TD

Studying Social Network Sites with the Combination of Traditional Social Science and Computational Approaches

DOCTORAL THESIS

Anastasios Spiliotopoulos

DOCTORATE IN INFORMATICS ENGINEERING
SPECIALTY IN HUMAN-COMPUTER INTERACTION

UNIVERSIDADE da MADEIRA

A Nossa Universidade

www.uma.pt

December | 2019

Cofinanciado por:

MADEIRA 14-20

REGIÃO AUTÓNOMA DA MADEIRA

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Social Europeu

arditi

UBQ MADEIRA

Studying Social Network Sites with the Combination of Traditional Social Science and Computational Approaches

DOCTORAL THESIS

Anastasios Spiliotopoulos

DOCTORATE IN INFORMATICS ENGINEERING
SPECIALTY IN HUMAN-COMPUTER INTERACTION

ORIENTATION
Ian Oakley

CO-ORIENTATION
Pedro Campos

ABSTRACT

Studying Social Network Sites with the Combination of Traditional Social Science and Computational Approaches

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. First, 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.

The objective of this dissertation is to demonstrate how a suitable combination of these two approaches can lead to a better understanding of user behavior on SNSs and can enhance the design of such systems. To this end, I present two two-part studies that act as four pieces of evidence in support of this objective. In particular, these studies investigate whether a combination of survey and API-collected data can provide additional value and insights when a) predicting Facebook motivations, b) understanding social media selection, c) understanding patterns of communication on Facebook, and d) predicting and modeling tie strength, compared to what can be gained by following a traditional social science or a computational approach in isolation.

I then discuss how the findings from these studies contribute to our understanding of online behavior both at the individual user level, e.g. how people navigate the SNS ecosystem, and at the level of dyadic relationships, e.g. how tie strength and interpersonal trust affect patterns of dyadic communication. Furthermore, I describe specific implications for SNS designers and researchers that arise from this work. For example, the work presented has theoretical implications for the Uses and Gratifications (U&G) framework and for the application of Rational Choice Theory (RCT) in the context of SNS interactions, and design implications such as enhancing SNS users' privacy and convenience by supporting reciprocity of interactions. I also explain how the results of the conducted studies demonstrate the added value of combining traditional social science and

computational methods for the study of SNSs, and, finally, I provide reflections on the strengths and limitations of the overall research approach that can be of use to similar research efforts.

Keywords: social network sites, computational social science, Facebook API, uses and gratifications, online disclosure.

RESUMO

O Estudo de Redes Sociais Combinado com Ciências Sociais Tradicionais e Abordagens Computacionais

As Redes Sociais (*SNSs - Social Network Sites*) estão a mudar de form fundamental a maneira como os seres humanos estabelecem ligações entre si, como comunicam e como relacionam-se uns com os outros, tendo atraído uma considerável quantidade de atenção investigativa. Em geral, duas abordagens de investigação distintas foram seguidas na procura de resultados nesta área de investigação. Em primeiro lugar, os já estabelecidos métodos tradicionais das ciências sociais, tais como inquéritos e entrevistas foram amplamente utilizados na investigação baseada em *SNSs*. Contudo, o surgimento mais recente das Interfaces de Programação de Aplicações (*APIs - Application Programming Interfaces*) tem permitido abordagens centradas em dados que têm culminado em estudos de "dados extensos", livres de teoria. Ambas estas abordagens têm vantagens, desvantagens e limitações que precisam de ser consideradas nos estudos de *SNS*.

O objectivo desta dissertação é demonstrar como uma combinação adequada destas duas abordagens pode levar a uma melhor compreensão do comportamento do utilizador em *SNSs* e pode melhorar a concepção de tais sistemas. Para esse efeito, apresento dois estudos, em duas partes, que funcionam como quatro peças de prova em apoio a este objectivo. Estes estudos investigam, em particular, se uma combinação de dados recolhidos através de inquéritos e *API* pode fornecer valor adicional e conhecimentos ao a) prever as motivações do Facebook, b) compreender a selecção dos meios de comunicação social, c) compreender os padrões de comunicação no Facebook, e d) prever e modelar a força dos laços, em comparação com o que pode ser ganho seguindo uma ciência social tradicional ou uma abordagem computacional isolada.

Abordo em seguida como os resultados destes estudos contribuem para uma compreensão do comportamento online tanto a nível do utilizador individual, por exemplo, como as pessoas percorrem o ecossistema *SNS*, e ao nível das relações diádicas, por exemplo, como a força dos laços e a confiança interpessoal afectam os padrões de comunicação diádica. Além disso, descrevo as implicações específicas para os designers e investigadores do *SNS* que decorrem deste trabalho. Por exemplo, o trabalho apresentado tem implicações teóricas para o quadro de Usos e Gratificações (*U&G - Uses and Gratifications framework*) e para a aplicação da Teoria da Escolha Racional (*RCT - Rational Choice Theory*) no contexto das interacções *SNS*, e implicações de design, como o reforço da privacidade e conveniência dos utilizadores de *SNS*, com o apoio à reciprocidade das interacções. Explico também como os resultados dos estudos realizados demonstram o valor acrescentado de combinar as ciências sociais tradicionais e os métodos computacionais para o estudo de *SNS*, e, por fim, apresento reflexões sobre os pontos fortes e limitações da abordagem global de investigação que podem ser úteis a esforços de investigação semelhantes.

Palavras-chave: redes sociais, ciência social computacional, Facebook API, usos e gratificações, divulgação online.

The work reported in this dissertation has been supported in part by the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia - FCT) research grant SFRH/BD/65908/2009.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1. INTRODUCTION

1.1 INTRODUCTION AND MOTIVATION

Social Network Sites (SNSs), such as Facebook[1] and Twitter[2], 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 currently reports 2.38 billion monthly active users with 65.5% of them accessing the site daily (Facebook, 2019). In comparison, Twitter reports 330 million monthly active users, with 79.4% of accounts coming from outside the USA (Twitter, 2019). These two sites are only part of an ecosystem of SNSs that provide a growing amount of features to users and are becoming increasingly diffused in our everyday lives; a recent survey reports that the median American uses three social media platforms (Pew Research Center, 2018). 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" (boyd & Ellison, 2007), 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).

Understanding people's behavior on SNSs is a complex and challenging topic and has been the subject of extensive research. For example, SNS scholars have examined user motivations (e.g., Joinson, 2008), privacy (e.g., boyd & Hargittai, 2010), social capital (e.g., Burke, Kraut, & Marlow, 2011), personality (e.g., Quercia, Lambiotte, Stillwell, Kosinski, & Crowcroft, 2012), influence (e.g., Cha, Haddadi, Benevenuto, & Gummadi, 2010), tie formation (e.g., Golder & Yardi, 2010)

[1] Facebook.com

[2] Twitter.com

and tie strength (e.g., Gilbert & Karahalios, 2009), network structure (e.g., Ugander, Backstrom, Marlow, & Kleinberg, 2012), and specific site features (e.g., Y. Lin, Margolin, Keegan, Baronchelli, & Lazer, 2013). At the same time, researchers have found similarities between the social structures in offline social networks and SNSs (Arnaboldi, Conti, Passarella, & Pezzoni, 2012) and have suggested that online social networks can represent effective proxies for hard-to-establish real world friendship networks (Hogan, 2010b). 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 and public availability of computational tools like Application Programming Interfaces (APIs) has enabled data-centric approaches that have culminated in theory-free "big data" studies. Throughout the remainder of this dissertation, I refer to the former approach as the *traditional social science* approach and corresponding methods, and the latter as the *computational* approach and corresponding methods. As the following chapters will show, however, both approaches have advantages, disadvantages and limitations.

This dissertation 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, I employ usage and network data collected via an API to complement survey metrics in two two-part 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. Studies on SNS motivations so far have solely relied on self-reported data about usage and have not taken into account users' personal network structure, so including this computational information has the goal to enhance the explanatory value of this kind of studies. The second study focuses on aspects of the interpersonal relationships on Facebook, and in particular tie strength, trust and disclosure. It employs the Facebook API to collect a range of data that quantify the intensity of communication between pairs of Facebook friends and examines how tie strength and interpersonal trust affect patterns of actual behavior (i.e., instead of reported behavior or behavioral tendencies) on Facebook. Furthermore, this study makes use of the API-collected wealth of data that characterizes Facebook friendships to contrast users' self-reported ratings of tie strength with computational models derived from the data.

The remainder of this chapter covers the challenges facing researchers when employing either a traditional social science or a data-centric approach separately for studying SNSs. I then formulate the argument that these problems can be meaningfully and effectively addressed by appropriately combining the two approaches. I describe the overall research approach, delineate the scope of the dissertation, outline the types of contributions of the work, present the origins of this dissertation, and describe the structure of the remainder of this document.

1.2 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 (Kraut et al., 2010; Lazer et al., 2009; Lewis, Kaufman, Gonzalez, Wimmer, & Christakis, 2008). 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 (Knoke & Yang, 2008; Lewis et al., 2008). 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 (Costenbader & Valente, 2003; Knoke & Yang, 2008; Kossinets, 2006). 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 reports on *perceived* ties with friends and family (Hogan, 2007). For example, Wellman and colleagues (Wellman, Carrington, & Hall, 1988) 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 (Lewis et al., 2008), while 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 (Brooks, Welser, Hogan, & Titsworth, 2011) or using longitudinal surveys matched to Facebook server logs to study social capital (Burke et al., 2011), the overall picture shows that the majority of SNS studies typically fail to harness information that could be relevant and useful. For example, a recent literature review of self-disclosure on SNSs reports that 92% of the sampled articles had chosen a survey method to answer their research question (Abramova, Wagner, Krasnova, & Buxmann, 2017).

Third, scholars in these studies have struggled with issues such as recall bias (Brewer, 2000), interviewer effects (Paik & Sanchagrin, 2013), and other sources of measurement error that may accompany survey research (see Lewis et al., 2008 for more discussion). As a result, research has found significant discrepancies between self-reported and actual Facebook use (Junco, 2013), and scholars have started recommending the study of people's behavior in realistic situations instead of lab experiments with self-reported behavioral data (e.g., Knijnenburg, Kobsa, & Jin, 2013; Kokolakis, 2017). 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 (Paik & Sanchagrin, 2013).

1.3 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 (Lazer et al., 2009). 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 (Hogan, 2010b; Lazer et al., 2009). 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., Brooker, Barnett, & Cribbin, 2016; Hogan, 2010b; Rieder, 2013), 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 "trace" or "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 (Lazer, Kennedy, King, & Vespignani, 2014) 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 finding 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.

Researchers have started to acknowledge that relying solely on a computational approach raises ethical (Fairfield & Shtein, 2014), as well as social and methodological (boyd & Crawford, 2012) concerns. For example, Lazer and colleagues (2014) identify two big "traps" in big-data analysis. 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" (Lazer et al., 2014) has led to studies that have possibly overlooked considerable information that could be extracted by traditional social science methods. As a corollary, researchers have warned that when studying electronic traces of social interactions, there is a risk that we will study what is easy to study rather than what is important to study (Resnick, Adar, & Lampe, 2015). 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 (Lazer et al., 2014). 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. In this regard, Gerlitz and Rieder (2018) argue that any picture obtained by API data would be incomplete without taking into account the details of specific (third-party) clients used for organizing, interpreting and engaging with social media content. What's more, 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.

1.4 Research Approach and Questions

Based on the above, this thesis 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.

Notably, scholars and practitioners have increasingly started to find credence in this combined approach, mainly as a response to problems identified in data-centric efforts. Focusing specifically on Social Network Analysis (SNA), Howison and colleagues (2011) identify 10 validity issues when using digital trace data (in the place of surveys that are typically used in SNA) for this kind of research. Interestingly, the solutions they propose for addressing these validity issues in many cases include the triangulation with more traditional science approaches, such as interviews and ethnography. Similarly, a review of studies on Twitter trace data has found inconsistencies in the explanations of the validity of trace data interpretations and suggest data researchers to draw on qualitative research methods to address this problem (Freelon, 2014). When looking at self-tracking data as digital traces, Kneidinger-Müller (2018) argues that a set of contextual factors need to be considered, such as motivations (e.g., individual versus social factors), modes of data selection (e.g., automatic versus manual) and outcomes of data sharing (e.g., individual and social consequences). Attempting to look at the bigger picture, a position paper argues for the establishment of a ”Rosetta Stone” that maps behavioral signatures of population behavior to meaningful social categories by joining behavioral and survey data (Margolin, Lin, Brewer, & Lazer, 2013). A forthcoming academic journal special issue is themed “Integrating Survey Data and Digital Trace Data” (Stier, Breuer, Siegers, & Thorson, 2019) and describes this integration/combination as an emerging field. As another example, a call for papers for a journal special issue scheduled to be published in 2021 devotes itself to studies utilizing new types of data that can be used “in conjunction with surveys, in place of surveys, or to address questions that cannot be addressed by surveys” in the area of public opinion (Conrad, Keusch, & Schober, 2020). From the industry side, in a recent media article Google explains that "data science doesn't cover how humans think through a decision" and argues that we can make better decisions by augmenting data science with softer sciences like psychology, neuroscience, economics, and managerial science, introducing what they have termed Decision Intelligence Engineering as a new discipline (Fast Company, 2018).

Throughout this work, I am 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. Towards this goal, this thesis follows a *proof-by-demonstration* approach, i.e. attempts to provide acceptable pieces of evidence, "in support of a 'proof', where proof is taken to be any convincing argument in support of a worthwhile hypothesis" (Nunamaker, Chen, & Purdin, 1990). Following this approach, I report on two two-part studies that realize this combination of methods and data.

The first study, Study 1, employs the Uses and Gratifications (U&G) approach 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 (Katz, Gurevitch, & Haas, 1973). U&G is a theoretical framework for studying these motives and outcomes – fundamentally, the "how" and "why" of media use (Stafford, Stafford, & Schkade, 2004). 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) (Papacharissi, 2008). While traditional U&G studies elicit the motivations for use by employing a survey instrument, in the first part of the first study (i.e., Study 1A) I propose to expand the methodological scope of U&G by combining a typical survey tool with data captured using the Facebook API[1]. In particular, a Facebook application collected 13 Facebook usage metrics and 8 network metrics that were used as novel forms of outcomes and antecedents in the U&G framework, respectively. In the second part of this study (i.e., Study 1B) I expand the work conducted in Study 1A to investigate how people choose

[1] It is worth mentioning that Facebook has been increasingly limiting access to social graph data (i.e. the specific connections among persons and between people and digital entities on the platform) via the Facebook API. Most notably, since 2015 a standard third party can only access a user's friends if those friends also use the app. The data collection in this dissertation was carried out before these changes took effect. Hogan (2018) provides some interesting further discussion on the details and the implications of these changes.

between different SNSs. In particular, this study combines motives for using Facebook and Twitter and examines the differences between Twitter users and non-users, based on the behavioral data collected.

The second study, Study 2, employs a Facebook application to investigate several facets of Facebook users' friendships. In particular, the application collected 18 activity variables that characterize user friendships and requested participants to rate these friendships in terms of tie strength, trust, and expected reciprocity. The first part of the study, Study 2A, uses 11 of these variables to quantify the actual text- and photo-related communication that has taken place between a participant and a Facebook friend and uses the survey answers to explicate the ways that reported tie strength and trust affect patterns of communication. The second part of the study, Study 2B, uses all 18 collected variables to develop a model of tie strength based on the participant responses and then evaluates the accuracy of this model, while identifying specific friendship characteristics and Facebook features that influence the reported tie strength.

The above areas were selected for study in this thesis as 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 (Kraut et al., 2010). Such a research approach requires a multidisciplinary integration across several levels; interaction levels (e.g., the individual, dyads, 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) (Kraut et al., 2010). The two studies are designed in a way that attends to this plurality, while employing both traditional social science and computational methods (see Figure 1-1). 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, on the theoretical level they examine user motivations, media selection, tie strength, interpersonal trust, and information disclosure.

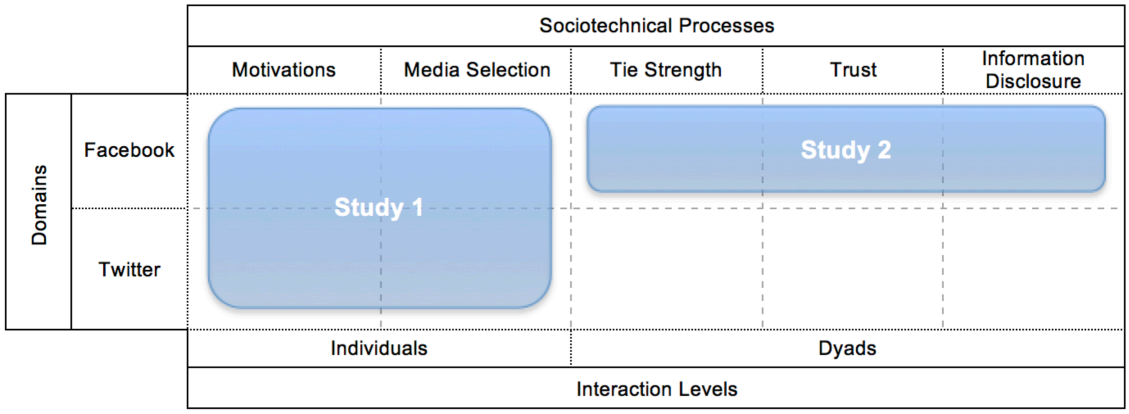

FIGURE 1-1. THE TWO STUDIES PROVIDE A MULTI-PERSPECTIVE EXPLORATION OF THE RESEARCH AREA

Overall, recent research suggests that such a combination of approaches is an interesting and emerging field of research that shows significant potential (e.g., Stier et al., 2019). Thus, the **objective** of this dissertation is to demonstrate how the combination of traditional social science and computational approaches can lead to a better understanding of user behavior on SNSs and can enhance the design of such systems. Based on this, the overarching question that this dissertation seeks to answer is:

RQ1: What value and insights for SNS researchers and designers can be gained by a suitable combination of traditional social science and computational approaches over and above what can be achieved from each approach in isolation?

Following the aforementioned proof-by-demonstration approach, this dissertation intends to provide four pieces of evidence that provide support for the argument that a suitable combination of these approaches can enhance our understanding of user behavior on SNSs and the design of such systems. The evidence provided take the form of the combination of the approaches either a) providing additional explanatory or practical value compared to following each approach separately, or b) answering questions that were not possible to answer effectively or at all by following each approach separately.

More specifically, the studies described in this dissertation seek to answer the following four questions:

RQ2: How can a combination of survey and API-collected data provide additional value and insights when

a) predicting Facebook motivations (Study 1A)

b) understanding social media selection (Study 1B)

c) understanding patterns of dyadic communication on Facebook (Study 2A)

d) predicting and modeling tie strength (Study 2B)

As explained, the work conducted in the dissertation by design explores a multitude and diversity of concepts and processes. Therefore, in addition to the above overarching research questions that overall guide the work, the studies presented also have more specific research questions and hypotheses. These more specific questions and hypotheses are described in the context of each study when each study is presented in detail in Chapters 3 and 4 and contribute to answering the respective RQ2 sub-question for each study.

1.5 Scope of the Dissertation Contributions

The work described in this dissertation is formally and pragmatically situated in the area of Human-Computer Interaction (HCI). As Hewett and colleagues (1992) explain, "Human-computer interaction is a discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them". Using this framework, the current dissertation primarily positions itself in the second part of this definition, i.e. studies phenomena surrounding the interaction of humans and systems, and in this case SNSs. Implied in the above definition, however, is that the results of the study of the interaction can potentially be used for the design, evaluation, and implementation of such systems. With this in mind, this dissertation aims to provide the following three types of contributions.

Contributions describing and explaining the studied phenomena surrounding interaction. In this case, this refers to contributions to our new understanding of online behavior on SNSs. This is encapsulated in what Harrison and colleagues (2007) describe as the second paradigm of HCI, which refers to a classical cognitivism/information processing approach that "emphasizes (ideally predictive) models and theories and the relationship between what is in the computer and in the human mind". This entails contributions grounded in the results of the conducted studies themselves, such as an enhanced understanding of people's motivations to use SNSs, social media selection, patterns of dyadic SNS communication, and specific constructs like tie strength, trust and reciprocity. It is important to note, however, that in the aforementioned context of the proof-by-demonstration approach, the results of these studies are intended as demonstrations that are used to support the central claim of the dissertation (as expressed by the main research question). Therefore, extensive commentary on the individual (e.g., psychological) or social (e.g., sociological) implications of these results is outside the remit of this dissertation.

Contributions concerned with the design, evaluation, implementation, and management of the systems under study. In this case, this refers to identifying and describing contributions aimed at SNS designers and practitioners. Contributions of this type correspond to the first paradigm of HCI (Harrison et al., 2007), which is a human-factors based, largely a-theoretic and pragmatic approach, that conceptualizes interaction as a form of "man-machine coupling", with the goal of optimizing the fit between humans and machines. It is worth noting that, in answering the main research questions of the dissertation, contributions of this type are rather secondary. Deriving an exhaustive or defensible set of design guidelines would formally entail an additional instrumental component such as an implementation of an intervention/interface/design that tests or otherwise evaluates these design guidelines. Such an undertaking is out of the scope of the current dissertation. Instead, the goal of the dissertation in this context is to report on empirically-derived insights obtained from the conducted studies that can inform research and design of SNSs. In some cases, these insights indeed take the form of recommendations for the design or management of systems.

Contributions grounded in the value gained by combining traditional social science methods with computational ones. Contributions of this type derive from the conducted work in two ways. First, the studies are explicitly designed and results are analyzed to provide findings that highlight and explicate the value gained by combining traditional social science methods with computational ones, in a way that goes beyond demonstrating that this approach simply "works" (i.e., as would be the case with the first type of contributions). For example, Study 1A quantifies the effect of including API-collected data as antecedents and outcomes in the U&G frame of analysis. Second, the dissertation reflects on the research approach followed and, thus, contributes to SNS research by describing some practical lessons learned from combining traditional social science methods with computational ones that may be useful to other scholars. These lessons learned take the form of practical recommendations of things to be considered when conducting studies of this kind, as well as an account of the applicable limitations. The insights presented involve sampling and recruitment, study design, and data analysis issues.

1.6 Dissertation Outline

The remainder of the dissertation is organized as follows:

Chapter 2 provides an overview of related work in order to situate and contextualize the original research conducted and presented in the dissertation. Chapters 3 and 4 present each of the two two-part studies, Study 1 and Study 2 respectively. The two studies are distinguished conceptually by their focus on different interaction levels. As such, they can be considered the two main research threads of the dissertation.

