strength, interpersonal trust, privacy, and information disclosure.

The work proposed examines the nuances and demonstrates the value of the combination of traditional social science and computational approaches to studying SNSs. However, it is important to note that this work is inevitably limited in scope and care should be taken when attempting to interpret or generalize the findings in different contexts. First, the two studies described are domain specific; although Facebook and Twitter exhibit widespread use, high diffusion, and include a large number of features, other SNSs may display very different characteristics, features, user interfaces, APIs, and usage norms, something that may limit the value of our findings in their study. Second, the proposed work covers a number of important topics and theoretical constructs, including motivations for use, tie strength, trust, and information disclosure. However, the value of the findings is limited to supporting designers and developers in creating SNSs that can react and respond accordingly only to these specific characteristics and to supporting researchers in studying the nuances of these specific topics. Third, this work is limited in scope in that it only examines phenomena and theories at the individual and the dyad level. Theories that describe behavior at different levels of interaction, such as group coordination and leadership, community membership, diffusion of information, viral marketing, and tipping point mechanics are out of the scope of the proposed work. Interestingly, recent research [28] suggests that effects from one interaction level can percolate to other levels and create interesting phenomena and encourages the study of these effects. For example, behavioral psychology suggests that the provision of incentive mechanisms can enhance community participation, while game-theoretic interaction and tradeoffs can have an effect at the societal level. Although we make an indirect effort in this direction by contrasting personal motivations for use and network structure (i.e., possible friending approaches), there is clearly room for a lot of interesting work in this regard.

IV. EXPECTED CONTRIBUTIONS

The central claim of the proposed work is that the combination of traditional social science and computational approaches to studying SNSs can provide explanatory value over and above what can be gained through these approaches alone. In order to aid the adoption of this research approach, this work examines the practicability and the potential of combining traditional social science and computational methods and data and aims to provide necessary research

guidelines for the scientific community. At the same time, the value of this approach can be instantiated in the form of design guidance that specifically targets the creators of SNSs. By generating this kind of design knowledge, the proposed PhD work expects to contribute practical, hands-on, how-to recommendations about how future SNSs should be constructed. In this thesis, this knowledge is presented in the form of design guidelines, a common form of output in Human-Computer Interaction (HCI) work. These expected contributions are described in more detail below.

A. Guidelines for the Scientific Community Studying SNSs

The work proposed leverages the Facebook API to gather a set of data that is far larger and more diverse than that in a typical U&G study. A wide range of activity and network data take the place of the monolithic, unidimensional measure of SNS use that is typically employed in U&G studies. Similarly, while studies spanning the same participants' use of more than one SNSs have so far only relied either solely on survey answers or solely on computational data, this work combines the two types of data.

Even though social network analysis has proven useful in providing an understanding of social media [e.g., 15], it has also been suggested that appropriate use of network analysis depends on choosing the right network representation for the problem at hand [47]. It is possible that studying alternative network representations, such as those whose links are weighted based on tie strength, trust, or propensity for information disclosure instead of binary friendships, has the potential to be the basis for substantial advances in understanding user behavior in social media. The second study in this PhD work hopes to lay the groundwork and provide insights to researchers attempting to employ such alternative network representations in their studies.

The two main studies presented in this thesis share a common theme of recruiting participants online and requesting access to their data, but they do so in different manners. An expected contribution in the final thesis is useful insights from the online recruiting procedures and the resulting samples that can benefit researchers engaging in similar activities, focusing on issues such as data privacy, ethics, validity, and effectiveness in recruiting.

B. Design Guidelines for SNSs

Effectively predicting motives and highlighting interesting behaviors on Facebook has several practical implications, because the motives for Facebook use can help us understand why someone behaves the way they do. System designers can use the identified motives of use as part of the user modeling process that is employed to personalize and adapt the system interfaces and the user experience. The different kinds of users, based on their motivations for use, can be identified from their usage data and have their needs catered for. In addition, the relationships identified between specific Facebook motives of use and the way users perceive their privacy, and act on privacy-sensitive issues, can aid the association of specific types of users with the level of privacy or publicity that makes them feel more comfortable and, thus, enable an improved user experience. The identified motives can also be very useful information for marketers who want to promote their products

or services to the users who visit Facebook with a particular goal in mind. For example, users with different needs and goals when visiting the SNS (e.g., receiving information, connecting with other people, or just spending some time) can be targeted with appropriate information, products, or services.

The study on the use and motivations of Facebook and Twitter in tandem is also expected to provide useful practical insights. Examining the combination of the usage data and the motives of the two services together should give us a better understanding of the underlying dynamics that drive social media selection, adoption, use, interconnection, and popularity. For instance, understanding the context in which the two SNSs exhibit complementary use and the context in which they exhibit overlapping use can aid designers in creating features and services that "play well" with each other.