Chapter 3 covers the first research thread, which focuses on the study of the individual SNS user. This thread investigates the concepts and processes regarding motivations for using SNSs and how individuals navigate the SNS landscape by engaging in social media selection. Accordingly, the data collection process focuses on the individual user. Section 3.2 first presents the overall study set-up and data collection, and then Section 3.3 presents Study 1A which examines motivations for using Facebook, and Section 3.4 presents Study 1B which examines SNS selection.

Chapter 4 covers the second research thread, which focuses on the study of SNSs at the level of the dyad. This thread investigates the concepts and processes regarding information disclosure, expected reciprocity, tie strength and interpersonal trust. Accordingly, the data selection process is focused on dyadic relationships. Section 4.2 first presents the overall study set-up and data collection, and then Section 4.3 presents Study 2A which examines the ways that tie strength and trust affect patterns of dyadic communication on Facebook, and Section 4.4 presents Study 2B which investigates the feasibility and practicability of predicting tie strength from Facebook API-collected data.

Chapter 5 concludes the dissertation by providing an account of the contributions of the work presented and discussing possible future directions.

1.7 ORIGINS AND ATTRIBUTION

The original work presented in this dissertation has been previously presented in the following publications:

i. **Spiliotopoulos, T.**, & Oakley, I. (2012). Applications of Social Network Analysis for User Modeling. In International Workshop on User Modeling from Social Media at ACM IUI 2012, Lisbon, Portugal.

ii. **Spiliotopoulos, T.**, & Oakley, I. (2013). Understanding Motivations for Facebook Use: Usage Metrics, Network Structure, and Privacy. In proceedings of ACM CHI 2013, Paris, France. [acceptance rate: 20% (392/1963)]. **RepliCHI award*

iii. **Spiliotopoulos, T.**, & Oakley, I. (2013). Replicating and Extending a Facebook Uses & Gratifications Study: Five Years Later. In proceedings of the CHI2013 Workshop on the Replication of HCI Research, Paris, France.

iv. **Spiliotopoulos T.**, Karnik M., Oakley I., Venkatanathan J, & Nisi V. (2013). Towards Understanding Social Media: Two Studies Exploring the Uses and Gratifications of Facebook. In proceedings of HCI Korea 2013, Gangwon, South Korea.

v. **Spiliotopoulos, T.**, Pereira, D., & Oakley, I. (2014). Predicting Tie Strength with the Facebook API. In proceedings of PCI 2014, Athens, Greece. **best paper award*

vi. **Spiliotopoulos, T.** (2015). Studying Social Network Sites via Computational Methods. In proceedings of IEEE RCIS 2015, Athens, Greece. **doctoral consortium paper*

vii. **Spiliotopoulos, T.**, & Oakley, I. (2015). An Exploratory Study on the Use of Twitter and Facebook in Tandem. In proceedings of British HCI 2015 Works-in-Progress, Lincoln, UK.

viii. **Spiliotopoulos, T.**, & Oakley, I. (2016). Post or Tweet: Lessons from a Study of Facebook and Twitter Usage. In Following user pathways: Using cross platform and mixed methods analysis in social media studies workshop at ACM CHI 2016, San Jose, CA, USA.

ix. **Spiliotopoulos, T.**, & Oakley, I. (2020). An Exploration of Motives and Behavior across Facebook and Twitter. Journal of Systems and Information Technology, 22(2), 201-222.

x. **Spiliotopoulos, T.**, & Oakley, I. (2019). Altruistic and Selfish Communication on Social Media: The Moderating Effects of Tie Strength and Interpersonal Trust. Behaviour & Information Technology. (in press)

The work presented in publications (i), (ii), (iii), (vi), (vii), (viii), (ix) and (x) was conceived, designed, conducted, analyzed and reported by me, under the active guidance and supervision of Ian Oakley. Publication (iv) was written by Ian Oakley and synthesizes work presented in publication (ii) and (Karnik, Oakley, Venkatanathan, Spiliotopoulos, & Nisi, 2013). The work presented in publication (v) was conceived, designed and conducted by Diogo Pereira, jointly analyzed by Diogo Pereira and me, and reported by me. During this process, Diogo Pereira was jointly guided and supervised by Ian Oakley and me. The work presented in publication (x) is based on a data collection effort by Diogo Pereira, and otherwise conceived, designed, conducted, analyzed and reported by me, under the active guidance and supervision of Ian Oakley.

Publication (vi) is a doctoral consortium paper that presented and discussed the overall framing of the dissertation. The contribution of publication (vi) lies primarily in Chapter 1 of the dissertation. Besides this, the contribution of each of the above publications in the dissertation is discussed in detail in the introductions of Chapters 3 and 4.[1]

[1] As explained, the research presented in chapters 3 and 4 was conducted with the contribution of co-authors. Therefore, these sections are written in the "we" form.

Chapter 2. Background

2.1 Introduction

This chapter provides an overview of related research in order to situate and contextualize the work presented in the next chapters. Guided by the two main threads of research in this dissertation that examine the different levels of analysis in the study of SNSs, the related work is presented in two main sections respectively; Section 2.2 focuses on the study of SNSs at the level of the individual, and Section 2.3 on the study of dyadic relationships.

In terms of the study of the individual SNS user, I present the U&G theoretical and analytical framework that is employed in Studies 1A and 1B and discuss the uses and gratifications of social media, with a focus on the two SNSs that are studied in this dissertation, Facebook and Twitter. This section also describes how the U&G framework can explain social media selection, which is the focus of Study 1B, and presents examples of cross-platform U&G studies. As Study 1A aims at introducing computationally collected usage metrics as novel outcomes to the U&G framework, I then present related work on measuring Facebook usage. Study 1A also investigates the introduction of personal network metrics as novel antecedents to the U&G framework, so I describe the personal network metrics that are used in this study. As mentioned earlier, Study 1B expands the work conducted in Study 1A to examine multiple SNSs. Therefore, I survey the literature that examines how people use multiple SNSs and I identify factors that influence how SNS users navigate the current social media ecosystem. Study 1B leverages computational data to contrast users and non-users of Twitter, so I also provide details of the literature that tackles use and non-use of technology and SNSs.

The second thread of the dissertation investigates aspects of dyadic relationships on SNSs, and Facebook in particular. Study 2A focuses on the antecedents of communication at the dyadic level, thus I provide an overview of the research that examines disclosure on SNSs. Because this study differentiates between text-related and photograph-related disclosures, I also describe the nuances of communication around photographic content on social media. Next, I describe how research has attempted to frame online communication as a function of expected reciprocity, and I present the literature that has studied the concepts of tie strength and interpersonal trust, i.e. the two concepts that Study 2A hypothesizes that have an effect on the relationship between expected reciprocity and intensity of communication. Tie strength is also a core concept in Study 2B that attempts to predict reported tie strength from computationally collected data regarding Facebook relationships.

2.2 Studying SNSs at the Level of the Individual

2.2.1 User motivations and the U&G theoretical and analytical framework

Media are consumed for a wide range of purposes and individuals utilize different media channels to achieve very different ends. Understanding how different users make the decision to use different media requires the study of these purposes and ends and is especially germane in the area of SNSs, where an abundance of options are available, adoption and switching costs are low, and cross-platform studies are notably lacking.

Uses and gratifications is a media use paradigm from mass communications research that has been used extensively for the study of traditional media, such as newspapers, radio and television, has shown to adapt effectively to newer communication technologies, such as email and the internet (Ruggiero, 2000; Stafford et al., 2004), and has emerged recently as a particularly useful approach for the study of SNSs (Quan-Haase & Young, 2014; Sundar & Limperos, 2013). U&G follows an audience-based approach, grounded theoretically on the assumption that individuals select media and content to fulfill felt needs or wants, with these needs expressed as motivations for adopting particular medium use (Katz et al., 1973; Stafford et al., 2004). As such, the U&G frame of analysis places emphasis on the combination of *motives* for media use (such as social connection and entertainment) with social and psychological *antecedents* (such as demographics) and cognitive, attitudinal, or behavioral *outcomes* (such as usage patterns) (Papacharissi, 2008).

U&G has been valuable in exploring and explaining a wide variety of social media phenomena in a comprehensive way; in fact, Stafford and colleagues (2004) describe this approach as the "how and why" of media use. In addition, the U&G approach can be a very powerful tool for understanding media selection and provide useful insights into how people navigate the current media ecosystem and select which particular SNS (of the many available) they will spend their time and attention on. In this regard, Krcmar and Strizhakova (2009) have made the case that gratifications for using media may be considered as the first stage of media use and, thus, play a key role in understanding media selection. Quan-Haase and Young (2014) further explain that by treating the media audience as active seekers and users, the U&G perspective provides insights about individual preference and interchangeability of communication channels and thus allows for more explanatory power in understanding the contemporary media environment.

2.2.1.1 Uses and Gratifications of Facebook and Twitter

As the currently dominant SNS, Facebook has been the subject of extensive U&G research. Notable work on this platform has identified seven unique motives for Facebook use: social connection, shared identities, photographs, content, social investigation, social network surfing and status updating (Joinson, 2008). The same study also showed that user demographics, site visit patterns and privacy settings were associated with specific motives. Papacharissi and Mendelson (2011) found links between Facebook motives, social and psychological predispositions, and the production of different forms of social capital. Smock et al (2011) studied user motivations associated with the use of specific features of Facebook, while Giannakos et al (2013) found evidence of a more ritualistic use of Facebook expressed as a motive for wasting time. Basak and Calisir (2015) found entertainment and status seeking to have indirect significant effects on continuance intention to use Facebook, whereas information seeking and self-expression to have insignificant effects. Rae and Lonborg (2015) found that motivations for using Facebook moderated the association between Facebook use and psychological well-being.

Research on motivations for using Twitter has been somewhat sparser. Johnson and Yang (2009) made a distinction between informational and social motives of Twitter use and examined the relationships between gratifications obtained and Twitter usage to find that positive relationships existed only in the case of information gratifications and not the social gratifications. More recent work focused on scholars' use of Twitter echo these findings identifying distinct informational and social uses and gratifications (Quan-Haase, Martin, & McCay-Peet, 2015). Chen (2011) singled out Twitter users' need for connection with other users and studied how this need is gratified by using this particular medium. Coursaris et al (2013) studied how the motivations for information, relaxation, and social interaction showcased the differences between active and inactive Twitter users. A study focused on Twitter opinion leaders (C. S. Park, 2013) found that they have higher motivations for information seeking, mobilization, and public expression than non-leaders. Finally, Liu et al (2010) attributed continuance intention to use Twitter to content gratification and new technology gratification.

2.2.1.2 Cross-platform motivational studies

Although the overwhelming majority of motivational research has been conducted on single platforms, U&G researchers have recognized that the current media environment requires the study of SNSs across platforms and have strongly argued for cross-platform U&G studies (e.g., Papacharissi and Mendelson 2011). It is also important to note that in addition to navigating the complexity of the media ecosystem, there are theoretical implications for cross-platform U&G studies; as Ruggiero (2000) explains, "a wide range of gratifications have been proposed across single-platform studies, with distinct and diffuse typologies, and this disparity in the literature has made it difficult for scholars to compare research findings and to develop internally coherent theoretical frameworks". Studying media in cross-platform studies, as opposed to comparing motivations for different media elicited from different single-platform studies, can effectively address this problem.

Initial work that looked into motivations for using more than one social media did not differentiate among the media studied and grouped them all under the term "social media" (e.g., Raacke & Bonds-Raacke, 2008; Urista, Dong, & Day, 2009). Soon researchers realized that, as has been the case with traditional media, people use social media strategically to fulfill specific needs and that, therefore, there is a need to differentiate among the distinct media in cross-platform studies in order to understand social media selection. A comparative analysis of Facebook and instant messaging gratifications found that users were motivated to use the former for having fun and knowing about the social activities occurring in their social network, whereas use of the latter was geared more towards relationship maintenance and development (Quan-Haase & Young, 2010). Johnson and Kaye (2015) compared motivations for using blogs, Facebook and Twitter, but focused their research strictly on the use of the media for political information, therefore providing only limited insights into the broader media selection process. A study examining stakeholder motives for corporate Facebook, Twitter and YouTube pages aimed to differentiate between digital natives and politicians to find that the former preferred Facebook to interact with companies, whereas the latter preferred Twitter (Ruehl & Ingenhoff, 2015). A comparison of motivations for following and engaging with brands on SNSs showed differences across platforms in this regard (Phua, Jin, & Kim, 2017a). Another study compared social capital on four SNSs to find that Twitter users had the highest bridging social capital, while Snapchat users exhibited the highest bonding social capital (Phua, Jin, & Kim, 2017b). A mixed-design survey found differences between image-based platforms (such as Instagram) and text-based platforms (such as Twitter) with regard to users' loneliness, happiness and satisfaction with life (Pittman & Reich, 2016). In a recent study of four SNSs, participants reported using all four platforms equally to share information, while gave entertainment and convenience the highest rating among eight identified motivations for using the platforms (Alhabash & Ma, 2017).

2.2.2 *UNDERSTANDING AND MEASURING FACEBOOK USAGE*

Usage of SNSs has most commonly been captured by self-report methods using surveys, with typical questions including time spent on site and visit frequency (e.g., Joinson, 2008; Quan-Haase & Young, 2010; Raacke & Bonds-Raacke, 2008; Urista, Dong, & Day, 2009). In the case of Facebook, researchers acknowledging a lack of rigor in such ad-hoc methods have suggested more representative measures of usage, such as the Facebook Intensity Scale (Ellison, Steinfield, & Lampe, 2007) which captures the extent to which users are emotionally connected to Facebook and the extent to which the site is integrated into their daily activities. Other studies have argued for unbundling media use to its constituent features and presenting it with more than unidimensional measures (Smock et al., 2011). At the same time, SNS research is putting increasing emphasis in the study of specific Facebook features, such as direct communication (Y.-C. Wang, Hinsberger, & Kraut, 2016), groups (Karnik et al., 2013), photograph sharing (Malik, Dhir, & Nieminen, 2016), and Facebook likes (Levordashka, Utz, & Ambros, 2016), however, this work is typically feature-centric and does not compare usage across features, thus offering limited insights into media selection.

Furthermore, scholars have identified the need to not only unpack SNS usage into its constituents, but to move away from self-reported measures of user activity altogether in favor of computationally collected usage data. A study comparing self-reported and actual Facebook use (Junco, 2013) found significant discrepancies between the two measures, while network researchers have argued that computationally collected usage data can avoid sources of measurement error that may accompany survey research (Lewis et al., 2008), such as recall bias (Brewer, 2000) and interviewer effects (Paik & Sanchagrin, 2013). This sentiment is echoed by research on information disclosure that has verified a discrepancy between stated privacy attitudes and actual behavior, with scholars suggesting the study of people's behavior in realistic situations instead of lab experiments with self-reported behavioral data (Knijnenburg et al., 2013; Quinn, 2016). Further studies on Facebook have addressed this concern by employing the Facebook API to gather broader and more granular data about users' online social activities (Luarn & Chiu, 2015; Rieder, 2013).

2.2.3 *PERSONAL SOCIAL NETWORK METRICS*

Studies of the structure of personal networks, i.e., the networks comprised by the social relationships a participant (ego) maintains with other people (alters), have revealed that network structure can provide a very useful perspective for understanding important theoretical constructs. In fact, a basic tenet of the field of social network analysis is that an individual's position in a network can provide a better understanding of "what's going on" or "what's important" than that person's individual attributes, and it has been argued that exclusively focusing on actor attributes leads to the loss of many important explanatory insights provided by network perspectives on social behavior (Knoke & Yang, 2008).

Results from network studies have found striking similarities between the social structures in offline and online personal social networks (Arnaboldi et al., 2012), and it has been argued that Facebook networks represent complete and unbiased proxies for hard-to-establish real world friendship networks (Hogan, 2010b). Reflecting this perspective, Facebook personal network structure has been associated with many important social constructs and phenomena, such as social capital (Brooks et al., 2011), personality (Quercia et al., 2012), and diffusion of information (Bakshy, Rosenn, Marlow, & Adamic, 2012). The advent of SNSs has greatly facilitated the capture of personal social network data and a wide range of useful metrics can now be calculated automatically and in real time (Hogan, 2010b). Commonly used metrics include:

- *Network Size*: The number of nodes in a participant's egocentric network, i.e., the number of friends that an individual has. Correlations have been shown between network size and personality (Quercia et al., 2012) and social capital (Brooks et al., 2011).

- *Network Density*: The extent that nodes in an egocentric network are interconnected – essentially, how many of an individuals' friends know each other. This is calculated as the ratio of the number of ties to the number of possible ties.

- *Average Degree*: Mean number of mutual friends in an egocentric network. Higher values on this statistic have previously been associated with bonding social capital and higher socioeconomic status (Brooks et al., 2011).

- *Average Path Length*: The average geodesic distance between all pairs of nodes in a network.

- *Diameter*: The longest geodesic distance within the network, i.e., maximum distance between two nodes.

- *Network Modularity*: A scalar value between −1 and 1 that measures the density of links inside communities as compared to links between communities (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008).

- *Number of Connected Components*: The number of distinct clusters within a network. This has been interpreted as the number of an individual's social contexts (Ugander et al., 2012) and associated with bridging social capital (Brooks et al., 2011) and social contagion (Ugander et al., 2012).

- *Average Clustering Coefficient*: The clustering coefficient is a measure of the embeddedness of a node in its neighborhood. The average gives an overall indication of the clustering in the network, and high values are associated with a "small-world" effect (Watts & Strogatz, 1998).

2.2.4 USING MULTIPLE SNSS

As SNSs are very popular and diffused in the population, there is an abundance of single-platform SNS studies. However, researchers have noted a lack of cross-platform studies (Hall, Mazarakis, Chorley, & Caton, 2018; Lampinen, 2016; X. Zhao, Lampe, & Ellison, 2016), especially in the area of media selection. This is an important oversight, as people take part in a converging media environment where SNSs present functional alternatives to each other (Papacharissi, 2008). For example, Zhao, Lampe, and Ellison (2016) made the case for further cross-platform studies when describing the current SNS ecosystem in terms of two tensions that participants had to manage when communicating using multiple SNSs; a tension between maintaining boundaries between platforms or allowing content and audience to permeate across these boundaries; and a tension between remaining in a stable SNS ecosystem or taking up new SNSs driven by the emergence of new tools, practices and contacts. In this tense and competitive SNS environment research has identified many factors that come into play when a user decides which service or combination of services is more effective for meeting their information and communication needs and selects to spend their time and attention on.

One of these factors is that users consider how the technical attributes and the selection of features provided by a SNS "affords" different types of activities (Kaur et al., 2018; B. Kim & Kim, 2019; Trepte, Scharkow, & Dienlin, 2020; Valenzuela, Correa, & Gil de Zúñiga, 2018; Vitak & Kim, 2014). Stemming from research in psychology (Gibson, 1979) and taken up by HCI (Norman, 1999), this *affordance* perspective makes reference to the perceived, actionable properties that are visually suggestive of the nature of user interaction with the medium (Sundar

& Limperos, 2013). For example, Facebook affords the ability to organize photographs into albums. In turn, this allows users both to curate their photographs for personal archiving (Richardson & Hessey, 2009; X. Zhao & Lindley, 2014) and to more carefully and strategically present themselves online (Hogan, 2010a; Marder, Joinson, & Shankar, 2011). Recent work focusing on perceived affordances for self-presentation has found significant variation across social media platforms; for example, Facebook was found to afford high levels of identity persistence and high visibility control, thus allowing for more granular management of content and identity, while Twitter was characterized by high perceived content persistence and content association affordances, thus considered more suitable as a broadcast environment with public visibility (DeVito, Birnholtz, & Hancock, 2017). Notably, Sundar and Limperos (2013) argue that affordances shape not only how we use a medium, but also how we assemble meaning from it, and, as a result, how we construct and gratify our needs from it.

Another approach for understanding and describing multi-SNS use can be traced to the faceted-identity theory, which posits that people maintain social boundaries and display different facets of their character depending on the social context (Farnham & Churchill, 2011). However, current SNSs make it difficult to achieve this flexibility as, by default, they collapse multiple audiences into single contexts capable of presenting only a single perspective, leading to a problem of *context-collapse* (Binder, Howes, and Sutcliffe 2009; Marwick and boyd 2010). One way for users to deal with this problem has been to compartmentalize their social media use and address different audiences with different services and content (Frederic & Woodrow, 2012; Marder, Joinson, Shankar, & Thirlaway, 2016; Ozenc & Farnham, 2011; Wilken, 2015; Zhong, Chan, Karamshuk, Lee, & Sastry, 2017), but the details of this process remain an open research issue.

Researchers have also pointed to information, communication and feature *overload* on a single channel as a potential challenge and a factor that affects SNS use. As the size and diversity of a user's network grows, the volume of social demands from a SNS may become overwhelming. In turn, this may cause psychological and physiological strain and lead to the selective usage of SNSs as a coping mechanism (Archambault & Grudin, 2012; A. R. Lee, Son, & Kim, 2016; Yao, Phang, & Ling, 2015). Recent research even suggests that SNS users may theorize about the overload experience of their audience in order to infer how their audience will behave (Moll, Pieschl, & Bromme, 2017).

Other elements that potentially shape SNS selection are *demographics* and *cultural characteristics*; for example, Facebook and Twitter have taken the back seat in China in favor of services like Weibo and Renren (Chiu, Ip, & Silverman, 2012), while Tuenti is especially popular among young Spanish people (García-Martín & García-Sánchez, 2015). A study of Facebook users of five countries found an effect of culture on their motivations, instrumental uses and the time they invested on the site (Vasalou, Joinson, & Courvoisier, 2010).

Finally, multiple-SNS use can be affected by the fact that it is becoming easy to use different services together from a *technical* perspective. Many SNSs now expose their APIs to the developer community, allowing and encouraging users to use their credentials on one site to log into another site so that they can automatically post the same content across sites or import their contact lists from other services (X. Zhao et al., 2016).

2.2.5 SNS USE AND NON-USE

Multi-platform SNS studies often perform comparisons on different samples of participants for each platform (e.g., H. Lin & Qiu, 2013; Yu, 2016), or they employ a single sample but the inquiry is directed at the use of a single platform, i.e. asking participants to select one SNS to elaborate on one platform without reflecting on the others (e.g., Phua et al., 2017b). Buccafurri and colleagues (2015) have discussed the drawbacks of utilizing this approach to sampling and strongly advocate the use of a common sample when examining behavioral data computationally extracted from the web, while U&G research has also started following this recommendation (e.g., Alhabash & Ma, 2017). However, to our knowledge, no multi-platform SNS study examines both motivations and behaviors by employing a common set of users across platforms. This is particularly important because a common set of users is more likely to provide useful insights into the nuances of media selection than distinct samples for each site. Indeed, as in the contemporary media environment all media potentially present functional alternatives to each other (Papacharissi, 2008), a common set of users helps to shed light into how people select these functional alternatives.