The other branch of the proposed research allows the live calculation of important constructs that describe Facebook relationships (tie strength, trust, information disclosure) from usage data gathered via the API. This opens the door to a wide range of novel applications based on adaptable and customized services. For instance, social media systems could recommend information items and filter newsfeeds based on the values of these constructs. Different types of information, such as questions or status updates, could be broadcast to different contacts for more efficient answers or information diffusion. The default values of privacy controls, or recommended privacy settings for new connections can be set based on these values. Better recommendations for new connections could be made based on the calculated tie strength and trust values.

V. BENEFITS FROM PARTICIPATION IN THE DOCTORAL CONSORTIUM

The RCIS 2015 doctoral consortium will provide me with the opportunity to discuss my research approach with peers and senior colleagues who have different backgrounds, expertise and perspectives. I expect that this interaction will help me gain a better understanding of the possible implications of my work and assist in the overall framing of my PhD. Participating in the doctoral consortium at this time will allow me to identify any gaps in my research prior to completing my final user studies. It will also provide a very valuable networking opportunity resulting in connections with colleagues that would be mutually favorable in our future work.

My PhD work is inherently interdisciplinary, drawing on theory and methods from human-computer interaction, social science and information systems. In fact, it attempts to tackle this interdisciplinarity with a critical approach. I believe that this characteristic will allow me to provide feedback and commentary on a wide range of topics in an analytical manner that can be particularly useful to the other consortium participants.

ACKNOWLEDGMENT

I would like to thank my supervisor, Ian Oakley, and my PhD committee for their feedback and support throughout this research. The work reported in this paper is supported by FCT research grant SFRH/BD/65908/2009.