However, a common set of users can lead to methodological challenges, as not all participants make use of all sites that are being studied, and therefore the study of non-use of a site needs to be taken into account. U&G scholars examining media selection have given particular attention to the concept of media dependency, i.e. the tendency for a user to rely heavily on a particular communication medium for the fulfillment of their needs and wants (Papacharissi, 2008; Rubin & Windahl, 1986). Other researchers have studied aspects of non-use either for technology in general (Baumer, Ames, Burrell, Brubaker, & Dourish, 2015; Satchell & Dourish, 2009), or for a single SNS (Baumer et al., 2013; Coursaris et al., 2013; Lampe, Vitak, & Ellison, 2013;

Schoenebeck, 2014). From the perspective of continuance intention, the U&G theory suggests that if individuals perceive the obtained gratifications of a medium to be satisfactory, they will continue their usage and not engage in abandonment or non-use (Krasnova, Veltri, Eling, & Buxmann, 2017; Ku, Chen, & Zhang, 2013). Other research has found that previous usage behavior and a network effect (i.e., connections already present on a platform) are the most important determinates of continuance intention (K.-M. Lin, 2016), suggesting a "stickiness" effect of a SNS or an "inertia" effect for using alternative media. Importantly, self-report studies have suggested that differentiation and richness of features are factors that lead to non-use of SNSs (Grandhi, Plotnick, & Hiltz, 2019), but this remains to be validated with behavioral data. These findings highlight the importance of studying non-use for understanding media selection and call for more research in this area.

2.3 Studying Dyadic Relationships

2.3.1 Disclosure on SNSs

People make use of SNSs to share a diversity of content to multiple audiences. SNS users share personal information to their connections in the platform not only actively, such as via status updates, comments, and photographs, but also passively through information revealed in their profiles, such as dates of birth, relationship information and events they are interested in attending. Furthermore, even information that is forwarded or reshared from third parties, although not personal in content, can have personal implications; for example, sharing a specific news story may imply that the sharer endorses or agrees with the content and that a receiver will find it worthwhile for their attention. Thus, communication is often studied in terms of acts of self-disclosure, traditionally defined as "any message about the self that a person communicates to another" (Wheeless & Grotz, 1976) with a clear implication that this communication is deliberate (Greene, Derlega, & Mathews, 2006). Online self-disclosure can reduce the uncertainty of dyadic interactions (Tidwell & Walther, 2002) and it has been shown that people like those who self-disclose to them (Jiang, Bazarova, & Hancock, 2011; Kashian, Jang, Shin, Dai, & Walther, 2017). As their friend networks increase in size over time and comprise different and potentially conflicting social spheres, SNS users can find it challenging to manage their sharing strategies and behaviors (Binder et al., 2009; Marder et al., 2011; Vitak, 2012). In response to this problem, SNSs allow their users to fine-tune sharing by creating predefined lists or "circles" of connections, or to select the recipients of their messages on an ad-hoc case-by-case basis (Kairam, Brzozowski, Huffaker, & Chi, 2012; Kelley, Brewer, Mayer, Cranor, & Sadeh, 2011).

2.3.1.1 Photograph-related sharing

In addition to text communication, photo sharing has emerged as a very popular activity on SNSs. This trend is partly fuelled by the proliferation of smartphones that allow users to take pictures with the camera of their devices and quickly share them on the mobile versions of SNS applications; in fact, a recent report shows that 95.1% of active user accounts access Facebook via smartphone (Statista, 2018). Furthermore, more than half of internet users post or share photos online, with 52% of users overall being creators of photographic content, i.e., have posted photos they have taken themselves, and 42% being curators, i.e., have posted photos they have found online (Pew Research Center, 2013). Sharing photographs captured by camera-phones has been described as a distinct form of self-impression management, in that it allows the dynamic reconfiguration of private/public boundaries by disclosing more information about oneself than verbal communication (D.-H. Lee, 2009). Specifically for Facebook, sharing, tagging and viewing photographs have been grouped into a distinct motivation for using the service (Joinson, 2008). Tosun (2012) has argued that active and passive ways of involvement with photos on Facebook are motivated by separate factors, while other research has found that different patterns of photo-related activity are associated with different personality characteristics (Eftekhar, Fullwood, & Morris, 2014). A qualitative analysis of college students' Facebook photos described photos as a means for strategic representation of a social group and social life with a focus on the connection and effective communication among the students, something that goes beyond merely documenting college life (Mendelson & Papacharissi, 2010). An online survey identified six gratifications for digital photo sharing on Facebook, namely, affection, attention seeking, disclosure, habit, information sharing, and social influence (Malik, Dhir, et al., 2016). Finally, a recent study found that photograph sharing on Facebook varies with relationship type, thus highlighting the importance of the relationship between the discloser and the recipient in photo-related sharing (Houghton, Joinson, Caldwell, Marder, & Collins, 2018).

2.3.2 COMMUNICATION AS A FUNCTION OF EXPECTED RECIPROCITY

While much evidence suggests that privacy is a universal human need and needs to be upheld, self-disclosure confers numerous objective and subjective benefits (Acquisti, Brandimarte, & Loewenstein, 2015). In fact, current privacy and communication scholarship is often traced back to the Rational Choice Theory (RCT) (Scott, 2000) and its application to social interactions, the Social Exchange Theory (Cook, Cheshire, Rice, & Nakagawa, 2013; Homans, 1958), which posit that human relationships are formed by applications of a subjective cost-benefit analysis. This suggests that individuals engage in a decision-making process whereby they weigh the perceived benefits of their disclosure activity against the potential privacy risks (Joinson & Paine, 2007;

Laufer & Wolfe, 1977), a process that has led to the development of a Privacy Calculus model (Dienlin & Metzger, 2016; Dinev & Hart, 2006; Krasnova, Spiekermann, Koroleva, & Hildebrand, 2010). However, in the context of online interactions, this rational-actor approach has also been criticized that it lacks nuance and reduces the complexity of social relations to a utilitarian rationality (Pelaprat & Brown, 2012). In fact, although research has found both SNS use, in general, and disclosure on SNSs, in particular, to be associated with numerous objective and subjective benefits, not all of these benefits can be explained as results of goal-directed actors making self-interested decisions. Instead, many of them may be considered products or externalities resulting from more complex social processes. Indicatively, studies show that certain motivations and patterns of Facebook use and self-disclosure are associated with increased social capital (Ellison et al., 2007), formation, maintenance and development of relationships (Krasnova et al., 2010; Tosun, 2012), social support (Huang, 2016), relational intimacy (N. Park, Jin, & Annie Jin, 2011), self-esteem (Steinfield, Ellison, & Lampe, 2008), subjective well-being (Burke, Marlow, & Lento, 2010; Huang, 2016; Islam & Patil, 2015; J. Kim & Lee, 2011), positive emotional states (Neubaum & Krämer, 2015), college adjustment (Yang & Brown, 2015) and political expression (Yu, 2016). What's more, these perceived benefits may be at odds with one another; for example, someone may post their political opinions online in order to attain the personal gratification of political expression, but this action may in turn alienate part of their audience. This observation highlights the complexity of the relationship between perceived benefits and intentionality of interaction, and suggests an examination of interaction at the dyadic relationship level.

Expected reciprocity has been identified as one of the central characteristics and drivers of self-disclosure, i.e. we disclose information because we want others to disclose in turn (Contena, Loscalzo, & Taddei, 2015; Cook et al., 2013; Greene et al., 2006; Kollock, 1999; Taddicken, 2014). For example, Barak and Gluck-Ofri (2007) found positive correlations between the measures of self-disclosure in messages and responses to them in discussion forums and Joinson (2001) found that participants in a study divulged a higher quantity of information about themselves when they had received some self-disclosing information about the experimenter beforehand (albeit their answers were not more revealing or intimate). More recent work further corroborates this positive relationship demonstrating that SNS features showing large quantities of other users disclosing increased self-disclosure (Trepte et al., 2020). In fact, reciprocity has been established as a distinct gratification users attain from using SNSs and as an antecedent of SNS adoption (Pai & Arnott, 2013). Further research has revealed a positive relation between receiving a great number of likes and comments from Facebook friends and the level of life satisfaction (Mayol & Pénard, 2017). On the other hand, receiving few responses from one's Facebook friends was found to threaten the needs for belonging, self-esteem, control, and meaningful existence (Greitemeyer, Mügge, & Bollermann, 2014).

Proponents of the rational choice approach for explaining interpersonal communication have put expected reciprocity at the heart of people's decision-making process. This approach argues that all social phenomena can be explained as the aggregation of discrete, isolated decisions made by individuals, and that these individuals behave as rational actors pursuing their own self-interest (Scott, 2000; Sen, 1997). At the level of these isolated decisions of interaction, behavior is considered to be dominated by the expectation of reciprocity (Kollock, 1999). Thus, this assumption effectively argues that online interactions are predominantly selfish, i.e. motivated by the expectation of reciprocity from the recipient, and doubts the possibility of otherwise altruistic motivations, i.e. without the expectation of reciprocity. Further work, however, has argued that a rigid, direct application of this cost-benefit analysis underplays the importance of many factors that influence our online behavior, and that privacy and disclosure online are, in fact, contextually determined (Nissenbaum, 2009; Pelaprat & Brown, 2012; Quinn & Papacharissi, 2018). Pelaprat and Brown (2012), for example, refer to concepts such as culture, history, relationships and moral commitments that may subvert this assumption of a self-interested rational communicator. A recent literature review of information disclosure on SNSs finds that this rational-actor approach based on social exchange theory currently represents the dominant theoretical perspective, however identifies hints that suggest developing a more holistic approach to account for distortions in the rational thinking when exploring disclosure behaviors on SNSs (Abramova et al., 2017).

2.3.3 TIE STRENGTH

Tie strength was introduced by Granovetter (1973) as a combination of the amount of time, emotional intensity, intimacy (measured as mutual confiding), and reciprocal services devoted to a relationship, with all these factors being independent but correlated. More simply, tie strength can refer to the bonding level or closeness between two people and a tie is typically characterized as *strong* or *weak*. Strong ties are the people that are structurally (Ellison et al., 2007; Friedkin, 1980) and emotionally (Marsden & Campbell, 1984; Wellman & Wortley, 1990) close to someone, such as family and close friends, while weak ties are looser or shallower relationships, i.e. acquaintances. With regards to computer-mediated communication, research has argued that strong ties can influence each other to adapt and expand their use of media to support the exchanges important to their tie, but weak ties are dependent on common means of communication and protocols established by others (Haythornthwaite, 2002). More recent research has examined how the dimensions of tie strength map onto social media usage (Gilbert & Karahalios, 2009; Jones et al., 2013; Luarn & Chiu, 2015).

Early seminal research has shown clear and distinct benefits from communicating both with strong and weak ties. Granovetter (1973) demonstrated the value of weak ties; because they are in contact with different social circles, they can be bearers of novel information and can be useful in tasks such as looking for a job. Wellman and Wortley (1990) illustrated the value of strong ties for the provision of different kinds of social support, such as emotional aid, small services, and companionship. Interestingly, more recent studies have provided evidence of a more nuanced and tangled view of the effects of tie strength on SNSs. While sociological studies have indicated that weak ties can provide better and more novel information (e.g., Granovetter, 1973), answers to questions that were asked through the status message feature of Facebook from strong ties provided a subtle increase in useful and novel information over answers from weak ties (Panovich, Miller, & Karger, 2012). Communication with strong ties was also found to be more predictive of finding employment within three months than communication with weak ties (Burke & Kraut, 2013). The same study found that communication with strong ties over social media has been generally associated with improvements in stress levels, social support, and bridging social capital. Tie strength was positively associated with the feeling of happiness and benign envy when browsing Facebook, as opposed to malicious envy which was found to be independent of tie strength (R. Lin & Utz, 2015). Weak ties, on the other hand, play an important role for information diffusion in SNSs due to the bridge structural effect in the network (J. Zhao, Wu, & Feng, 2011). Bearing in mind the nuances described above, the literature, for the most part, suggests a positive connection between tie strength and the motivation and action of communicating and sharing information online (Haythornthwaite, 2002; Y.-C. Wang, Burke, & Kraut, 2016), a relationship that also holds for the sharing of photographs specifically (Gilbert & Karahalios, 2009; Mendelson & Papacharissi, 2010).

At the same time, Facebook users will be ostensibly more interested in receiving communication from their closest friends, indicating a link between tie strength and expected reciprocity. In fact, Granovetter's definition of tie strength makes a reference to the "reciprocal services which characterize a tie" as a factor in building, maintaining, and measuring tie strength (Granovetter, 1973). Reciprocity has been linked to SNS members' common ground (Pai & Arnott, 2013), which is a significant factor of tie strength, while the mutual exchange of wall posts has been used for the computational calculation of tie strength in data mining studies (Alhazmi & Gokhale, 2016). Furthermore, if we consider question asking as a form of self-disclosure, since the fact that one is interested in something is information about them, then eliciting answers to questions on SNSs also constitutes disclosure with an expectation of reciprocity. In this case, tie strength has been also found to affect reciprocity; a survey study of status message Q&A behavior on SNSs found that closeness of a friendship was a motivator to answer questions (Morris, Teevan, & Panovich, 2010b) and a small study comparing information seeking between search engines and question asking on Facebook found that many participants' questions were answered by friends they rated as close (Morris, Teevan, & Panovich, 2010a). It is worth noting, however, that while a positive link between tie strength and expected reciprocity seems intuitive, researchers very early showed that the connection is more nuanced; Altman (1973) noted that the norm of disclosure reciprocity may be stronger early in a relationship than in later stages, and Derlega and colleagues (1976) reported that strangers display more disclosure reciprocity than friends in a social encounter.

2.3.4 INTERPERSONAL TRUST

Trust has been characterized as an integral part of human interactions, as it allows people to engage in exchanges that leave both parties better off, as well as reduces the cost of these transactions (Resnick, 2002; Riegelsberger, Sasse, & Mccarthy, 2005). Golbeck and Hendler (2006) have provided a definition of interpersonal trust that is particularly suitable for characterizing relationships on SNSs, explaining that "trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome". Reputation has been described as a useful and important tool for determining the trustworthiness of another person for internet interactions (Cheshire & Cook, 2004), however such interpersonal trust is inherently a personal opinion that can be influenced by several factors, such as past experiences with the other person and their friends, our opinions of actions the person has taken, rumors, and influence by others' opinions (Golbeck, 2005).

A number of studies have consistently shown that trust is a necessary condition for disclosing information and has a positive effect on disclosure either in the case of a website or organization (Mesch, 2012; Yanbo Wang, Min, & Han, 2016; Zimmer, Arsal, Al-Marzouq, & Grover, 2010), or in the case of dyadic relationships offline (Wheeless & Grotz, 1977) and on SNSs (Millham & Atkin, 2016; Sheldon, 2009). Researchers, however, have pointed out that the relationship between trust and self-disclosure may be more complex, suggesting that trust has a mediating or moderating effect on the relationship between privacy and self-disclosure (Joinson, Reips, Buchanan, & Schofield, 2010; Taddei & Contena, 2013). This means that trust can reduce perceived privacy risks, thereby encouraging SNS users to engage in more disclosure behaviors and in the sharing of more personal information with people they trust (X. Chen, Pan, & Guo, 2016; Dwyer, Hiltz, & Passerini, 2007; Zimmer et al., 2010). Thus, a high degree of trust in the recipient of the disclosure should be even more important in risky situations, such as sharing content that can be more sensitive in nature, like photographs (Gilbert & Karahalios, 2009; Malik, Hiekkanen, Dhir, & Nieminen, 2016). Research has also studied the link between interpersonal trust and expected reciprocity, as Resnick (2002) explains that "[a]n expectation of continued interaction in the future is helpful in maintaining trust". Pai and Arnott (2013) expand on this link and argue that without some level of trust in the reciprocity of others, SNS users are reluctant to use the platform for communications that are, to a large extent, highly personal and revealing.

Chapter 3. Studying the Individual: Motivations and Media Selection

3.1 Introduction

This chapter describes the work conducted in the context of what I described earlier as Study 1 and consists of two parts, Study 1A and Study 1B. This thread of research studies the combination of traditional social science and computational approaches for the study of motivations for using SNSs. Study 1A identifies motives for Facebook use by employing the U&G framework and investigates the extent to which these motives can be predicted through usage and network metrics collected via the Facebook API. Study 1B similarly draws from the U&G framework and combines survey and behavioral data, but aims at comparing motivations across Facebook and Twitter to understand media selection and examines the findings in the context of technology non-use.

Although they seek to address different research questions, both studies shared the same data collection process, and thus, for reasons of simplicity and presentation, they are combined in one chapter. Section 3.2 describes the overall set-up, detailing the data collection procedure, the study participants, the survey content and the behavioral data collected. Section 3.3 introduces Study 1A, provides the motivation for the study and the research questions, presents the results and discusses the findings and the implications of the work. Accordingly, Section 3.4 introduces Study 1B, provides the motivation for the study and the research questions, presents the results and discusses the findings and the implications of the work.

Study 1A reports primarily on the work presented in (Spiliotopoulos & Oakley, 2013b), but also includes some findings and discussion from (Spiliotopoulos & Oakley, 2013a), (Spiliotopoulos, Karnik, Oakley, Venkatanathan, & Nisi, 2013) and (Spiliotopoulos & Oakley, 2012), as well as additional/alternative analyses that have not been presented elsewhere. Study 1B reports primarily on work that was very recently accepted for publication (Spiliotopoulos & Oakley, 2020), while preliminary findings were presented and discussed in (Spiliotopoulos & Oakley, 2015) and (Spiliotopoulos & Oakley, 2016).

3.2 METHOD AND STUDY SET-UP

3.2.1 PROCEDURE

Participants were recruited with a request to complete an online survey. Approximately 1/3 of participants were recruited via posts on SNSs, 1/3 via posts to online forums, mailing lists and online study repositories, and 1/3 via a Facebook ad campaign. The Facebook campaign consisted of two ads with similar wording targeted at self-reported English-speaking Facebook users from 12 countries. Facebook automatically manages the visibility of ads in an auction-like way. Thus, the Facebook ads resulted in the recruitment of a relatively large number of users from India, possibly due to the lower cost (and therefore higher frequency) of ads distributed to this group. The Facebook ads overall had a 0.059% click-through rate.

The participants were recruited by answering a request to complete an online survey and were directed to a comprehensive study description page that clearly framed the experiment as an academic study, explained the data collection process, provided the contact details of the researchers, and requested users' consent. The description page contained a link that invited participants to login with their Facebook credentials and access the survey, an action that is equivalent to installing a Facebook application. In addition to our description, Facebook displays all data-access permissions granted to an application during installation, thus ensuring that the participants had a comprehensive account of the data captured by the study. The app required the basic data-access permission (as defined by the Facebook API) and two extended permissions: access to newsfeed and friendlists. A number of participants (25.5%) refused the extended permissions, and so these variables were excluded from the analysis. 67.1% of the people that clicked the link to go to the app accepted the "basic info" permission dialog. Participants whose responses exhibited discrepancy between the demographic variables (e.g., gender, age) that were collected through the API and those reported by them in the survey were considered unreliable and removed. This resulted in an 8% discard rate. Figure 3-1 shows the two Facebook ads that were used for recruiting participants and Figure 3-2 shows the basic permission dialog.

After logging in, participants were directed to a survey capturing demographics and their motivations for using Facebook and were, then, prompted to answer whether they were also

Twitter users. If the reply was positive, they were presented with an additional set of questions eliciting their motivations for using Twitter. In the background, a number of metrics about each participant's actual Facebook usage were collected with the use of the publicly available Facebook API. The participants had the choice to opt out of the study at any time.

FIGURE 3-1. THE TWO FACEBOOK ADS USED FOR RECRUITING PARTICIPANTS

FIGURE 3-2. THE BASIC PERMISSION DIALOG

3.2.2 PARTICIPANTS

A total of 232 usable responses were collected, with 126 (54.3%) being male and 106 (45.7%) female. The participants reported a mean age of 23.9 years (median = 20, SD = 8.68) and came from 32 different countries, with 94 (40.5%) from the USA and 70 (30.2%) from India. The majority of the sample (n = 174, 75%) were full-time students, 51 (22%) were employed, and 7 (3%) unemployed. Most participants (82.7%, n = 192) were daily users of Facebook, with approximately half of this group (n = 95) reporting using the site multiple times each day. On the days that they use Facebook, participants reported spending a mean of 78 (median = 45, SD = 97.4) minutes on the site. Out of the 232 participants in the study, 103 (44.4%) reported using Twitter. On the days that they use Twitter, participants reported spending a mean of 29.1 (median = 15, SD = 42.9) minutes on the service. 33% (n = 34) of the Twitter sample reported being daily Twitter users with approximately half of them (n = 18) using the site multiple times per day.

3.2.3 MEASURES

3.2.3.1 Motivations for Facebook and Twitter use

Motivations for using Facebook were measured by presenting participants with a list of 28 statements based on Joinson (2008) and asking them to answer "How important are the following uses of Facebook to you personally?" on a 7-point Likert scale from "very unimportant" to "very important". Similarly, motivations for Twitter use were measured with the set of 15 items from Johnson and Yang (2009) and the question "How important are the following uses of Twitter to you personally?".

3.2.3.2 Behavioral data

The Facebook API was used to access a range of usage information for each participant. In addition, the participant's Facebook friendship network was also collected via the application. This is essentially a 1.5-degree egocentric network (i.e., the friends and all the mutual friendships among them) with ego (i.e., the participant) removed. Table 3-1 presents descriptive statistics for the demographics, usage, and network data collected.

	Mean	Median	SD
Age	23.9	20	8.7
Time spent on site (mins/day)	78	45	97.4
Facebook usage metrics			
Groups joined	22.7	11	31.6
Events currently attending	1.37	0.5	2.25
Check-ins posted	2.84	0	6.35
Likes given (to pages)	337.6	129.5	542.8
Interests/activities mentioned	13.3	3	34.5
Photos uploaded	331.1	161.5	431.3
Photo albums uploaded	13.7	12	8.0
Photos tagged in[1]	84.7	35	252.2
Status updates posted[2]	62.8	21.5	115.5
Comments made	96.7	43.5	145.1
Likes given (to posts)	180.8	70.5	320.3
Links posted[2]	55.8	8	173.9
Questions posted	0.38	0	1.35
Network metrics[3]			
Size (nodes)	427	362.5	295.3
Average degree	55.5	30.6	59.2
Diameter	7.1	7	2.2
Density	0.132	0.111	0.092
Modularity	0.40	0.41	0.17
Connected components	14.7	9	32.1
Average clustering coefficient	0.56	0.56	0.089
Average path length	2.60	2.45	0.66

[1]in the past 12 months, [2]in the past 6 months

[3]based on the personal networks with ego and their ties removed

TABLE 3-1. DEMOGRAPHICS, USAGE AND NETWORK METRICS COLLECTED

3.3 Study 1A: Understanding Facebook Motivations with the Inclusion of Activity and Network Metrics

3.3.1 Introduction

The U&G framework is a well-established, popular and effective way of studying media, including social media. The U&G approach essentially studies media by studying the individuals that use them. As such, the U&G frame of analysis 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). A typical U&G study employs a survey instrument (or occasionally interviews or focus groups) for the collection of all relevant data. However, as a theoretical framework, U&G does not mandate that any particular empirical methods be used and, therefore, this study examines the inclusion of computationally captured data in the U&G framework of analysis.

Of course, motives are subjective and have to be solicited with the use of a survey or similar instrument. However, the Facebook API can be a rich and potentially useful source of data that can describe the outcomes and the antecedents. In the study described below, I investigate the value of including API-collected usage metrics as behavioral outcomes, and personal network metrics as antecedents. Thus, in effect, I am proposing and evaluating an extended U&G frame of analysis that includes these additional measures. Based on the above, this study has been guided by the following research questions:

1) Are computationally collected *activity* metrics a useful addition to the U&G frame of analysis? In particular, does the addition of activity metrics as behavioral outcomes explain additional variance in the regression models predicting motives for Facebook use?
2) Are computationally collected *network* metrics a useful addition to the U&G frame of analysis? In particular, does the addition of personal network metrics as antecedents explain additional variance in the regression models predicting motives for Facebook use?
3) How well can we predict motives for using Facebook from API-collected data?

3.3.2 Data Analysis and Results

Exploratory factor analysis based on the items used in previous literature (Joinson, 2008) led to the identification of the motives for Facebook use. The scores for each factor were calculated for each participant, and then a series of multiple regressions were carried out, in order to investigate the effect of Facebook usage metrics and network metrics on the motives for Facebook use. [1]

3.3.2.1 Identifying motives of Facebook use

In order to identify the motives for Facebook use, we conducted an exploratory factor analysis with orthogonal rotation (varimax) on the 28 items corresponding to the Facebook questions. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .856. This value confirms the sample size as "great" (Field, 2009; Kaiser, 1974) for this analysis. Bartlett's test of sphericity χ^2 (378) = 3491, p < .001, indicated that correlations between items were sufficiently large (Field, 2009). Seven factors were found with eigenvalues over Kaiser's criterion of 1, explaining in combination 68% of the variance. The seven factors exhibited good reliability (Cronbach's α values ranged from .717 to .900). A cut-off value of .5 for factor loadings led to the exclusion of three items (*using advanced search to look for specific types of people*, *receiving a friend request* and *meeting new people*) that did not load highly on a factor or loaded highly on two or more factors. Table 3-2 shows the factor loadings after rotation.

[1] A note on the sample and the metrics: The data collection produced a sample of 232, but due to a technical issue we collected the network data of only 208. Thus, for Study 1A we use the sample of 208 and for Study 1B the full sample of 232. However, for reasons of simplicity, presentation and consistency across the dissertation, as well as for taking advantage of the slightly larger sample, I report the results of Study 1A in a slightly different manner than what was originally reported in (Spiliotopoulos & Oakley, 2013b). In particular, I report the results of the factor analysis for N = 232 here and use these results to calculate the factor scores for the regressions. Furthermore, I use a lower cut-off value for factor loadings compared to the published work (.5 versus .6) and a slightly updated set of activity metrics. Hence, there are some small discrepancies in the results compared to the published results in (Spiliotopoulos & Oakley, 2013b). The combined effect of these changes is rather small; the results are similar and the findings identical.

Items	Mean	SD	Factor Loadings
Entertainment/Content (α = .886)			
Applications within Facebook	2.58	1.79	.856
Playing games	2.05	1.72	.826
Discovering apps because you see friends have added them	2.20	1.60	.817
Quizzes	2.14	1.58	.782
Photographs (α = .876)			
Being tagged in photographs	3.57	1.95	.861
Tagging photographs	3.29	1.87	.836
Sharing / posting photographs	4.32	1.85	.756
Viewing photographs	4.88	1.58	.652
Social Network Surfing (α = .900)			
Looking at the profiles of people you don`t know	2.70	1.90	.823
Viewing other people`s friends	2.94	1.84	.815
Browsing your friends` friends	2.93	1.83	.775
Social Connection (α = .788)			
Connecting with people you otherwise would have lost contact with	5.19	1.54	.772
Reconnecting with people you`ve lost contact with	4.90	1.71	.738
Finding people you haven't seen for a while	4.81	1.54	.715
Finding out what old friends are doing now	4.56	1.58	.626
Maintaining relationships with people you may not get to see very often	5.60	1.42	.564
Contacting friends who are away from home	5.65	1.47	.551
Shared Identities (α = .769)			
Organizing or joining events	3.69	1.92	.815
Joining groups	3.15	1.83	.799
Communication with likeminded people	3.86	1.98	.692
Status Updates (α = .785)			
Seeing what people have put as their status	4.41	1.76	.760
The news feed	5.00	1.77	.688
Updating your own status	4.13	1.95	.577
Social Investigation (α = .717)			
Virtual people-watching	2.96	1.96	.749
Stalking other people	2.63	1.96	.677

Notes. KMO = .856. All items shared a common prompt: "How important are the following uses of Facebook to you personally?" and were measured with a 7-point Likert-type scale ranging from "very unimportant" to "very important". The factors are ordered based on variance explained.

TABLE 3-2. SUMMARY OF FACTORS AND INDIVIDUAL ITEMS DESCRIBING MOTIVES FOR FACEBOOK USE

3.3.2.2 Predicting Facebook motives

A series of hierarchical multiple regressions (forced entry method) were conducted with the seven motives (i.e., factor scores) of Facebook use as outcome variables. Age, reported time spent on site, gender (male = 1), occupation (recoded as a dichotomous variable, student = 1) and nationality (recoded as a dichotomous variable, USA = 1) were included in the first step. The Facebook usage metrics were added in the second step, and the personal network metrics were added in the third step. The goal of this design is to tease out the effects of the usage metrics (step 2) and the network metrics (step 3) in predicting the motives for Facebook use. This is done by contrasting the explanatory value (i.e., the variance explained by the model or R^2) between the steps.

The correlation matrix revealed a number of strong relationships among the predictor variables, but only the relationship between *likes given (to posts)* and *comments made* exceeded the .8 benchmark that indicates potential multicollinearity ($r = .822$, $p < .001$), leading to the exclusion of the former variable from the regression analysis. The next highest correlation found was between *network diameter* and *average path length* ($r = .789$, $p < .001$), which is to be expected as both metrics rely on path length, but indicate a different distribution of path lengths in a network. As they refer to distinct network properties, it was decided to keep both variables in the analysis. Furthermore, examination of the Variance Inflation Factor (VIF) for every predictor variable found a highest value of VIF = 6.54, which is well below the benchmark value of 10 that indicates multicollinearity. Therefore, we are confident that the regressions carried out were free from multicollinearity concerns.

Table 3-3 shows the results of the regressions. For presentation purposes, only the predictor beta coefficients of the final step are shown, but the model statistics are shown for each step.

Predictors	Motives for Facebook use						
	Entertainment /Content	**Photographs**	**Social Network Surfing**	**Social Connection**	**Shared Identities**	**Status Updates**	**Social Investigation**
Model intercept	0.995	-1.276	1.725*	-0.615	-1.591*	0.831	-0.601
Age	.020	.104	-.103	-.003	.357***	-.115	-.073
Time spent on site	.039	-.068	.145^{+}	.020	-.053	.059	.201*
Gender (male)	.014	.084	.183*	-.194*	.091	.100	.006
Occupation (student)	.105	.173^{+}	-.086	.042	.113	-.100	-.080
Nationality (USA)	-.141	.188*	-.182*	-.056	-.212*	.031	.120
Model significance (F value)	4.39***	4.13**	4.77***	2.31*	5.84***	1.20	1.60
ΔR^2 and significance	.098***	.093**	.106***	.054*	.126***	.029	.038
Groups joined	-.093	-.051	-.026	.040	.082	-.081	.035
Events currently attending	-.017	.067	.005	.126	.153^{+}	-.040	.007
Check-ins posted	-.033	-.029	.018	-.045	-.140^{+}	.067	-.026
Likes given (to pages)	.149^{+}	.053	-.131	-.019	.086	-.226*	.037
Interests/activities mentioned	.012	-.010	.124	-.002	-.087	-.044	.002
Photos uploaded	-.343***	.161	.033	.009	.052	.013	.148
Photo albums uploaded	.162	.218*	-.136	.009	.051	.040	.005
Photos tagged in	.029	.046	.002	.021	-.060	.042	.090
Status updates posted	-.108	.015	-.195*	.044	-.132	.302***	-.194*
Comments made	-.087	.007	-.040	-.036	.020	.065	.042
Links posted	-.006	-.040	.096	.006	.155*	.049	-.108
Questions posted	.100	-.125^{+}	-.091	.003	.012	-.001	.042
Model significance (F value)	3.66***	2.80***	2.44**	1.17	3.18***	1.93*	1.46
ΔR^2 and significance	.149***	.108**	.074	.040	.095*	.118*	.077
Total R^2	.247	.200	.179	.094	.222	.147	.115
Network size	.011	-.106	.014	.283^{+}	.022	.010	-.193
Average degree	-.098	.156	.066	-.276^{+}	.044	-.157	.236
Diameter	-.079	-.144	-.028	.060	.039	.023	.191
Density	.010	-.027	.090	.125	-.194	.056	-.074
Modularity	-.105	.135	.046	.182	-.111	-.158	.129
Connected components	.030	.046	-.031	-.086	.161^{+}	-.186*	.138
Average clustering coefficient	-.172	.024	-.206^{+}	-.045	.203^{+}	-.115	.151
Average path length	.134	-.001	-.013	.000	-.180	.132	-.296^{+}
Model significance (F value)	2.94***	2.13**	1.91**	1.20	2.54***	1.77*	1.42^{+}
ΔR^2 and significance	.041	.026	.028	.047	.037	.048	.048
Total R^2	.288	.226	.208	.142	.258	.196	.163

Notes. $^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$, beta coefficients are standardized

TABLE 3-3. HIERARCHICAL MULTIPLE REGRESSION MODELS COMPARING THE EFFECTS OF DEMOGRAPHICS, FACEBOOK USAGE MEASURES AND NETWORK MEASURES FOR THE PREDICTION OF MOTIVES FOR FACEBOOK USE

3.3.3 Discussion

3.3.3.1 Motives for Facebook use

The exploratory factor analysis yielded seven factors, corresponding to motives for Facebook use, which are similar to those identified in the previous literature[1] (Joinson, 2008). This was expected, since the same set of items were used. The differences between the factors identified in the two studies are in the three items that did not load clearly.

3.3.3.2 Predicting the motives for Facebook use with the inclusion of behavioral data as outcomes and network data as antecedents

Overall, seven hierarchical models were tested for predicting the seven motives for Facebook use. At step 1 of each regression, which included only the five demographic variables, five of the seven models were statistically significant (i.e., $p < .05$). The models predicting the motives of *status updates* and *social investigation* were not predicted by the demographics in a statistically significant way. The variance explained (i.e., the value of R^2) by the predictors in step 1 ranged from .054 to .126 for the significant models.

Addition of the 12 activity variables in step 2 increased the variance explained substantially. After step 2, again five of the seven models were significant. However, the model predicting the motive of *social connection* was not significant any more, while the model predicting the motive of *status updates* became significant. The incremental variance of step 2 was significant at the $p < .05$ level in four of the seven cases, with this incremental variance (i.e., the value of ΔR^2) ranging from .095 to .149 for the statistically significant cases.

Addition of the eight network variables in step 3 did not produce major changes in the significance of the overall models; the same five out of the seven models at step 3 were significant at the $p < .05$ level, although the model predicting the motive of *social investigation* also became significant at the $p < .1$ level. However, in all seven cases the incremental variance resulting from the addition of the network metrics at step 3 was not statistically significant.

[1] In Study 1B that extends this analysis we decided to rename the *content* factor from the previous literature to *entertainment/content*. During the reviewing process for our follow-up paper (currently under review) we received a comment from a reviewer stressing that this motive is largely related to entertainment and that would be a more apt description of the motive (as opposed to *content* which is more feature-related). Indeed, entertainment is not explicitly represented as a motive in this analysis and it is a common motive identified in such studies, including our own Twitter U&G study described later. Although in our paper that discusses this work (Spiliotopoulos & Oakley, 2013b) we used the original name for the factor, I retroactively made the change in the dissertation for reasons of consistency and presentation.

The above findings can provide some high-level answers to the three research questions of this study. We see that the demographics, which are typically used as antecedents in U&G studies, are generally significant predictors of motives for using Facebook indeed (at least in five of the seven cases), albeit rather weak in terms of explaining variance in the models. Addition of the activity metrics increased the explanatory value of the models significantly and substantially. On average, addition of the activity metrics more than doubled the variance explained by the models. We find that a few isolated network metrics emerged as significant predictors and, overall, addition of the network metrics in step 3 slightly raised the variance explained by the models and did not harm the significance of the models. However, this change in explained variance was not statistically significant in any of the seven models. This may be due to the overall exploratory nature of this research, i.e. we included network metrics that are generally meaningful in network research without relying on specific hypotheses. A more focused approach that would associate specific metrics with expected findings and strategically select metrics for inclusion may provide statistically significant results, especially if coupled with a larger and more homogeneous sample.

3.3.3.3 Specific effects of Facebook usage, social and network antecedents on the motives for Facebook use

The *entertainment/content* motive, which includes items for Facebook applications and games, was strongly and negatively associated with only one predictor variable: uploaded photographs. This highlights the possibility of a user population on Facebook that is focused on highly interactive content and disinclined to use and share more traditional media. This finding also reinforces the notion that Facebook uses can be very distinct and that there is a need to differentiate among particular uses when examining the site (Smock et al., 2011).

Participants from the USA were positively correlated with the *photographs* motive, pointing perhaps to the high diffusion of camera-equipped smartphones in that market. Being a student was also a positive predictor of this factor. Interestingly, the number of photo albums uploaded emerged as a significant predictor, whereas the number of photos uploaded was (marginally) not significant. In a follow-up analysis (not presented), when the number of albums was removed from the model, the number of photos emerged as a very significant predictor. This indicates that, while the two variables share a lot of variance, the number of albums is a better predictor for this motive, possibly demonstrating that people who are really interested in photographs organize them carefully in albums.

Gender emerged as a significant predictor of both the *social connection* and the *social network surfing* motives, albeit in opposite directions. Females were associated with the *social connection* motive (as in (Joinson, 2008)), the items of which indicate connections and links to past relationships. On the other hand, males were associated with the factor whose items indicate a tendency for acquiring more information about acquaintances or strangers. Network size, i.e., the number of friends, was also positively correlated with the *social connection* motive; users interested in connecting with others tend to have larger networks. However, average degree of the network was negatively associated with this factor, suggesting that users motivated by social connection have a more spread-out network with friendships that tend not to overlap.

Older participants and those from outside the USA were more motivated by the opportunity to be associated with like-minded individuals, as described by the *shared identities* factor. The number of events attending on Facebook was a significant positive predictor, suggesting that users attempt to gratify this need for connecting with like-minded individuals by using this feature and broadcasting their interest in specific events where they might meet such people. Interestingly, the number of check-ins posted was negatively associated with this motive. This indicates that although the *shared identities* motive seeks to gratify a need for connecting with specific people, advertising or showing off one's presence in a particular location does not express or gratify this need. The number of links posted was positively correlated with this factor, illustrating that (re)sharing information can be a way of connecting with like-minded people. Interestingly, two network measures were found to have a significant positive effect on this motive: the number of connected components and the average clustering coefficient. The former has been interpreted as the number of an individual's social contexts (Brooks et al., 2011; Ugander et al., 2012), and in this sense explains the motivation of these people to belong to distinct groups. A high average clustering coefficient is an indication of networks with modular structure and, at the same time, small distance among the different nodes; in other words, like-minded people will tend to form groups and attend events (based on their similar interests) and will tend not to engage in isolated friendships. In all, the model for the *shared identities* motive has significant predictors from all three variable types, accounting for 25.8% of the variation.

The motive of *status updates* has two significant usage predictors, "likes" given to pages and status updates posted. It is worth noting that these two major and popular Facebook features predict this motive in opposite direction, again reinforcing the idea that it is important to unbundle Facebook usage to its respective features (Smock et al., 2011). For example, the use of likes may indicate someone who tends to respond more to media clips rather than status updates, which, in turn, may seem more appealing to users interested in conversation. Furthermore, the number of connected components in a user's personal network was negatively correlated with this motive. As component count has been viewed as a measure of structural diversity (Ugander et al., 2012), with each component hinting at a distinct social context, this correlation may indicate that Facebook users with a very large number of diverse social groups get less value from their newsfeed - it may be overloaded, or the content too wide-ranging and tertiary to be of substantial interest.

Time on site was positively associated with the *social investigation* motive, possibly suggesting that this kind of activity can be "addictive" and occupy large amounts of time. On the other hand, the number of status updates posted was negatively associated with this motive, as well as with *social network surfing*. This reinforces the notion of a distinction between users who are interested in contributing content to the site and those that are not, e.g. lurkers (Lampe, Wash, Velasquez, & Ozkaya, 2010).

Looking at the overall picture of the analysis, it stands out that the number of status updates emerged as a significant predictor for 3 out of the 7 motives for Facebook use. This suggests that this feature remains one of the most important aspects on the site, despite the continuous inclusion of new functionality, the shift in the demographics of users and the general evolving ecosystem of Facebook.

The size of a Facebook user's personal network emerged as a significant predictor for one of the seven factors, even though it has traditionally been the most common, and usually the only, network measure in SNS studies. Three more sophisticated network measures, the number of connected components, the average clustering coefficient and the average path length also show a significant effect on motives for use. Thus, the impact of the network size appears to have been lessened with the introduction of more complex network measures, suggesting they capture aspects of the structure that are more important and meaningful for understanding motives.

Finally, recent research has suggested that appropriate use of network analysis depends on choosing the right network representation for the problem at hand (Butts, 2009). Indeed, a previous study of the different "connection strategies" among Facebook users has found that they differentiate between all Facebook friends and "actual" friends as approximately 25% of that total (Ellison, Steinfield, & Lampe, 2011). Since the underlying relations (i.e., Facebook friendships) of networks can vary substantially, it may be that standard network metrics are not directly comparable across Facebook users. Taking the idea of systematically introducing personal network measures in studies of SNS motives a step further, it may be valuable to study alternative network representations, such as those whose links are weighted based on tie strength

(Gilbert & Karahalios, 2009) or interpersonal trust (Golbeck & Hendler, 2006). Such networks may result in metrics and analyses with greater explanatory power.

3.3.3.4 Theoretical and methodological contributions to U&G

Although the U&G framework has been used extensively in the communications sciences, one of its main criticisms is that it relies heavily on self-reported data (Katz et al., 1973; Papacharissi, 2008). This study addressed this limitation by eliciting extensive data about the patterns of use and several social and network antecedents programmatically through the Facebook API. These data should be more accurate than self-reported data about usage or network structure, as well as free from possible cognitive and recall biases.

In fact, previous research (Smock et al., 2011) revealed that users' motivations for using Facebook predict their use of different features, such as status updates and wall posts, but features that share similar capabilities do not necessarily share underlying motivations for use. When these results are contrasted against models employing unidimensional measures of Facebook use, differences were found between motivations for both general Facebook use and that of specific site features. This suggests that unidimensional measures of SNS use obfuscate motivations for using specific features. The current study took this analytic approach further by looking not only at the reported use of specific Facebook features, but by examining a broad range of Facebook usage data. In particular, a comprehensive set of data corresponding to Facebook usage was gathered computationally, comprising 12 distinct variables as opposed to the one or two variables (time on site, frequency of visits) that are typically gathered through self-reports in similar studies.

Furthermore, this study expanded the methodological arsenal of U&G studies by leveraging the Facebook API to gather a set of data that is by far larger and more diverse than that in a typical U&G study. In addition, the network structure was gathered and eight representative network metrics were computed for each participant. This introduced the network antecedent as a possible consideration in the U&G frame of analysis, next to the social and psychological antecedents usually employed.

As a result, the activity metrics, to a larger extent, and the network metrics, to a smaller extent, increased the explanatory value of the models and at least one predictor variable for every motive was found to have a significant effect. Overall, all three types of predictor variables - social antecedents, usage metrics, and personal network measures - were useful in predicting motives and identifying trends that point to future research, supporting the validity of this broad data-centric approach.

3.3.3.5 Advantages and limitations of the sampling procedure

The sampling procedure that was employed resulted in a participant sample that exhibited certain particularities. The combination of recruitment methods led to a sample that was diverse in terms of demographic and geographic distribution, compared to similar studies that typically take place within universities and study students. Since motives for Facebook use will likely vary substantially across cultures, ages, and educational backgrounds, the diversity of the sample used in this work may better match the traditionally exploratory nature of U&G studies.

However, as with other web-based survey studies, the current work was subject to a self-selection bias. Basically, the group of people who opted to participate in the study may not adequately represent typical users. This bias may have been strengthened by the study's requirement that participants install a Facebook application that openly admitted it would access personal details; many users may have been frightened off. On the other hand, these same processes may have discouraged spurious participants (e.g., careless, dishonest, or mischievous web surfers). These advantages and limitations, common to similar studies (e.g., Quercia et al., 2012), pose interesting implications for future work using the Facebook API or comparable data-intensive techniques.

3.3.3.6 Practical implications

Typically, in a U&G study, after the gratifications are gathered, the analysis examines the effect of the social/psychological antecedents and gratifications on the uses. However, since this analysis is purely correlational, it is methodologically sound to reverse the directionality of analysis and attempt to predict the gratifications from the variables describing antecedents and uses, which is the approach adopted in the current work.