REFERENCES

- [1] Facebook Newsroom – Company info. <http://newsroom.fb.com/company-info/>
- [2] About Twitter. <https://about.twitter.com/company>
- [3] d. boyd, “Why youth (heart) social network sites: The role of networked publics in teenage social life,” in *MacArthur Foundation Series on Digital Learning – Youth, Identity, and Digital Media Volume*, D. Buckingham, Ed. Cambridge, MA: MIT Press, 2007.
- [4] M. Burke, R. Kraut, and C. Marlow, “Social capital on Facebook: Differentiating uses and users,” in *CHI 2011*, 2011, pp. 571–580.
- [5] Z. Papacharissi and A. Mendelson, “Toward a new(er) sociability: Uses, gratifications and social capital on Facebook,” in *Media Perspectives for the 21st Century*, S. Papathanassopoulos, Ed. Routledge, 2011.
- [6] D. Quercia, R. Lambiotte, D. Stillwell, M. Kosinski, and J. Crowcroft, “The personality of popular facebook users,” in *CSCW 2012*, 2012.
- [7] M. Burke and R. Kraut, “Using Facebook after losing a job : Differential benefits of strong and weak ties,” in *CSCW 2013*, 2013.
- [8] E. Gilbert and K. Karahalios, “Predicting tie strength with social media,” in *CHI 2009*, 2009.
- [9] A. Joinson, “‘Looking at’, ‘looking up’ or ‘keeping up with’ people?: Motives and use of facebook,” in *CHI 2008*, 2008, pp. 1027–1036.
- [10] A. D. Smock, N. B. Ellison, C. Lampe, and D. Y. Wohn, “Facebook as a toolkit: A uses and gratification approach to unbundling feature use,” *Comput. Human Behav.*, vol. 27, no. 6, pp. 2322–2329, Nov. 2011.
- [11] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg, “Structural diversity in social contagion,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 16, pp. 5962–6, Apr. 2012.
- [12] C. Lampe, J. Vitak, R. Gray, and N. Ellison, “Perceptions of Facebook’s value as an information source,” in *CHI 2012*, 2012.
- [13] K. Panovich, R. Miller, and D. Karger, “Tie strength in question & answer on social network sites,” in *CSCW 2012*, 2012.
- [14] d. boyd and E. Hargittai, “Facebook privacy settings: Who cares?,” *First Monday*, vol. 15, no. 8, 2010.
- [15] B. Brooks, H. T. Welsler, B. Hogan, and S. Titsworth, “Socioeconomic Status Updates: Family SES and emergent social capital in college student Facebook networks,” *Information, Commun. Soc.*, vol. 14, no. 4, pp. 529–549, Jun. 2011.
- [16] B. Hogan, “Visualizing and interpreting Facebook networks,” in *Analyzing Social Media Networks with NodeXL*, D. L. Hansen, B. Shneiderman, and M. A. Smith, Eds. Burlington, MA: Morgan Kaufmann, 2010, pp. 165–180.
- [17] S. A. Golder and S. Yardi, “Structural predictors of tie formation in twitter: Transitivity and mutuality,” in *SocialCom 2010*, 2010, pp. 88–95.
- [18] d. boyd, S. A. Golder, and G. Lotan, “Tweet, Tweet, Retweet: Conversational aspects of retweeting on Twitter,” in *HICSS-43*, 2010.
- [19] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, “Measuring user influence in Twitter: The million follower fallacy,” in *ICWSM 2010*, 2010.
- [20] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a social network or a news media?,” in *WWW 2010*, 2010, pp. 591–600.
- [21] Y. Lin, D. Margolin, B. Keegan, A. Baronchelli, and D. Lazer, “#Bigbirds never die: Understanding social dynamics of emergent hashtags,” in *ICWSM 2013*, 2013, pp. 370–379.
- [22] S. Yardi and d. boyd, “Tweeting from the town square: Measuring geographic local networks,” in *ICWSM 2010*, 2010, pp. 194–201.
- [23] Y. Hu, S. Farnham, and A. Monroy-Hernández, “Whoo.ly: Facilitating information seeking for hyperlocal communities using social media,” in *CHI 2013*, 2013.
- [24] D. Quercia, D. Séaghdha, and J. Crowcroft, “Talk of the city: Our tweets, our community happiness,” in *ICWSM 2012*, 2012.
- [25] M. De Choudhury, N. Diakopoulos, and M. Naaman, “Unfolding the event landscape on twitter: Classification and exploration of user categories,” in *CSCW 2012*, 2012.
- [26] G. M. Chen, “Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others,” *Comput. Human Behav.*, vol. 27, no. 2, pp. 755–762, Mar. 2011.
- [27] V. Arnaboldi, M. Conti, A. Passarella, and F. Pezzoni, “Analysis of ego network structure in online social networks,” in *SocialCom 2012*, 2012.
- [28] R. Kraut, M. L. Maher, J. Olson, T. W. Malone, P. Pirolli, and J. C. Thomas, “Scientific Foundations: A case for technology-mediated social-participation theory,” *Computer*, vol. 43, no. 11, pp. 22–28, Nov. 2010.
- [29] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, “Life in the network: the coming age of computational social science,” *Science*, vol. 323, no. 5915, pp. 721–3, Feb. 2009.
- [30] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, “Tastes, ties, and time: A new social network dataset using Facebook.com,” *Soc. Networks*, vol. 30, no. 4, pp. 330–342, Oct. 2008.
- [31] B. Hogan, “Using information networks to study social behavior: An appraisal,” *IEEE Data Eng. Bull.*, vol. 30, no. 2, pp. 6–14, 2007.
- [32] D. Knoke and S. Yang, *Social Network Analysis*, 2nd ed. California: Sage Publications, Inc, 2008.
- [33] G. Kossinets, “Effects of missing data in social networks,” *Soc. Networks*, vol. 28, no. 3, pp. 247–268, Jul. 2006.
- [34] E. Costenbader and T. W. Valente, “The stability of centrality measures when networks are sampled,” *Soc. Networks*, vol. 25, no. 4, pp. 283–307, Oct. 2003.
- [35] B. Wellman, P. J. Carrington, and A. Hall, “Networks as personal communities,” in *Social Structures: A Network Analysis*, B. Wellman and S. D. Berkowitz, Eds. Cambridge, UK: Cambridge University Press, 1988, pp. 130–184.
- [36] C. Lampe, N. Ellison, and C. Steinfield, “A familiar face(book): profile elements as signals in an online social network,” in *CHI 2007*, 2007, pp. 435–444.
- [37] D. Brewer, “Forgetting in the recall-based elicitation of personal and social networks,” *Soc. Networks*, vol. 22, pp. 29–43, 2000.
- [38] A. Paik and K. Sanchagrin, “Social isolation in America: An artifact” Available SSRN <http://ssrn.com/abstract=2101146>, 2012.
- [39] B. Rieder, “Studying Facebook via data extraction: The Netvizz application,” in *ACM WebScience 2013*, 2013.
- [40] D. Lazer, R. Kennedy, G. King, and A. Vespignani, “Big data. The parable of Google Flu: traps in big data analysis,” *Science*, vol. 343, no. 6176, pp. 1203–5, Mar. 2014.
- [41] D. Margolin, Y. Lin, D. Brewer, and D. Lazer, “Matching data and interpretation: Towards a rosetta stone joining behavioral and survey data,” in *The 2nd When the City Meets the Citizen Workshop, ICWSM 2013*, 2013, pp. 9–10.
- [42] E. Katz, M. Gurevitch, and H. Haas, “On the use of the mass media for important things,” *Am. Sociol. Rev.*, vol. 38, pp. 164–181, 1973.
- [43] Z. Papacharissi, “Uses and Gratifications,” in *An Integrated Approach to Communication Theory and Research*, M. Salwen and D. Stacks, Eds. Lawrence Erlbaum, 2008, pp. 137–152.
- [44] T. Spiliotopoulos and I. Oakley, “Understanding motivations for Facebook use: Usage metrics, network structure, and privacy,” in *CHI 2013*, 2013.
- [45] T. Spiliotopoulos and I. Oakley, “Replicating and extending a Facebook Uses & Gratifications study: Five years later,” in *RepliCHI 2013 workshop at ACM CHI 2013*, 2013.
- [46] T. Spiliotopoulos, D. Pereira, and I. Oakley, “Predicting tie strength with the Facebook API,” in *PCI 2014*, 2014.
- [47] C. T. Butts, “Revisiting the foundations of network analysis,” *Science*, vol. 325, no. 5939, pp. 414–6, Jul. 2009.