In this study, a number of predictor variables that can be collected and measured automatically by an API were used to establish potentially predictive links to valuable subjective data that can only be collected via a survey instrument. In particular, the motives for Facebook use that were the outcome of this analysis can 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, advertisements of digital cameras can be shown to users who score highly on the *photographs* motive, or applications, games and online services can be suggested to users interested in *entertainment/content*. In addition, opportunities for social connection can be shown more prominently to users interested in connecting and interest- or event-based recommendations may more effectively target people scoring highly on the *shared identities* factor.

The study found users with large numbers of connected components (i.e., separate social contexts) to be less motivated to use their feeds, independently of overall network size. This hints at information overload – a problem that needs to be addressed in future versions of this feature (Horrigan, 2016). Furthermore, status updates were also negatively associated with two motives, *social investigation* and *social network surfing*. This suggests that individuals who post few status updates are not necessarily inactive on this site, but may be enthusiastic and regular users aiming to achieve specific, largely observational, goals.

Motives of use can also provide useful insights for features to incorporate into future system designs. For instance, motives can be directly incorporated into user personas in the requirements analysis and design phase of systems, leading to richer creative artifacts. In fact, some very recent work has demonstrated a way of developing personas from social media data (Salminen et al., 2018). On the interface level, adaptive systems 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.

3.3.4 CONCLUSION

This study set out to answer the first sub-question of the second overarching research question of the dissertation, i.e. RQ2a or whether a combination of survey and Facebook-API collected data can provide additional value and insights when predicting Facebook motivations above what can be gained from each approach in isolation. Overall, we found support for this combination of approaches, as the addition of the usage metrics as U&G outcomes substantially increased the explanatory value of the models. We also found that some isolated network metrics, which were intended as U&G antecedents, were significant predictors.

In terms of the specific research questions posed in the introduction of this study, we find support that both the computationally collected activity metrics and the network metrics can be considered a useful addition to the U&G frame of analysis (questions 1 and 2). Furthermore, we found that five of the seven regression models predicting Facebook motivations were significant at the .05 level, suggesting that there is credence is the claim proposed by question 3, although further work including alternative metrics or selecting specific metrics grounded on theory might produce even more accurate models.

3.4 Study 1B: Combining Motivations and Actual Behavior to Explain Social Media Selection

3.4.1 Introduction

The plurality and diversity of available SNSs that compete for people's time and attention increases the complexity of the decision that users have to make in order to select an appropriate medium to satisfy their needs for communication (e.g., X. Zhao et al., 2016). Comprehending this decision process can provide a more accurate understanding of people's behavior across SNSs and has the potential to inform SNS research, design and management. As explained earlier, one well-established approach for studying this process is to posit that users select the most appropriate service driven by their particular *motives* for use. This is theoretically and empirically grounded in the Uses and Gratifications (U&G) communication perspective, which asserts that people use media actively, purposefully and strategically to fulfill specific needs (Katz et al., 1973; Papacharissi, 2008; Quan-Haase & Young, 2014). Another common approach is to focus on people's *behavior* to determine how usage patterns of a SNS can affect whether someone will also use a different SNS. In this regard, it is becoming increasingly important to pay attention to people's actual behavior, instead of self-reported behavior or behavioral intentions (Junco, 2013).

Thus, this study sets out to answer the following research question:

1) How does the combination of motives and behavioral data explain social media selection?

More particularly, the current study focuses on Facebook and Twitter, two of the most impactful, popular and diffused in the worldwide population SNSs. Research generally suggests that understanding users' motives can provide useful insights into how people navigate the social media ecosystem and how they decide which SNS to use and spend their time on. Further research has highlighted the importance of studying non-use for understanding media selection and suggests utilizing behavioral data and a common sample. Based on this, the study design addresses the following specific research questions:

2) What is the relationship between motivations for using Facebook and motivations for using Twitter for the same users?

3) How do Twitter users and non-users differ in terms of their behavior on Facebook?

3.4.2 *DATA ANALYSIS AND RESULTS*

This study complements study 1A with the inclusion of the motivations for using Twitter and explores the relationships between the motivations of the same users across sites via three regression models. Then we examine the differences in Facebook usage for Twitter users and non-users, based on the computationally collected data.

3.4.2.1 Motivations for Facebook and Twitter use

In order to identify the motives for Facebook use, we conducted an exploratory factor analysis with orthogonal rotation (varimax) on the 28 items corresponding to the Facebook questions[1]. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .856. This value confirms the sample size as "great" (Field, 2009; Kaiser, 1974) for this analysis. Bartlett's test of sphericity χ^2 (378) = 3491, $p < .001$, indicated that correlations between items were sufficiently large (Field, 2009). Seven factors were found with eigenvalues over Kaiser's criterion of 1, explaining in combination 68% of the variance. The seven factors exhibited good reliability (Cronbach's α values ranged from .717 to .900). A cut-off value of .5 for factor loadings led to the exclusion of three items (*using advanced search to look for specific types of people*, *receiving a friend request* and *meeting new people*) that did not load highly on a factor or loaded highly on two or more factors. Table 3-2 on page 38 shows the factor loadings after rotation.

In order to identify the motives for Twitter use, we conducted another exploratory factor analysis with orthogonal rotation (varimax) on the 15 items corresponding to the Twitter questions. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .845. This value confirms the sample size as "great" (Field, 2009; Kaiser, 1974) for this analysis. Bartlett's test of sphericity χ^2 (105) = 784.32, $p < .001$, indicated that correlations between items were sufficiently large (Field, 2009). Three factors were found with eigenvalues over Kaiser's criterion of 1, explaining in combination 63.6% of the variance. The three factors exhibited very good reliability (Cronbach's α values ranged from .805 to .865). A cut-off value of .5 for factor loadings led to the exclusion of one item, *seeing what others are up to*, that loaded highly on two factors, *social* and *entertainment* motives. Table 3-4 shows the factor loadings after rotation.

---

[1] In order to identify the motives for Facebook use, we followed the same procedure as in Study 1A. For convenience and for easy comparison with the next paragraph, I repeat the details of the exploratory factor analysis procedure in this section, but the reader can refer to Table 3-2 earlier for more details on the specific items.

Items	Mean	SD	Factor Loadings
Social (α = .856)			
Meeting new people	3.17	2.02	.785
Participating in discussions	3.18	1.80	.732
Communicating with many people at the same time	4.12	2.02	.694
Keeping in touch with friends or family	3.24	2.13	.680
Communicating more easily	4.23	1.98	.608
Expressing yourself freely	4.63	2.06	.585
Giving or receiving advice	3.31	1.95	.557
Entertainment (α = .865)			
Passing the time	4.38	2.17	.886
Being entertained	4.85	1.98	.866
Having fun	4.20	1.96	.788
Relaxing	3.79	1.80	.634
Information (α = .805)			
Learning interesting things	5.16	1.64	.841
Getting information (facts, links, news, knowledge, ideas)	5.39	1.71	.813
Sharing information with others (facts, links, news, knowledge, ideas)	4.91	1.98	.812

Notes. KMO = .845. All items shared a common prompt: "How important are the following uses of Twitter to you personally?" and were measured with a 7-point Likert-type scale ranging from "very unimportant" to "very important". The factors are ordered based on variance explained.

TABLE 3-4. SUMMARY OF FACTORS AND INDIVIDUAL ITEMS DESCRIBING MOTIVES FOR TWITTER USE

3.4.2.2 Relationships between motivations for Facebook use and Twitter use

In order to understand the relationships between motivations for Facebook use and Twitter use, three multiple regressions (forced entry method) were conducted with the three Twitter motives (i.e., the factor scores for each participant) as the dependent variables and the seven Facebook motives as the predictors. Examination of collinearity diagnostics for the predictors showed VIF values well below 10 and the tolerance statistics well above 0.2, indicating no multicollinearity in the data (Field, 2009). All three models were significant. Table 3-5 shows the results of the regressions.

Motives for Facebook use	Motives for Twitter use		
	Social	Entertainment	Information
Entertainment/Content	.34***	-.06	.14
Photographs	.10	.34***	-.16
Social Network Surfing	.38***	-.12	.11
Social Connection	.12	.10	-.07
Shared Identities	.04	-.36***	.15
Status Updates	.02	.17*	.22*
Social Investigation	-.21*	.24**	.09
Intercept	-0.02	0.04	-0.02
Model significance (F value)	6.57***	7.54***	2.13*
R^2	.33	.38	.14

Notes. * $p < .05$, ** $p < .01$, *** $p < .001$, beta coefficients are standardized.

TABLE 3-5. MULTIPLE REGRESSION MODELS PREDICTING THE MOTIVES FOR USING TWITTER FROM THE MOTIVES FOR USING FACEBOOK

3.4.2.3 Differences in Facebook behavior between Twitter users and non-users

A multivariate analysis of variance (MANOVA) was performed to investigate the differences between Facebook-only users and users of both platforms across the variables collected by the Facebook API. The MANOVA test revealed a statistically significant multivariate effect, Hotelling's Trace T = 0.171, $F(12, 219) = 3.12$, $p < .001$, partial eta squared = .146, observed power = .993. Follow-up t-tests comparing the means of the 12 variables collected via the Facebook API for both groups found that Twitter users had substantially more Facebook **friends** (M = 616.3, *SE* = 43.96) than Twitter non-users (M = 393.5, *SE* = 25.03), $t(165.05) = 4.403$, $p < .001$, $r = .32$. Twitter users also attended more Facebook **events** (M = 1.728, *SE* = 0.277) than Twitter non-users (M = 1.085, *SE* = 0.144), $t(155.59) = 2.061$, $p < .05$, $r = .16$. Furthermore, Twitter users made more **check-ins** to locations (M = 4.272, *SE* = 0.780) than Twitter non-users (M = 1.690, *SE* = 0.391), $t(152.14) = 2.959$, $p < .01$, $r = .23$. The other activity variables were not found to be significantly different between the two groups. Figure 3-3 shows the differences in the means of the 12 variables for both groups. Preliminary analysis (not shown) of the data found no statistically significant differences between the two groups in terms of demographics (age, gender, nationality, student status).

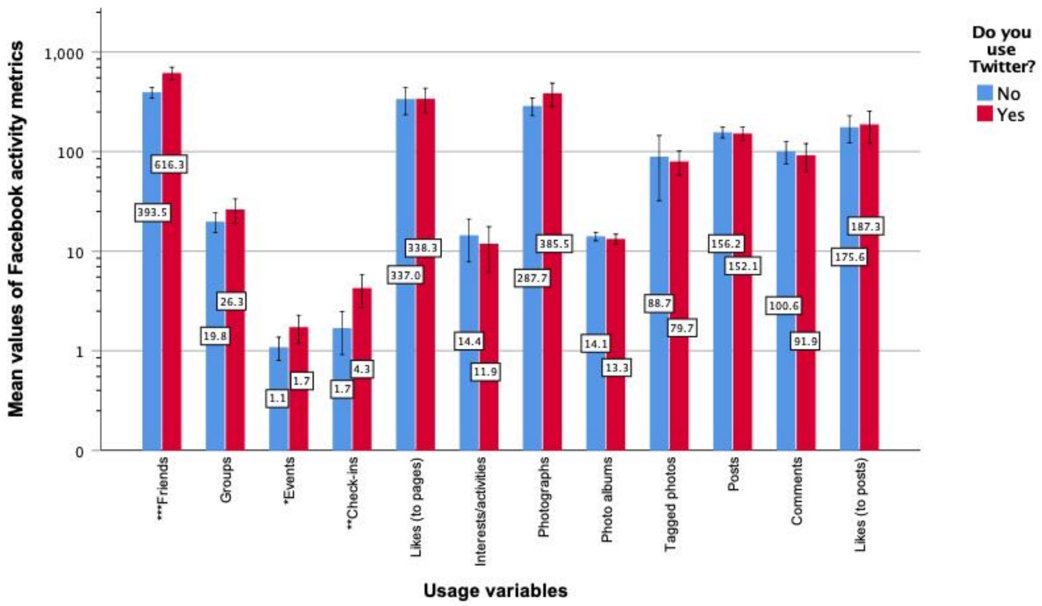

FIGURE 3-3. MEAN DIFFERENCES IN THE FACEBOOK ACTIVITY DATA (LOG SCALE) BETWEEN USERS AND NON-USERS OF TWITTER. STARRED VARIABLE NAMES INDICATE STATISTICALLY SIGNIFICANT DIFFERENCES (* P < .05, **P < .01, *** P < .001, TWO- TAILED). ERROR BARS INDICATE 95% CONFIDENCE INTERVAL.

3.4.3 DISCUSSION

3.4.3.1 Explaining the interplay between motivations for Facebook use and Twitter use

The exploratory factor analysis conducted on the answers to the set of Facebook questions yielded seven factors, corresponding to motives for Facebook use, which are generally in line with those identified by Joinson (2008). This was largely expected, since the same set of items was employed. Interestingly, the results of the analysis conducted on the answers to the Twitter questions did not fully reflect the results of the study of Johnson and Yang (2009) whose items were used in the study. In particular, while only one item was discarded, our analysis suggests the existence of an additional motive, *entertainment*, to the two already identified ones, *social* and *information* motives. This type of apparent divergence is not uncommon as U&G is an exploratory approach, rather than a confirmatory one, and can be attributed to possible differences in the sample and recruiting procedure or even changes to people's perception and use of a medium over time. For instance, Coursaris et al. (2010) also identified three motives for Twitter use; *social interaction*, *information*, and *relaxation* in the place of entertainment. Nevertheless, the examination of the items and their interpretation into factors clearly support the motives identified in the current study.

The three regression models reveal a noteworthy mix of complementary and antagonistic motivations for using the two SNSs. The Facebook *entertainment/content* motivation, which includes the only Facebook items in the study that are conceivably not social (Giannakos et al., 2013), namely using applications, playing games and doing quizzes, is a positive predictor for the Twitter *social* motivation. Interestingly, the motivation for using these apps, games and quizzes is not associated with the Twitter *entertainment* motivation, suggesting that, at least as entertainment is concerned, social media users may prefer to focus on a single medium in order to gratify this need. The *social* Twitter motive is also positively predicted by the *social network surfing* and negatively predicted by the *social investigation* Facebook motive. This finding has two important implications. First, it highlights the differences between these two Facebook motives; *social network surfing* is more focused on investigating people one is not currently a friend with on the site, while the *social investigation* motive has more emphasis on the surveillance of people one is already acquainted with and has possibly befriended on the service. It is worth noting here that the composition of the *social investigation* factor in our study is slightly different to the one from Joinson (2008), suggesting a deeper surveillance aspect of this factor and resulting in more emphasis to social browsing and less to social searching as described by Lampe and colleagues (2006). The second implication of this association of the Twitter *social* motive with the two

Facebook motives concerns the affordances of the two sites (Utz, Muscanell, & Khalid, 2015; Vitak & Kim, 2014). Facebook's richer and more structured content allows for a level of surveillance that is deeper than what Twitter allows. So, the deeper level of surveillance that characterizes the *social investigation* motive is not afforded by Twitter, but the more superficial level described in the *social network surfing* motive is.

The *entertainment* motive for using Twitter is positively predicted by the *photographs*, the *status updates*, and the *social investigation* Facebook motives, while it is negatively associated with people primarily motivated by *shared identities* on Facebook. While the factor analysis grouped several items thematically under the motive of *photographs*, further examination of the constituent items of this factor suggests that this umbrella term contains a more nuanced account of people's motives. Posting and sharing photographs on Facebook is a predominantly social activity that has been associated with a diverse set of gratifications (Malik, Dhir, et al., 2016), while a study focusing on photo-tagging on Facebook has also identified a number of gratifications with entertainment being specifically identified as one of them (Dhir, Chen, & Chen, 2017). Moreover, viewing Facebook photographs can have an entertainment element, either when interpreted as "light" surveillance when browsing photographs in one's newsfeed that their friends have posted, or when interpreted as a "deeper" form of surveillance when people engage in virtual-people watching and stalking other people as it is also described in the *social investigation* factor. The *status update* motive, primarily concerned with one's newsfeed and their friends' timelines, may also exhibit distinct entertainment value, especially when considering how the Twitter *entertainment* factor is described by its constituent items; viewing updates, links, check-ins and photographs from one's friends and pages they follow can lead to passing the time, relaxing, having fun, and generally being entertained. This finding reflects the dimension of entertainment gratified through browsing the Facebook newsfeed that has been identified in previous research (R. Lin & Utz, 2015), describes that a similar mechanism may be at play in Twitter, and, through the positive correlation found, implies that this specific dimension of entertainment acts in a complementary manner for the two sites, i.e. that users will aim to gratify this entertainment need through both Facebook and Twitter.

The *shared identities* motive is primarily involved with Facebook features that are not available on Twitter, like organizing and joining events and groups. The fact that it is associated with only one Twitter motive, and that association is negative, is another indication that users interested in a specific feature or use of a SNS will make a selection to use that SNS at the expense of a possible alternative that lacks that feature, further highlighting the importance of technological affordances for explaining SNS use.

The *information* Twitter motive is predicted by only one Facebook motive, *status updates*, and this is a positive association. Examination of the constituent items reveals a clear parallel between the two factors. Learning interesting things, getting information, and sharing information with others on Twitter are very similar activities to seeing what other people have shared on their timelines, browsing or curating the newsfeed, and updating one's status on Facebook. This positive association between two similar motives on two different services indicates a complementarity. This complementarity suggests that the motive of *information* is so strong that overcomes the negative effects that information overload can have on usage (Koroleva, Krasnova, & Günther, 2010) and is in line with research demonstrating that, at least in some contexts, information seekers utilize multiple sources in the process of acquiring information (Rains & Ruppel, 2016). Furthermore, this artifact may be an indication that individual information filtering tools, such as the Facebook newsfeed, have mitigated the effect of information overload (Y.-C. Chen, Shang, & Kao, 2009). It is also possible, however, that a personal antecedent may act as a confounding factor; for example, a recent Pew Research Center survey found that the Americans that are more technologically inclined are less likely to report a feeling of information overload (Horrigan, 2016), thus suggesting that people that feel comfortable using the full extent of the features of multiple SNSs may be less burdened in this regard. This finding may also be an indication of online social compartmentalization (Wilken, 2015; X. Zhao et al., 2016); aiming at more effective identity management (Frederic & Woodrow, 2012) or driven by concerns of context collapse (Marwick and boyd 2010), one's Facebook connections may be substantially and qualitatively different to their Twitter connections, so it makes sense to receive information from both. Another explanation of this information complementarity may reflect inherently different types of information that users are looking for on Facebook and Twitter; for instance, outside the realms of friends and family, while the same portion of users have reported getting news from both sites, the proportion of users following breaking news on Twitter is nearly twice as high as those who say they do so on Facebook (Pew Research Center, 2015). Besides receiving information, these two motives also comprise items that refer to sharing information with others. In this regard, this positive association between the two motives echoes recent findings on personal content sharing, which suggest that SNS users may combine multiple channels to create composite sharing features (Sleeper et al., 2016).

It is worth noting that in the current study we opted to elicit different sets of motivations for the two platforms, instead of assuming that users have the same motivations for using the SNSs in varying degrees of importance. Although this approach does not allow for a direct comparison between the motivations for using the two platforms (e.g., by comparing their mean values), our approach is arguably more in line with the exploratory nature of the U&G framework. The main benefit of our approach is that it encourages and facilitates the expression of the unique motivations for each platform based on their individual features and characteristics (Alhabash & Ma, 2017). For instance, the general need for sociality can be gratified with different motivations for Facebook and Twitter, or the items that make up the entertainment motivations point to subtle differences into how the two sites gratify the need for entertainment. Another advantage of eliciting different sets of motivations for each platform is that we remove any potential test-retest effect pertaining to the way participants respond to the questions, as different questions are being employed for each platform (Alhabash & Ma, 2017). Finally, expecting exactly the same motivations to be present in multiple platforms may be subject to certain validity concerns; for example Jordan (2018) points out the difficulties in constructing a sample that is simultaneously representative of all the platforms involved in multi-platform studies, something that would typically be necessary when drawing comparisons for the same motivations.

3.4.3.2 SNS non-use through the lens of media selection

Our tests comparing Twitter users and non-users revealed that having a high number of Facebook friends is associated with having a Twitter account. In fact, in our sample Twitter users had 223 more Facebook friends than non-users on average (616 versus 393 friends). This indicates that, at least with regards to the number of friends, the two SNSs are not competitive, but instead complementary, i.e. the friends one has on Facebook may be different to their followers on Twitter, something that further supports the case for online social compartmentalization (X. Zhao et al., 2016). Although previous research clearly suggests that the number of friends one has on a specific SNS is a strong predictor of how likely they are to join (Zafarani & Liu, 2014) or to continue using it (K.-M. Lin, 2016), our data show that this does not prevent them from joining other sites. An alternative reading of this finding can be that a third confounding factor affects both variables. This factor may be a primarily demographic or psychological antecedent, such as overall affinity with technology, personality or something more nuanced; for example Kim and Lee (2011) found a positive association between the number of Facebook friends and subjective well-being, a construct that has been also related to the use of image-based SNS platforms (Pittman & Reich, 2016).

On average, Twitter users attended substantially more Facebook events (1.728 versus 1.085) and used Facebook to check-in to locations more than twice more often compared to non-users (4.272 versus 1.690). Interestingly, both of these activities represent functionality that is not available on Twitter. A simple approach to media selection theory would suggest that Twitter users interested in these activities would select to also use Facebook in order to have access to them and that their decision process would explain this artifact. It is also plausible that this finding may be attributed to a personal antecedent, such as affinity with technology or self-efficacy, i.e. more technologically inclined people will feel more comfortable both in using many SNSs and taking full advantage of their functionality (Bright, Kleiser, & Grau, 2015). These features also represent an *offline* dimension of social media, as they both refer to activities that take place offline. Importantly, this finding also highlights the importance of introducing behavioral data in U&G studies. Facebook events and check-ins would be outcomes associated with the *shared identities* Facebook motive as identified by Joinson (2008). Thus, Facebook users will aim to connect to, communicate and meet with "like-minded people" by participating in (or declaring their interest to) certain events and visiting (or declaring their endorsement to) particular places. However, the *shared identities* motive was not associated positively with any motive for using Twitter, and was, in fact, a negative predictor for one of the Twitter motives. In our study, this suggests that although Twitter users report to not be particularly interested in gratifying the need to connect with like-minded people, their behavior when using Facebook clearly suggests that they are.

The underlying assumption of this analysis has been that non-use of a SNS is due to someone's explicit choice. Although Satchell and Dourish (2009) note *lagging adoption* as the most common form of non-use, we argue that the popularity of the two studied SNSs and the fact that our sample of Twitter non-users is comprised of people who are Facebook users minimizes the influence of lagging adopters, i.e. people who simply have not *yet* adopted a technology. Rather, the type of Twitter non-use in our study is more akin to what Baumer et al. (2013) describe in their research; people who do not use the site, have no intention of joining and provide well-reasoned explanations for their non-use.

3.4.3.3 Strengths and limitations

The reported research sets a starting point for exploring motivations and behaviors for using multiple SNSs, but focuses on only two sites – albeit two of the most popular ones currently. Clearly, inclusion of more social media platforms can paint a more complete picture of media selection in the social media ecosystem. It should be noted, however, that inclusion of more SNSs would bring new challenges with regard to participant recruiting and sampling. Unpacking user activity into its constituents and taking advantage of the full wealth of data that can be collected programmatically via the Facebook API was deemed more appropriate for a cross-platform study, because the granularity of the data would enable us to unearth specific nuances of use. However, although such computationally collected behavioral usage data are more objective, granular and accurate than self-reports of usage (Junco, 2013), researchers have lately started raising concerns about the quality of the API-collected data (e.g., Hogan, 2018; Lomborg & Bechmann, 2014). In the case of this study, for example, recent changes to the Facebook API mean that some variables may be replaced, merged, or even completely deprecated, and therefore it is possible that these kind of studies cannot be replicated with high accuracy. Finally, even though we attempted to respect and accommodate users' privacy concerns, it is apparent that our sample is subject to self-selection bias; not only participants self-selected to be included in the study, but they had to install a custom Facebook application and agree to offer some of their profile and activity data.

3.4.4 IMPLICATIONS

3.4.4.1 Implications for researchers

The findings from this study provide useful insights to SNS researchers interested in media selection as they expose and highlight specific details of the mechanics of SNS selection. Although previous research has employed either a motivational or a behavioral approach to describe and explain how people use multiple SNSs, the current paper extends U&G scholarship by combining survey and computational data. This way, we identified connections between motivations for using the two sites that help highlight nuances in these motivations and we have illustrated how the different affordances of the two sites inform the mechanics of the decision process of media selection. In particular, our analysis revealed both antagonistic and complementary use of the two sites based on different motives, and we found that six out of the seven Facebook motives emerged as statistically significant positive or negative predictors of Twitter motives. Furthermore, we showed how specific affordances of Facebook can affect whether one is also a Twitter user. These findings suggest that disentangling the media selection process in the current converging social media environment can benefit from moving beyond specific media-centric motivational studies and examine people's motivations and usage across SNSs.

Studying Twitter use and non-use *in the context of Facebook use* reflects back to the theory of media selection by examining the nuances of media dependency (Papacharissi, 2008; Rubin & Windahl, 1986). While the current body of research on SNS non-use focuses on single sites, this study informs non-use theory by studying non-use in conjunction with usage of another site, thus providing much-needed context and addressing a long-standing limitation of the non-use literature (Lampe et al., 2013). Furthermore, the use of a common sample and behavioral data for describing usage and non-use reinforces the assumption that any findings are due to people's explicit choices of media instead of other parameters (Knobloch-Westerwick, 2014). Researchers with a focus on adoption or continuance intention of technology should consider the study of non-use of technology in the context of the relevant ecosystem of technologies, preferably with the use of common samples and behavioral data.

Following methodologically from research that has found discrepancies between self-reported and actual Facebook use (Junco, 2013), as well as discrepancies between stated privacy attitudes and actual behavior (Taddicken, 2014), this study went beyond the single measure of self-reported usage that is the norm in U&G studies and computationally collected a range of Facebook activity variables. The wide diversity and granularity of the API-collected data allowed detailed comparisons between the two groups in our sample and resulted in unearthing meaningful and specific connections that would have probably remained hidden had self-report measures been used. Social media researchers would be encouraged to consider taking advantage of computational methods for collecting data whenever available.

3.4.4.2 Implications for practitioners

Overall, we found evidence that users interested in a specific feature or use of a SNS will make a clear selection to use that SNS at the expense of a possible alternative that lacks that feature. However, if similar functionality is available in multiple services, in some cases users will use those features in only a single SNS, while in other cases they will combine sites. More specifically, our findings show that users will use both Facebook and Twitter to gratify a need for entertainment when there is a social element to it, but will not hesitate to focus on a single medium to gratify a need for entertainment when this is not particularly social (e.g., playing games, using applications and doing quizzes). This suggests that entertainment through SNSs is not monolithic and there is a need for future studies to unpack this concept at least in its social and private constituents, if not along more dimensions. This finding is important for designers and SNS providers who should plan to target their users with more differentiated types of entertainment. The current study also highlights the differences between the social network surfing and the social investigation Facebook motives by exposing a relationship of opposite direction between each of them and the Twitter social motivation. The implication of this is that people interested in “lighter” surveillance will use both sites to achieve it, while Facebook users interested in “deeper” surveillance are not motivated to use Twitter. Reflecting back on the affordances perspective, the provision of deeper surveillance features from Facebook can be a driver for adopting and using the site. Another finding suggests that SNS users will seek to gratify their need for information from both sites. This indicates that there is room for information-focused services in the current SNS ecosystem; new services providing high-quality or domain-specific information and news may act complementarily to the currently established SNSs.

Furthermore, information providers should also keep in mind that people combine SNSs to gratify their information needs, something particularly significant as previous research has shown that exposure to multiple sources can be more important than multiple exposures from the same source (González-Bailón, Borge-Holthoefer, Rivero, & Moreno, 2011).

Finally, adding to previous research that suggests that a network factor is a significant contributor to the "stickiness" of a SNS (C.-P. Lin & Bhattacherjee, 2008; K.-M. Lin, 2016), we found that this relationship can be more complex. Our findings show that Facebook users that are more embedded in the site (i.e., have more friends) are significantly more likely to also have a Twitter account. Although "critical mass" has long been recognized as a key factor in media acceptance and selection (Markus, 1987), our study suggests that, in the current social media ecology, the network externalities that characterize the critical mass mechanic and the "stickiness" to a site do not necessarily prevent users from joining another site. We also found that users of features that are more particular to Facebook and are related to an offline dimension, such as check-ins and Facebook events, were more likely to own a Twitter account. This complementarity corroborates the argument that people do not hesitate to use multiple SNSs to fulfill different goals, thus putting the service loyalty perspective (Shankar, Smith, & Rangaswamy, 2003) into question and suggesting that it is now meaningful to consider media use in a feature-specific, instead of a medium-specific, manner. It may be the case that the low barrier of entry to SNSs and the low cost of switching should prompt a rethinking of these concepts of media adoption and continuance intention.

3.4.5 CONCLUSION

This study set out to answer the second sub-question of the second overarching research question of the dissertation, i.e. RQ2b or whether a combination of survey and Facebook-API collected data can provide additional value and insights when understanding social media selection above what can be gained from each approach in isolation. Overall, we found support for this, as the comparison of Facebook and Twitter motives combined with the comparison of the two sites in terms of a range of API data provided results that were not possible to obtain with survey data only. This also answered the first specific question that was posed in the context of this study.

The results also find significant relationships between motivations for using Facebook and motivations for using Twitter, thus answering question 2. We also identified specific differences between Facebook-only users and users of both platforms (question 3).

CHAPTER 4. STUDYING DYADIC RELATIONSHIPS: TIE STRENGTH, TRUST, AND DISCLOSURE

4.1 INTRODUCTION

This chapter describes the work conducted in the context of what I described earlier as Study 2 and consists of two parts, Study 2A and Study 2B. This thread of research employs survey data and computationally collected data to study dyadic relationships on Facebook. The data collection was based on a Facebook application that presented participants with a random set of their friends and asked them to rate their friendships in terms of certain aspects, while collecting a number of metrics about each pair of connections using the Facebook API. Study 2A challenges the popular assumption that people on social media act as self-interested rational actors that communicate based on expected reciprocity from their connections, by hypothesizing and testing for a moderating effect of tie strength and interpersonal trust on this decision-making process. Study 2B examines the effectiveness of predicting the reported tie strength between two Facebook users from data collected computationally from the Facebook API.

Although they seek to address different research questions, both studies shared the same data collection process, and thus, for reasons of simplicity and presentation, they are combined in one chapter. Section 4.2 describes the overall set-up, detailing the data collection procedure, the study participants, the survey content and the behavioral data collected. Section 4.3 introduces Study 2A, provides the motivation for the study and the research questions, presents the results and discusses the findings and the implications of the work. Accordingly, Section 4.4 introduces Study 2B, provides the motivation for the study and the research questions, presents the results and discusses the findings and the implications of the work.

The work presented in this chapter builds on a data collection effort by Diogo Pereira during his MSc thesis that was co-supervised by me and Ian Oakley (see Pereira, 2014). Study 2A is reported in a recently accepted journal article (Spiliotopoulos & Oakley, 2019). Study 2B was presented at (Spiliotopoulos, Pereira, & Oakley, 2014).

4.2 Method and Study Set-up

4.2.1 Procedure

Participants were recruited with a request to complete an online survey, primarily through posts in social media, but also a relevant online forum and an online study repository. The online survey was in the form of a Facebook application. The first page of the application included a comprehensive description of the study, clearly framed the experiment as an academic study, explained the data collection process, provided the contact details of the researchers, and requested users' consent. In addition to our description, Facebook displays all data-access permissions granted to an application during installation, thus ensuring that the participants had a comprehensive account of the data captured by the study. Apart from the basic profile information, the application requested a single extended permission, "Access posts in your News Feed". Users can deny extended permissions when they install an application, but the study participants were instructed to accept this permission and, in fact, the application was designed so that it would not proceed unless they did so. The participants had the choice to opt out of the study at any time.

After completing demographic questions, each participant was presented with the name and profile picture of a randomly selected Facebook friend and asked to answer three questions[1] about them, essentially rating their relationship by moving a horizontal slider (see Figure 4-1 for an example), similar to the approach followed by previous studies (Gilbert & Karahalios, 2009; Luarn & Chiu, 2015). The slider for each question had to be moved in order for the application to proceed to the next person as a means to ensure that the participant had rated the friendship before moving on to the next. The position of the slider was internally translated into a value between 0 and 1 with a granularity of 0.01. In the meantime, the application gathered a range of data about the interactions between the two people via the Facebook API. After rating 20 friends, the participant was presented with summary results and some light-hearted commentary

[1] The survey, in fact, consisted of eight questions for each friendship, however a preliminary factor analysis did not yield factors consistent with an original hypothesis (possibly due to the ad-hoc nature of some of the additional questions), therefore for the remainder of the dissertation the focus rests on these three questions that are more theoretically grounded.

about their rated friends. Participants were then given the option to rate more friends but were also able to quit the application. The application and survey were pilot-tested with two groups of ten participants each for technical or data collection issues and comprehension/ambiguity of the questions, respectively. These twenty participants answered questions about twenty of their friends each, i.e. 400 friendships, and are not included in the main survey.

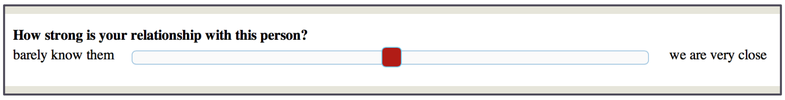

FIGURE 4-1. EXAMPLE OF THE HORIZONTAL SLIDER USED FOR RATING FRIENDSHIPS

4.2.2 PARTICIPANTS

The survey was implemented and deployed in both the English and Portuguese languages and was targeted to speakers of either language. Participants with fewer than 20 and more than 1000 Facebook friends were excluded, resulting in a sample of 90 participants (59% male) who rated 1728 Facebook friendships in total. The participants had a mean age of 26.9 years (SD = 8.7), and came from 11 countries with the vast majority (n=77, 85.6%) from Portugal and 4.4% (n=4) from the USA. They had a mean of 355 Facebook friends (SD = 218.9, range = 28 – 872) and reported using Facebook for an average of 13.4 (SD = 15.1) hours per week.

4.2.3 MEASURES

4.2.3.1 Survey data

The study design required each participant to answer questions about a large number of their friends, so it was important to keep each set of questions short in order to prevent fatigue on the part of the participants. Thus, the three constructs of interest (tie strength, interpersonal trust, and expected reciprocity) were operationalized using single-item measures. Although single-item measures are not ideal, researchers have provided evidence that under certain conditions single items can function similarly to multiple items in terms of reliability and predictive validity (Alexandrov, 2010; Bergkvist & Rossiter, 2007; Wanous & Hudy, 2001). Tie strength was measured with the question "How strong is your relationship with this person?" with the rating on the slider spanning from "barely know them" to "we are very close" and no intermediate markings. Although Gilbert and Karahalios (2009) considered five questions and created five respective models of tie strength, they deemed this question the most general and representative one and decided to focus on this one question for further analysis. Similarly, Panovich and colleagues (2012) employed this single question to validate their tie strength model. Interpersonal trust was measured with the question "How much do you trust this person?" with the rating spanning from "I do not trust this person" to "I would trust this person with my life". This specific single-item measure has also been employed to rate interpersonal trust or trustworthiness in many studies across disciplines, from neuroeconomics (Phan, Sripada, Angstadt, & McCabe, 2010) and organizational science (Evans, Hendron, & Oldroyd, 2015) to studies of social networks (Schensul & Burkholder, 2005). In order to measure the expected reciprocity, an ad-hoc item was formulated, "How much are you looking forward to receiving updates from this person?", with the rating spanning from "not at all" to "very much". It is worth noting that in the current study the concept of expected reciprocity is operationalized contextually, i.e. it refers to actions and attributes within Facebook. Research has shown that single-item measures are appropriate when a construct refers to a concrete, singular object or attribute (Bergkvist & Rossiter, 2007), as in this case.

4.2.3.2 Behavioral data

While participants were answering the survey questions for each friend, the application gathered a range of data about the content already shared between the two people. The application gathered 18 pieces of data for each rated pair (Table 4-1). For technical reasons, data collection was limited to participants' latest 200 photographs, 50 groups and 50 events. Family relationships and the difference in education level were modeled as numerical values of 0, 1, 2, and 3, as in prior work (Gilbert & Karahalios, 2009). The *days since last communication* measure refers to the most recent interaction between two users on Facebook from the day of data collection. The *days since first communication* would ideally be calculated based on the date that two users became friends. Since this information is only partially available on Facebook, it is based on the earliest interaction on record. The number of *intimacy words exchanged in wall (timeline) posts* was based on a relevant sentiment analysis dictionary (Nielsen, 2011), translated from English to Portuguese and used in both languages.

Data Collected	Mean	SD
Wall (timeline) posts exchanged	0.23	0.718
Comments exchanged on wall (timeline) posts	0.09	0.529
Comments on participant's photos [1]	0.05	0.675
Comments on photos where participant is tagged [1]	0.26	1.431
Likes on participant's wall (timeline) posts	0.36	3.718
Likes on photos where participant is tagged	0.23	1.363
Likes on participant's photos[1]	0.05	0.439
Number of mutual friends	34.79	43.38
Number of groups in common [2]	0.67	1.261
Mutual confirmed participation in events [3]	0.03	0.212
Family	0.03	0.217
Number of appearances together on photos	0.33	1.6
Number of wall (timeline) words exchanged	3.13	11.61
Days since first communication	844.1	831.4
Days since last communication	813.6	848.4
Difference in education level	0.59	0.657
Intimacy words exchanged in wall (timeline) posts	0.06	0.362
Participant-initiated wall (timeline) posts	0.22	0.68

Limits: [1]last 200 photographs, [2]50 groups, [3]50 events

TABLE 4-1. THE 18 VARIABLES GATHERED BY THE APPLICATION FOR EACH FACEBOOK FRIENDSHIP

4.3 Study 2A: Altruistic and Selfish Communication on Social Media: The Moderating Effects of Tie Strength and Interpersonal Trust

4.3.1 Introduction

Borrowing theoretical frameworks from economics, researchers have often attempted to describe and explain disclosure via the application of a subjective cost-benefit analysis, postulating that users act as rational, self-interested actors that constantly engage in a decision-making process where they evaluate the perceived personal benefits of a specific disclosure against the probability and severity of potential privacy risks (Dienlin & Metzger, 2016). This decision-making process is often expressed as a function of expected reciprocity, that is we engage in communication with a specific person because we want them to communicate back with us in return (Cook et al., 2013). In other words, this strand of research characterizes online interactions as predominantly selfish, i.e. motivated by the expectation of reciprocity, and questions the possibility of an otherwise altruistic motivation, i.e. without the expectation of reciprocity (Kollock, 1999). Further work, however, has argued that a rigid, direct application of this cost-benefit analysis underplays the importance of many factors that influence our online behavior, and that privacy and disclosure online are, in fact, contextually determined (Nissenbaum, 2009; Pelaprat & Brown, 2012; Quinn & Papacharissi, 2018). Other studies have identified specific dyadic characteristics that influence the disclosure of information, such as tie strength (Y.-C. Wang, Burke, et al., 2016) and interpersonal trust (Joinson et al., 2010). For example, when contemplating a potential online interaction with a very close friend or with an acquaintance recently met at a conference, one would consider communicating and eliciting reciprocal communication for different reasons and to different ends in each case. The current study aims to further investigate this assumption of a SNS user as a self-interested rational actor by modeling communication between pairs of Facebook friends as a function of expected reciprocity and examining this relationship for two types of content, text-related and photo-related. Driven by suggestions of previous research, I also aim to investigate possible moderation effects of tie strength and trust on this relationship, i.e. that for differing levels of tie strength and/or interpersonal trust the relationship between expected reciprocity and actual sharing will also differ. Thus, this study hypothesizes:

H1: Tie strength moderates the relationship between expected reciprocity and intensity of text-related communication.

H2: Tie strength moderates the relationship between expected reciprocity and intensity of photo-related communication.

H3: Interpersonal trust moderates the relationship between expected reciprocity and intensity of text-related communication.

H4: Interpersonal trust moderates the relationship between expected reciprocity and intensity of photo-related communication.

4.3.2 DATA ANALYSIS AND RESULTS

Although participants were encouraged to rate at least 20 friendships, they were allowed to rate as many as they wanted. Participants that rated fewer than five friendships (13 participants, 33 cases in total) were removed from the dataset. Further analysis of the responses showed 334 cases where a participant rated a friendship with zero on the tie strength question. This number is in line with a recent study in which participants could only accurately name 72.7% of their Facebook friends (Croom, Gross, Rosen, & Rosen, 2016). Because the current study focuses on the disclosures with Facebook connections that the participants actually know, these cases (19.7% of total) were also removed, resulting in a usable dataset of 1361 cases for further analysis, where 77 participants performed a mean of 17.7 (SD= 9.9) usable ratings each.

4.3.2.1 Survey data

Table 4-2 shows descriptive statistics of the answers to the three survey questions and Figure 4-2 the distributions of the answers.

Measure	Survey question	Mean	Median	SD
Tie strength	Q1. How strong is your relationship with this person?	.362	.30	.278
Expected reciprocity	Q2. How much are you looking forward to receiving updates from this person?	.355	.32	.282
Trust	Q3. How much do you trust this person?	.384	.37	.284

TABLE 4-2. MEASURES AND ANSWERS TO THE SURVEY QUESTIONS

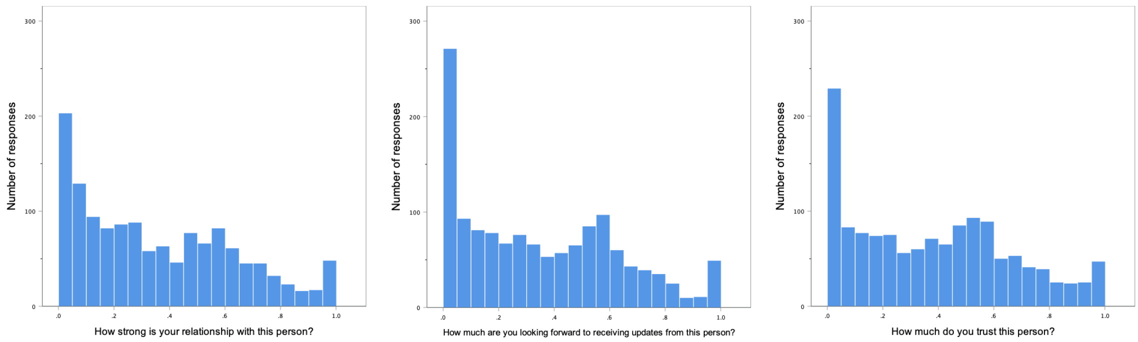

FIGURE 4-2. DISTRIBUTIONS OF THE ANSWERS TO THE SURVEY QUESTIONS

4.3.2.2 Behavioral data

In this study we make use of 11 out of the 18 variables collected, in particular those that refer to actual communication between pairs of friends. Thus, in order to measure text communication between the participant and each of their friends rated we collected six metrics (e.g., the number of timeline posts exchanged), while to measure communication related to photographs we collected five metrics (e.g., the number of likes on a participant's photographs).

In order to get an accurate composite measure of the text and photograph communication characterizing the friendships, a principal component analysis with orthogonal rotation (varimax) was conducted on these variables and the factor scores were used for further analysis. The correlation matrix revealed one case of extreme multicollinearity, namely the relationship between the number of *wall (timeline) posts exchanged* and the number of *participant-initiated posts* ($r = .968$, $p < .001$) leading to the elimination of the latter variable from further analysis. All of the behavioral variables follow power law distributions, and thus we used the logarithm (base *ln*, after adding a start-value of 1) of these variables to control for skew and then standardized by centering at the mean and dividing by the standard deviation. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the principal component analysis, KMO = .712. This value confirms the sample size as "good" (Field, 2009; Kaiser, 1974) for this analysis. Bartlett's test of sphericity χ^2 (45) = 5299, $p < .001$, indicated that correlations between the items were sufficiently large and suitable for this analysis (Field, 2009). Harman's single-factor test revealed that the variance explained by a single factor was less than 50% (34.12%), suggesting the data were free from common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The two factors corresponding to the text and photograph communication explained in combination 53.5% of the variance. Both factors exhibited good reliability with Cronbach's alpha values above .70.

Table 4-3 shows descriptive statistics of the behavioral data collected by the application and the factor loadings after rotation.

Items	Mean	SD	Factor Loadings
Text-related communication (α = .773)			
Number of wall (timeline) words exchanged	3.71	12.74	.89
Wall (timeline) posts exchanged	0.28	0.79	.79
Comments exchanged on wall (timeline) posts	0.10	0.58	.69
Intimacy words exchanged on wall (timeline) posts	0.07	0.41	.65
Likes on participant's wall (timeline) posts	0.43	4.18	.52
Photo-related communication (α = .762)			
Comments on photos where participant is tagged	0.32	1.59	.77
Likes on photos where participant is tagged	0.26	1.44	.75
Number of appearances together on photos	0.38	1.69	.72
Likes on participant's photos	0.04	0.37	.66
Comments on participant's photos	0.06	0.74	.61

TABLE 4-3. DESCRIPTIVE STATISTICS OF ITEMS AND SUMMARY OF FACTORS DESCRIBING THE TEXT AND PHOTOGRAPH-RELATED COMMUNICATION ON FACEBOOK BASED ON THE API-COLLECTED DATA

4.3.2.3 Testing moderation effects

In order to investigate moderation effects of the tie strength and trust variables on the relationships between expected reciprocity and the two types of communication (text and photo-related), two moderated multiple regression analyses were conducted. Moderated multiple regression includes the interaction of predictors as a term in the regression equation in order to examine whether or not the interaction of the predictors accounts for incremental variance in the dependent variable beyond the variance accounted for by main effects (Baron & Kenny, 1986;

Hayes, 2018). Before running the regressions, predictor variables were centered and the two interaction variables (expected reciprocity * tie strength, expected reciprocity * trust) were created. Thus, two hierarchical multiple regression models were tested predicting the actual text and photo-related interactions from the measure of expected reciprocity, with the interaction variables added in the second step of each model. Examination of collinearity diagnostics for the predictors showed VIF values well below 10 and the tolerance statistics above 0.2, indicating no multicollinearity in the data (Field, 2009). The Durbin-Watson statistic values were 1.772 and 1.978 confirming that the assumption of independence of errors for the two regressions has been met (Durbin & Watson, 1971; Field, 2009). Overall, both models including only the main effects were significant and the addition of the interaction terms in the second step of each regression resulted also in significant models and accounted for significantly more variance in both cases. Examination of the beta coefficients and their significance showed that two of the four hypotheses were supported.

The model predicting text-related communication from expected communication reciprocity was significant $F(3, 1357) = 43.4$, $R^2 = .087$, $p < .001$. Addition of the interaction terms in the second step also resulted in a significant model, $F(5, 1355) = 27.8$, $R^2 = .093$, $p < .001$, and accounted for significantly more variance, R^2 change = .006, $p = .016$, indicating potentially significant moderation of tie strength and trust on the relationship between expected reciprocity and actual text-related communication (Table 4-4). However, the two interaction effects were not statistically significant and, thus, H1 and H3 are not supported.

	Step 1		Step 2		Hypothesis tested
	β	t	β	t	
Expected reciprocity (REC)	.067	1.471	.053	1.158	
Tie strength (TS)	.093	1.820	.078	1.481	
Trust (TR)	.156**	3.142	.157**	3.163	
REC × TS			-.011	-0.202	H1 (not supported)
REC × TR			.089	1.668	H3 (not supported)
R^2	.087		.093		
Adjusted R^2	.085		.090		
F change	43.367***		4.144*		

Notes. * $p < .05$, **$p < .01$, *** $p < .001$. Beta coefficients shown are standardized. N=1361.

TABLE 4-4. MODERATED MULTIPLE REGRESSION ANALYSIS TESTING THE MODERATION EFFECTS OF TIE STRENGTH AND INTERPERSONAL TRUST ON THE RELATIONSHIP BETWEEN EXPECTED RECIPROCITY AND TEXT-RELATED COMMUNICATION.

The model predicting photo-related communication from expected communication reciprocity was significant $F(3, 1357) = 17.4$, $R^2 = .037$, $p < .001$. Addition of the interaction terms in the second step also resulted in a significant model, $F(5, 1355) = 14.5$, $R^2 = .051$, $p < .001$, and accounted for significantly more variance, R^2 change = .014, $p < .001$, indicating potentially significant moderation of tie strength and trust on the relationship between selfish motivation for communication and actual photo-related communication (Table 4-5). Both interaction effects were statistically significant, indicating support for both H2 and H4.

	Step 1		Step 2		Hypothesis tested
	β	t	β	t	
Expected reciprocity (REC)	-.091	-1.941	-.096*	-2.049	
Tie strength (TS)	.232***	4.405	.178***	3.307	
Trust (TR)	.030	0.597	.042	0.830	
REC × TS			.213***	3.806	H2 (supported)
REC × TR			-.113*	-2.077	H4 (supported)
R^2	.037		.051		
Adjusted R^2	.035		.047		
F change	17.370***		9.777***		

Notes. * $p < .05$, **$p < .01$, *** $p < .001$. Beta coefficients shown are standardized. N=1361.

TABLE 4-5. MODERATED MULTIPLE REGRESSION ANALYSIS TESTING THE MODERATION EFFECTS OF TIE STRENGTH AND INTERPERSONAL TRUST ON THE RELATIONSHIP BETWEEN EXPECTED RECIPROCITY AND PHOTO-RELATED COMMUNICATION.

In order to aid visualization and interpretation of the moderation effects we generated a set of estimates of the dependent variable (i.e., the factor scores for photo-related communication) from combinations of the moderators (i.e., tie strength and trust) and the main effect variable (i.e., expected reciprocity) using the unstandardized coefficients of the variables in the regression models (including the intercept) and plotted the dependent variable as a function of the moderators and the main effects. Per the recommendation of Hayes (2018 p.244), we used the 16th, 50th, and 84th percentile values (equivalent to a standard deviation below the mean, the mean, and a standard deviation above the mean if a variable is assumed to be normally distributed) to denote low, mid, and high values in the variables. Figure 4-3 and Figure 4-4 show visual representations of the two significant moderation effects.

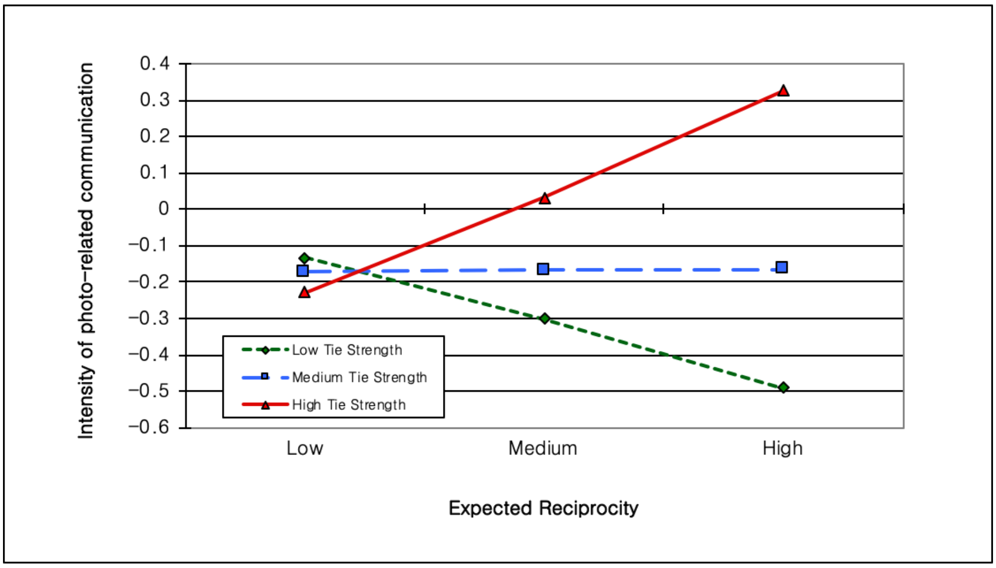

FIGURE 4-3. THE INTERACTION BETWEEN LEVELS OF TIE STRENGTH AND EXPECTED RECIPROCITY ON THE INTENSITY OF PHOTO-RELATED COMMUNICATION (HYPOTHESIS H2)

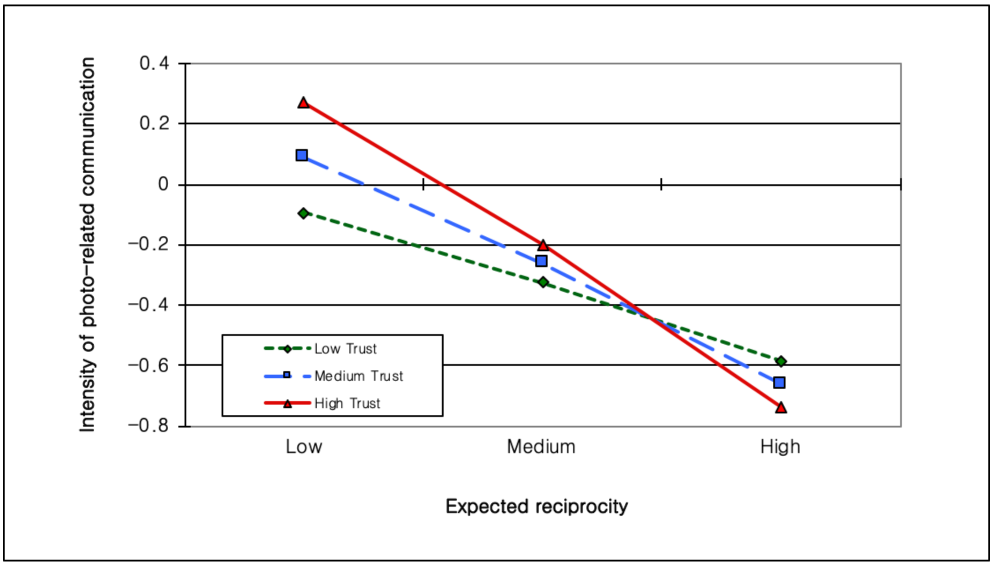

FIGURE 4-4. THE INTERACTION BETWEEN LEVELS OF TRUST AND EXPECTED RECIPROCITY ON THE INTENSITY OF PHOTO-RELATED COMMUNICATION (HYPOTHESIS H4)

4.3.3 DISCUSSION

This study examined the moderating effects of tie strength and interpersonal trust on the relationship between expected reciprocity as a motivation for interacting on Facebook and the actual interactions that take place between specific pairs of Facebook connections. Furthermore, we differentiated between two types of interactions based on content, namely text-related and photo-related communication. Results show significant moderating effects of tie strength and trust on communication in two out of the four examined cases of moderation, namely the two cases predicting photo-related communication.

Our results show that tie strength moderates the relationship between expected reciprocity and actual sharing of content around photographs. In particular, for low levels of expected reciprocity the intensity of communication is similar across all levels of tie strength. As the motivation for communication becomes more selfish (i.e. for higher levels of expected reciprocity), the intensity of communication rises for strong ties (such as close friends and family), remains steady for contacts in the medium tie-strength category, and actually decreases for weak ties (such as remote acquaintances) (Fig. 2). In other words, for the strong ties that we are particularly interested in receiving communication from, more actual communication indeed takes place. However, for the weak ties that we are particularly interested in receiving communication from (e.g. an important person that we are not close to, or specific content creator), the intensity of actual communication that takes place decreases. This means that, with regards to photo-related content, the model of a self-interested rational actor may not provide an adequate understanding and interpretation of behaviour, but instead should take into account measures of tie strength.

Interpersonal trust is revealed as a significant moderator that enhances the effect of the predictor on the outcome. Specifically, the more the participant was looking forward to receiving updates from their friend, the less actual photo-related communication was measured between them, and this effect was amplified by the trust the participant showed for their friend (Fig. 3). This moderating effect, however, is weaker than in the tie strength case. The difference in direction between the effects of the two moderators may be due to the structural differences of the two moderators. On one hand, tie strength is considered largely mutual and undirected (Granovetter, 1973), for example two close friends or relatives are expected to mutually report their relationship as strong and two distant acquaintances will report their relationship as weak. On the other hand, interpersonal trust can often be one-sided and directed (Golbeck & Hendler, 2006), for example one may show great trust towards a specific Facebook connection, be that a personal friend, a boss or a public figure, while the other person may not feel the same way and, thus, not be eager to reciprocate the communication. As we collected interaction information that also included two-way communication, this asymmetry in interpersonal trust presents a possible explanation for the moderating effect. Furthermore, the low-reciprocity, low-trust Facebook connections may represent cases where reciprocity is simply not generally expected. Posts about important positive life events, such as having a baby, getting married, or earning a degree are generally shared to larger audiences (Day, 2013) and are more likely to include photos (Bevan et al., 2015). Thus, the large audience for these cases may skew the dataset towards a high intensity of communication around photographs for the low-reciprocity, low-trust cases.

Our findings as a whole, reflect criticisms of RCT that have suggested that the relationship between expected reciprocity and the intensity of communication is not as straightforward as the theory suggests. For example, Pelaprat and Brown (2012) make the theoretical argument that "[online] social actions that solicit a return–action seek to neither profit nor benefit, but rather express a desire to draw in others into social life and relationships". The results in this paper provide empirical support for this argument; we find more actual photograph-related communication taking place between low tie-strength connections when expected reciprocity is low compared to expected reciprocity being high. A similar effect takes place for high trust connections. Weak ties and trusted individuals are persons that we would like to draw further into our social life and relationships, since they can provide novel information and connections (in the case of weak ties), and reduce the risk of disclosure (in the case of trusted individuals).

The experimental set-up and data collection approach of the current study have both benefits and limitations. On one hand, this work answers the call of many scholars recommending the study of people's behaviour in realistic situations instead of lab experiments with self-reported behavioural data (e.g. Knijnenburg, Kobsa, and Jin 2013) by employing a Facebook application to collect objective, accurate and granular data about participants' online interactions. This approach is especially important for the study of online disclosure, as previous research has found significant discrepancies between self-reported and actual Facebook use (Junco, 2013), as well as individuals' intentions to disclose personal information and their actual behaviours online (Norberg, Horne, & Horne, 2007). Furthermore, this experimental format is particularly suitable for empirically studying online disclosure at the dyadic level, something that is a long-standing limitation of disclosure studies that typically focus on the individual as the salient unit of analysis (Smith et al., 2011). On the other hand, it is worth noting that researchers have lately started raising concerns about the quality of API-collected data (e.g. Lomborg & Bechmann, 2014; Weiler, 2018). In the case of this study, for example, changes to the Facebook API since the data were captured mean that some variables have been replaced or deprecated, and, in fact, API access to friends' data has been limited, making it possible that these kind of studies cannot be easily replicated with high accuracy in the future (Hogan, 2018). Finally, even though we attempted to respect and accommodate users' privacy concerns, it is clear that our sample is subject to self-selection bias; not only participants self-selected to be included in the study, but they had to install a custom Facebook application and agree to offer some of their data.

4.3.4 IMPLICATIONS

4.3.4.1 Theoretical implications

This work provides insights for communication research by investigating the application of RCT for understanding users' behaviour on SNSs. In particular, this paper puts into question the assumption of a SNS user as a self-interested rational actor and shows that the relationship between expected reciprocity and SNS communication is, in fact, moderated by tie strength and interpersonal trust in specific ways. While previous criticisms of RCT for describing disclosure have emphasized individual differences (Hann, Hui, Lee, & Png, 2007), environmental cues (John, Acquisti, & Loewenstein, 2010) and platform interface cues (Gambino, Kim, Sundar, Ge, & Rosson, 2016), our approach contributes to this body of research by focusing on characteristics of dyadic relationships. Our findings are important for social media researchers studying and modelling SNS behaviour. Future studies of dyadic online interactions should keep in mind the ways that tie strength and interpersonal trust influence the links between motivations for communication and actual behaviour, and include such measures in their models or control for the differences between strong ties and weak ties in their sampling and analyses. The findings in this paper are also important to researchers in economics aiming to understand the limits, applications and possible extensions of RCT (see Sato, 2013 for discussion on these broader topics). Especially, the fledging research area that focuses on the application of behavioural economic theories and practices (such as soft paternalism and nudging) for understanding and motivating/incentivizing SNS behaviour can be of particular benefit (Wang et al., 2014).

While our study found significant effects of our variables of interest (tie strength and interpersonal trust) on the relationship between expected reciprocity and actual photo-related communication, this was not the case for text-related communication. This calls attention to the ways that photographic content on SNSs can be inherently different to text content and highlights the need for more studies in this area. This distinction between the two types of content is further supported by our finding that Facebook usage can be effectively dimensionalized into photograph-related and text-related. This supports the argument that explanations of online interaction should refrain from treating interaction on a specific platform in a monolithic way, but instead could benefit from focusing on specific modes of interaction, such as text and photographs.

4.3.4.2 Practical implications

Previous theoretical work has argued for the importance of reciprocal interactions for understanding and supporting online activity (Kizilcec, Bakshy, Eckles, & Burke, 2018; Pelaprat & Brown, 2012). Our findings show that expected reciprocity does not directly translate to actual communication, but is instead moderated by tie strength and interpersonal trust. This means that simply designing for expected reciprocity is not enough to support online communication, but instead the interactions of tie strength and trust with expected reciprocity should be taken into account. Previous research has identified ways to enhance reciprocity by increasing expected reciprocity on social media, such as designing for "encounter", providing public visibility of specific actions motivated by reciprocity, and facilitating symbolic exchanges (Pelaprat & Brown, 2012). The design recommendations arising from the current paper suggest that such design decisions aimed at supporting reciprocity would be more effective when targeted at specific SNS connections based on the characteristics of the relationship with the connection, namely tie strength and interpersonal trust. These recommendations can be used as inputs to drive models of behaviour and algorithms that suggest connections to share specific content with or manage visibility of interactions (e.g. in newsfeed-like features). This can enhance the design of SNS platforms and third-party tools that connect to the platforms, as well as SNS users' privacy and convenience.

Other SNS research has gone beyond the directed type of reciprocity addressed in this paper, in which the receiver may feel obligated to give something back to the giver, and has found that actions aimed at reciprocity may also lead to general reciprocity, which generates a desire to be more broadly generous (e.g. to "pay it forward" to someone else) (Kizilcec et al., 2018). The visibility of reciprocal actions has also been found to have a strong effect on reciprocity (i.e. observing generous or reciprocal actions between people leads to more generous actions) via a hypothesized mechanism of social learning that affects social norms of reciprocity (Kizilcec et al., 2018; Pelaprat & Brown, 2012). The coupling of these two effects means that enhancing the link between expected reciprocity and actual behaviour by optimizing the visibility of social interactions based on measures of tie strength and trust can lead to social networks with more engaged users.

4.3.5 CONCLUSION

This study modeled communication between pairs of Facebook friends as a function of expected reciprocity and examined this relationship for two types of content, text related and photo related. In doing so, it leveraged a range of programmatically collected data and survey data that were collected with the help of an application. Overall, we presented findings that enhance our understanding of dyadic communication on Facebook, thus addressing the pertinent sub-question of the second overarching question of the dissertation. In terms of addressing the four specific hypotheses of the study, we found support for two of them, namely that tie strength and interpersonal trust moderate the relationship between expected reciprocity and intensity of communication in the case of photo related communication.

4.4 STUDY 2B: PREDICTING TIE STRENGTH WITH THE FACEBOOK API

4.4.1 INTRODUCTION

As we saw in the related work, tie strength is a very important concept that characterizes dyadic relationships. Thus, automatic, and even real-time, calculation of tie strength on social media can have numerous and significant implications. This study makes use of the wealth of digital traces that describe dyadic relationships on Facebook to attempt to predict reported values of tie strength between two connections. This work aims both at demonstrating the prediction of tie strength in a way that can be useful input for further work, and at identifying specific aspects of social media relationships (potentially expressed through specific Facebook features) that can add to our understanding of the concept of tie strength in social media.

Thus, this work is guided by the following research questions:

1) How accurately can we predict reported tie strength from Facebook API data?

2) Which specific Facebook features, affordances, and participant actions are associated with tie strength?

4.4.2 DATA ANALYSIS, RESULTS AND DISCUSSION

A multiple regression was run 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. As expected, the two "days since" variables were highly correlated ($r = .99$, $p < .01$), as cases of 0 or 1 interaction result in the same "days since" number. Although high correlation among predictor variables is typically a problem in multiple regression, the fact that the two variables carry opposite effects in the regression support the theory from previous research (Gilbert & Karahalios, 2009) that these variables do not contain the same information about the dependent variable and therefore should both remain as predictors.

The two "days since" variables aside, the correlation matrix showed a number of strong relationships among the predictor variables, one of which exceeded the 0.8 benchmark that indicates potential multicollinearity: "wall (timeline) posts exchanged" and "participant-initiated wall (timeline) posts". Thus, the regression was run again excluding the second variable (with a total of 17 predictor variables). Examination of the Variance Inflation Factor (VIF) for each predictor variable found the highest value of VIF to be 2.667, well below the benchmark value of 10 that indicates multicollinearity. Therefore, we are confident that the regression carried out was free from multicollinearity concerns. The model was significant, $F(17, 1710) = 16.7$, $p < .001$. Table 4-6 shows the result of the regression. The standardized beta coefficients show the effect (i.e., the relative weight) of each dependent variable on tie strength.

Overall, in the regression model for tie strength 10 of the 17 employed variables emerged as significant predictors. *Days since first* and *days since last communication* are signs of the duration and intensity of the relationship, respectively, and were found to have a very strong effect on tie strength, as indicated by the literature (Gilbert & Karahalios, 2009). As expected, *family* was a significant positive predictor of tie strength, exhibiting the highest beta coefficient after the two *days since* variables.

The number of *wall (timeline) words exchanged* showed a significant positive effect on tie strength; writing text to each other in public may signify greater intimacy than the plain, lightweight communication achieved with likes. The *number of co-appearances in photographs* also emerged as a positive predictor, indicating that strong ties will typically also engage in offline relationships. Similarly, the *number of groups* that a participant had in common with a friend was a strong predictor of tie strength, hinting at the value of homophily; strong ties have similar interests and belong to the same groups. The number of a *friend's comments on photos that the participant was tagged in* was positively associated with tie strength, whereas no other types of comments were found to have a significant effect. This is likely a sign of intimacy, as photographs where the participant is present are conceivably more personal than other types of photos. The number of *likes on a participant's wall (timeline) posts* was also positively associated with tie strength.

Measure	**β**
Days since last communication	-.641**
Days since first communication	.550*
Family	.159***
Number of wall (timeline) words exchanged	.110**
Number of appearances together on photos	.093**
Comments on photos where participant is tagged	.072*
Number of groups in common	.068**
Number of mutual friends	-.051*
Likes on participant's wall (timeline) posts	.050*
Difference in education level	-.048*
Comments exchanged on wall (timeline) posts	.046
Likes on participant's photos	.027
Likes on photos where participant is tagged	-.023
Intimacy words exchanged in wall (timeline) posts	-.016
Mutual confirmed participation in events	-.010
Wall (timeline) posts exchanged	.010
Comments on participant's photos	.003
Intercept	0.292***
R^2	.142
Adjusted R^2	.134

Notes. * $p < .05$, ** $p < .01$, *** $p < .001$, beta coefficients are standardized.

TABLE 4-6. REGRESSION MODEL OF TIE STRENGTH

Interestingly, the number of *mutual friends* was negatively associated with tie strength, something that at first seems counter-intuitive. This can be explained by the fact that a Facebook user may belong to several social contexts, some of which can involve large clusters of acquaintances (such as school, university, workplace). People in these large clusters have a large number of mutual Facebook friends because of the very fact that they belong to the same cluster. However, some traditionally very strong ties, such as family members, childhood friends, or generally old and close friends with friendships that span many years of time may not belong in such clusters.

Finally, social distance, derived from the *difference in education* variable in our dataset, was also negatively associated with tie strength.

The unstandardised beta coefficients were used for the creation of a new linear model for calculating tie strength. Figure 4-5 shows the distribution of the tie strength of the relationships reported by the participants, as well as the distribution of the tie strengths calculated by our model. As in similar literature (Gilbert & Karahalios, 2009; Panovich et al., 2012), the range is normalized between 0 and 1 for each participant, where 0 is the weakest tie strength of a friend, and 1 is the strongest. In line with the findings of previous work (Gilbert & Karahalios, 2009) the model showed a bias towards underestimation of tie strength. More specifically, the mean of the reported tie strength was measured at 0.29 (median = 0.21) and the mean of the model's tie strength was 0.13 (median = 0.1). It is notable, however, that 19.7% of friendships rated by the participants were set to zero. This means that participants acknowledged that a substantial percentage of their Facebook relationships are essentially non-existent.

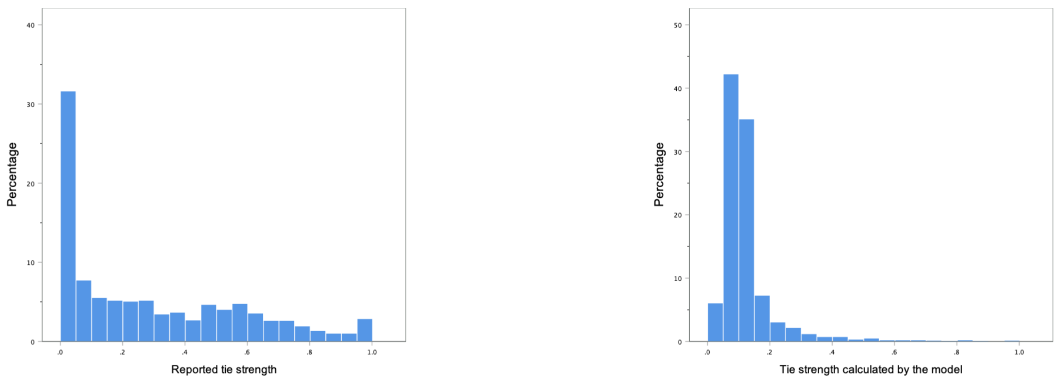

FIGURE 4-5. DISTRIBUTION OF REPORTED TIE STRENGTH OF PARTICIPANTS AND TIE STRENGTH CALCULATED BY THE MODEL

To demonstrate the practical applicability of this model, we reduced the tie-strength based relationships to two fundamental categories, weak ties and strong ties. In line with the approach by Gilbert and Karahalios (2009) we classified all friends above the model's mean value as strong ties and all those below as weak. Correct predictions are those where the participant's rating is correspondingly above or below the mean in the questionnaire dataset. Our tie strength model classified with 65.9% accuracy using this procedure, χ^2 (1, N = 3456) = 135.08, $p < .001$. However, given the large number of friendships per person that a Facebook application can have access to and the positive skew of the distribution of the reported and calculated tie strengths, it is also meaningful to examine the ability of the model to predict the few strongest ties. These would correspond to the close friends of the participant, those that the participant communicates more often and possibly more meaningfully with. Thus, we classified the strongest 10% of ties as very strong and ran the chi-square test again, showing that our model had an accuracy of 86.3% in differentiating between very strong and weaker ties, χ^2 (1, N = 3456) = 107.83, $p < .001$.

4.4.3 IMPLICATIONS

While prior work has demonstrated that tie strength can be derived from social media, this paper demonstrates the feasibility of doing this live and using a standard 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 tie strength of connections. Different types of information, such as questions or status updates, could be broadcast to different contacts for more efficient answers (Panovich et al., 2012) or information diffusion (Bakshy et al., 2012; J. Zhao et al., 2011). The default values of privacy controls, or recommended privacy settings for new connections can be set based on tie strength. Better recommendations for new connections could be made. For instance, if strong ties A-B and A-C exist, and if B and C are aware of one another, then it has been shown that a "psychological strain" may exist between B and C (Granovetter, 1973) and recommendations that these users become friends might best be avoided. Finally, even though social network analysis has proven useful in providing an understanding of social media (Brooks et al., 2011; Spiliotopoulos & Oakley, 2013b), it has also been suggested that appropriate use of network analysis depends on choosing the right network representation for the problem at hand (Butts, 2009). This paper suggests that studying alternative network representations, such as those whose links are weighted based on tie strength instead of binary friendships, has the potential to be the basis for substantial advances in understanding user behavior in social media.

There are also limitations to this work. The sample used in the study is relatively homogeneous, i.e. most participants are Portuguese. More diverse samples should be used in future studies in order to be more representative of the Facebook user population. There are also more conceptual issues. For example, although studies employing computationally collected usage metrics can provide many practical insights, arguably there is much about social media usage that falls outside the scope of such metrics. For example, previous work suggests that one reason why users choose not to post content is because they are in an offline social context, such as a meeting, where it would be inappropriate to do so (Sleeper et al., 2013). Since such behavior would not be reflected in the usage data, it is possible that a more complete picture could be obtained with the combination of computational and qualitative data. To deal with this issue, future work on this topic should explore such mixed methods approaches.

4.4.4 CONCLUSION

This study sought to predict and model tie strength from Facebook API-collected data, thus addressing the final sub-question of the second overarching research question of the dissertation. Building on the work of Gilbert and Karahalios (2009) who employed a browser extension to collect relevant data, this study provided a way of predicting and modeling tie strength from API data. We found that this approach can be useful in predicting strong and very strong ties, thus answering the first specific question asked in the introduction of this study. Furthermore, the study identified ten specific metrics that are significant positive or negative predictors or the reported tie strength (question 2).

CHAPTER 5. CONCLUSIONS

5.1 SUMMARY

An increasing volume of research has started acknowledging that a potential combination of traditional social science approaches with computational approaches can be valuable and practical for the study of interactive computer systems. In this dissertation, I have sought to add further credence and validity to this argument, as well as explore pertinent emerging issues, by conducting a series of studies that employ a combination of survey and API-collected data for the study of SNSs across domains, interaction levels, and sociotechnical processes.

This first thread of research in this dissertation (presented in Chapter 3) focuses on the study of individual users, investigating motivations for using Facebook and Twitter and providing insights into how people navigate the SNS landscape and select the most appropriate platform to spend their time and attention on. The second thread of research (presented in Chapter 4) focuses on dyadic relationships, and investigates antecedents of patterns of communication on Facebook and specific concepts like tie strength, interpersonal trust, and reciprocity. All studies employed a combination of survey and Facebook API-collected data. Taken together, the results of these studies provide support towards the central claim that a suitable combination of traditional social science approaches with computational approaches can provide value and insights for SNS researchers and designers that are over and above what can be gained from each approach in isolation.

Below, I discuss key insights derived from the research approach followed and provide an overview of the main contributions.

5.2 CONTRIBUTIONS

5.2.1 UNDERSTANDING ONLINE BEHAVIOR ON SNSS

5.2.1.1 Identifying motivations for using SNSs and contributions to the U&G theoretical and analytical framework

Study 1A elicited motivations for using Facebook, with the exploratory factor analysis identifying seven such motivations, namely *entertainment/content*, *photographs*, *social network surfing*, *social connection*, *shared identities*, *status updates*, and *social investigation*, in order of variance explained. These were in line with previous research (Joinson, 2008), something that was expected since the same set of items were used[1]. More interestingly, however, this study attempted to predict these motives from 18 API-collected variables describing usage and personal network characteristics of each participant. This has several important implications. First, it identifies specific usage and network variables that are associated with specific motivations, providing, thus, more insights into how specific users gratify certain needs through Facebook. For example, we found that the number of check-ins posted was negatively associated with the *shared identities* motive, suggesting that while this motive seeks to gratify a need for connecting with specific people, advertising or showing off one's presence in a particular location does not express or gratify this need. Second, this work highlights an important theoretical contribution to the U&G framework; we demonstrated that a range of usage metrics can take the place of the (typically self-reported and unidimensional) U&G outcomes, and that personal network metrics can act as novel forms of U&G antecedents.

In Study 1B we extended the above work to also examine motives for using Twitter. In this case, we identified three motives, namely *social*, *entertainment*, and *information*, in order of variance explained. Interestingly, the entertainment motive was not identified in the original study that was replicated (P. Johnson & Yang, 2009), however this motive has been identified in other Twitter U&G studies. This study also contrasted Facebook and Twitter motivations for the same users, revealing significant associations between the three motives for Twitter use and the seven motives for Facebook use. By studying the Facebook and Twitter motivations in tandem, we also

[1] While this was the expected outcome of the replication of the previous work, our workshop paper (Spiliotopoulos & Oakley, 2013a) and our main publication of this study (Spiliotopoulos & Oakley, 2013b) provide some additional details in the form of suggestions for items to be included in future Facebook U&G studies, and the relation of these motives to privacy concerns and behavior, respectively. Although these findings are outside the scope of this dissertation, the interested reader can refer to these publications for further particulars in this regard.

addressed the specific theoretical concerns of U&G researchers that single-platform motivational studies hinder conceptual development because they are too compartmentalized and produce separate typologies of motives (Ruggiero, 2000).

5.2.1.2 Contribution to research on SNS non-use

Our results from Study 1B contribute to the growing field of non-use research by contrasting Twitter users to non-users. In particular, we found that Twitter users have substantially more Facebook friends, attend more Facebook events and make more check-ins to locations than Twitter non-users. These findings also contribute to the non-use scholarship by providing an example of a non-use study of a site in conjunction with the usage of another site, thus providing much-needed context and addressing a long-standing limitation of the non-use literature (Lampe et al., 2013), as typically studies of non-use tend to examine a site or a technology in isolation.

5.2.1.3 Disentangling social media selection

As SNSs become richer, more diverse, popular and diffused in the population, they continuously compete for people's attention and time. In this environment, media present functional alternatives to each other and previous research has identified several factors that come into play when a user decides which SNS or appropriate combination of SNSs is more effective for meeting their needs. In Study 1B we focused on Facebook and Twitter in order to get insights into this decision-making process. Our results suggest that SNS users will utilize both sites to gratify their need for information, but will only do so for entertainment that has social characteristics. We also found that Facebook users that are more embedded in the site and use the site to support their offline life are more likely to also use Twitter. These results expose specific instances of the effects of site affordances, online social compartmentalization, information overload, and context collapse on this process.

5.2.1.4 Understanding communication on social media by highlighting the mediating effects of tie strength and interpersonal trust

People share a diversity of content on social media for a variety of reasons. In Study 2A we took as a point of departure previous research that has described and explained disclosure via the application of a subjective cost-benefit analysis framed around reciprocity, i.e. suggesting that people communicate selfishly motivated by the expectation of receiving something in return. With that in mind, we investigated the moderating effects of tie strength and interpersonal trust between two Facebook connections on the relationship between expected reciprocity and intensity of communication between these two specific people. The results showed significant moderating effects of tie strength and trust on the communication around photographs, but not around text. In this regard, this work made a significant theoretical contribution by identifying and investigating factors outside the remit of a traditional goal-oriented, self-interested cost-benefit analysis that influence online behavior (Pelaprat & Brown, 2012). An important implication of these findings is that future studies of online interaction would do well to consider the concepts of tie strength and interpersonal trust as important factors that influence online interaction at the dyadic level. For example, studies on dyadic online interaction should include measures of tie strength in their models or control for the differences between strong ties and weak ties in their sampling and analyses.

5.2.1.5 Highlighting Facebook characteristics that affect tie strength

Following on the research thread of studying dyadic relationships, in Study 2B we identified specific Facebook features, affordances and participant actions that are associated with the tie strength reported by the participants. In particular, we found ten variables that are significant positive or negative predictors of the reported strength of participants' ties. This work builds on and extends seminal previous work that ventured a similar link from the online to the offline (Gilbert & Karahalios, 2009).

5.2.2 Implications for SNS designers and practitioners

5.2.2.1 Predicting motives for SNS use

Study 1A described a way of predicting motives for using Facebook from a range of computationally collected data. This is particularly important, as motives can be very subjective, personal and have been acknowledged as concepts that characterize online behavior in very impactful ways. Prediction of motives from computational data can be useful to several stakeholders. Marketers may want to promote their products or services to the users who visit Facebook with a particular goal in mind. SNS designers can be facilitated in offering adaptive systems that can personalize the system interface and the user experience. Furthermore, earlier in the design process, motives can be used for informing user personas that are useful creative artifacts for the requirements analysis and design of interactive systems. SNS operators may also benefit from identifying users that are driven by specific motivations, because this can help distinguish between users with similar usage patters, such as non-active users and lurkers, and design their policies accordingly.

5.2.2.2 Leveraging findings on social media selection

Overall, we found evidence that users interested in a specific feature or use of a SNS will make a clear selection to use that SNS at the expense of a possible alternative that lacks that feature.

However, if similar functionality is available in multiple services, in some cases users will use those features in only a single SNS, while in other cases they will combine sites. In the text, I have provided numerous examples of this interplay in the case of Facebook and Twitter that can assist SNS designers and operators to find ways that a SNS can differentiate itself from other SNSs, or complement other SNSs when appropriate. For example, information providers can benefit from keeping in mind that SNS users will seek to gratify their need for information from multiple sites, and, thus, introducing new services that provide high-quality or domain-specific information and news may complement the currently established SNSs instead of acting antagonistically.

5.2.2.3 The nuances of designing for reciprocity

With expected reciprocity identified as a central motivator for engaging in dyadic communication, researchers have argued for designing for reciprocity, for example by providing public visibility of actions directed at specific others (Pelaprat & Brown, 2012). These design efforts can benefit from the findings of Study 2A that show that this relationship between expected reciprocity and dyadic communication is not that straightforward, but instead is affected by tie strength and interpersonal trust for different types of content in the specific ways explicated in our findings. For example, interfaces may be designed to present strong ties differently to weaker ones, or to take into account the trustworthiness of a person.

5.2.2.4 Predicting tie strength from API-collected data

While prior work has demonstrated that tie strength can be derived from social media (see Gilbert & Karahalios, 2009), Study 2B demonstrates the feasibility of doing this live and using a standard API. This opens the door to many novel adaptable and customized features and services based on tie strength, relating to person and information recommendations and filtering, social information searching (e.g., asking questions), information diffusion strategies, and privacy controls. Furthermore, this work can lead to the study of alternative network representations, such as those whose links are weighted based on tie strength instead of binary friendships, that have been largely advocated as having the potential to be the basis for substantial advances in understanding user behavior in social media (Butts, 2009).

5.2.3 The added value of combining traditional social science and computational methods

5.2.3.1 Augmenting the U&G framework

While U&G studies typically employ a survey instrument to collect all necessary data, in Study 1A we included 21 API-collected metrics to be used as outcomes and antecedents. The addition of the usage metrics (as outcomes) significantly and substantially increased the explanatory value of the framework; in fact, this addition more than doubled the variance explained by the models, on average. We also find that some isolated network metrics (intended as novel antecedents) emerged as significant predictors. Overall, this showed that the addition of computational data augmented the framework in a significant way and suggests that future work should consider similar or other computational approaches to augment survey data.

5.2.3.2 Enhancing understanding of media selection

Besides comparing Facebook and Twitter motives, Study 1B differentiated between Twitter users and non-users based on a set of 12 Facebook API-collected variables. This allowed a comparison of the two sites on more dimensions and with more accurate and objective data than what would be possible with survey data only. This is particularly interesting, as the demographic variables on their own could not account for statistically significant differences between the two groups. Furthermore, employing these computational data to understand Facebook usage is especially important when studying decision-making processes (in this case media selection), because they account for people's actual behavior instead of intended behavior and, thus, allow the differentiation between preferences and actual choices. Overall, the differences between the two groups that emerged from the computational data, coupled with the insights from comparing motives for using the two sites, enhanced our understanding of the SNS selection process.

5.2.3.3 Differentiation of data collection methods

In Study 2A we used 11 API collected variables to measure text and photo-related communication between pairs of Facebook friends. Again, the collected data are particularly useful, because they refer to actual behaviors instead of behavioral intentions. Furthermore, asking the participants about their willingness to communicate with specific friends (or to reflect on how often they have actually communicated with them) on the same survey instrument that asked questions about tie strength and trust would expose extra biases, in addition to the ones already inherent in survey studies. Thus, disentangling the dependent variable of interest (i.e., intensity of communication) from the independent ones (i.e., reported tie strength, trust, and expected reciprocity) by following distinct data collection methods led to more representative and valid results.

5.2.3.4 A more granular depiction of Facebook relationships

Study 2B was heavily influenced by the work of Gilbert and Karahalios (2009), who employed a browser extension to collect a range of data for each friendship, instead of the API (which didn't exist at the time). Thus, the contribution of this study in this regard is in validating the findings of the previous research, i.e. providing a more detailed understanding of the construct of tie strength via a more granular depiction of Facebook relationships.

5.2.4 REFLECTIONS ON THE STRENGTHS AND LIMITATIONS OF THE RESEARCH APPROACH

In this section I summarize and outline some specific reflections and practical lessons learned from the overall research approach of augmenting survey data with Facebook API-collected data that was followed in the studies presented in the dissertation. These reflections represent both strengths and limitations of the approach, as in several cases these are connected.

5.2.4.1 Sampling and recruitment

- As with other web-based survey studies, the work presented here was subject to a self-selection bias. This means that the people who opted to participate in the studies may not adequately represent typical users. Furthermore, the data collection procedures followed required participants to install a Facebook application and share some of their data, exacerbating this bias. On the other hand, though, these same processes may have discouraged spurious participants, such as careless, dishonest, or mischievous users.
- Asking people to install a Facebook application to participate in the studies definitely discouraged many potential and willing participants, which led to a smaller sample size in both studies.
- Although our sampling approach that used a Facebook ad campaign led to a diverse sample and increased the outreach of the study, it also led to an interesting sampling artifact. In particular, the auction-like mechanism that Facebook uses to distribute its ads meant that ads were shown where the cost of advertising was lower. This resulted to the ads being shown to (and clicked by) a demographic that based on the Facebook advertising algorithm has low purchasing power (as our ads were competing with commercial ones). Hence, we had a disproportionately large number of participants from India in our sample for Study 1.

- Research has shown that online-recruited participants can provide high quality data, while bringing a high degree of diversity (Casler, Bickel, & Hackett, 2013). Arguably this diversity is especially important in studies that follow an exploratory approach (as opposed to a confirmatory one), such as our application of the U&G framework in Study 1.

5.2.4.2 Study design

- Employing the API for the collection of network data in Study 1 ensured that there were no problems with regard to missing network data. Missing data have been shown to be a significant problem for (social) network-oriented studies (Kossinets, 2006).
- Asking each participant to rate 20 friendships as default meant that we had to restrict the number of questions in order to keep the survey short. This led to using single-item measures for the constructs in question, which is a quite uncommon practice. In the case of our studies, however, the single-item measures used were adequately supported by theory.
- In Study 2 participants rated their friends, which means that our dataset was subject to the Friendship Paradox effect (Feld, 1991). This paradox states that "on average, your friends have more friends than you do", an effect that is a common occurrence in social network studies both offline (Feld, 1991) and specifically in the case of the Facebook network (Ugander, Karrer, Backstrom, & Marlow, 2011). Although conceptually this wasn't important for our study, future studies that follow a similar set-up and investigate concepts that may be theoretically associated with the popularity of SNS users should keep this effect under consideration.
- The application in Study 2 showed a picture and some basic details of the friend that was being rated. Thus, there was less bias from difficulty to recall, something that is a common concern in social network studies (Brewer, 2000).
- Changes to the Facebook API since the data were captured mean that some variables have been replaced or deprecated. Especially in the case of Study 2, API access to friends' data has been severely limited, making it possible that these kinds of studies cannot be easily replicated with high accuracy in the future.
- There is a clear and comprehensive push lately for sharing research data in the interest of transparency and enabling the replication of studies. It should be noted, however, that sharing API-collected data without the users' explicit consent is against Facebook policy. Furthermore, the combination of extensive, personal and potentially private data, the possible opportunities that exist for triangulating data, and the negative publicity of such

practices done commercially, such as in the example of Cambridge Analytica (Isaak & Hanna, 2018), have resulted in a general reluctance of publicly sharing these kind of datasets from the part of researchers.

- The potential privacy and ethical implications of computational SNS studies mean that certain aspects of study design must be carefully considered. In particular, special care must be taken in the design and communication of participant consent forms, the prevention of unintended sharing of data, and the measures to ensure the security of data and prevent breaches.

5.2.4.3 Data analysis

- Using dyadic relationships as the unit of analysis in Study 2 had some interesting implications. On one hand, we could make use of a very large sample (i.e., of relationships), as each participant was asked to rate a number of friendships. On the other hand, from a statistical perspective, this may be considered a possible violation of the independence assumption (i.e., not all ratings were fully independent from each other as some were made by the same person) that is required for certain statistical tests. However, this does not formally fall under the definition of the assumption of independence and similar studies in the literature in many occasions have decided to ignore it. In our case, we employed appropriate statistical tests to check for possible effects of this, in particular independence of errors and common-method bias.
- Some API-collected data exhibited power law distributions (i.e., the variables violated the assumption of normality) and had to be log-transformed before they were used in certain statistical tests. This is something that should be kept in mind for similar studies.
- Besides the extra breadth, granularity, accuracy, and objectivity provided by the computational approach to collecting API data, employing an application-based survey in Study 2 allowed more granularity even for the non-API collected data. In particular, the theoretical constructs of interest (tie strength, interpersonal trust and expected reciprocity) were measured with a slider, which provided more granular control compared to the typically used likert scale.

5.3 Future Directions

Through the two two-part studies presented, I have sought to provide four pieces of evidence that support the central claim of the dissertation, that a suitable combination of traditional social science and computational approaches can provide value and insights for SNS researchers and designers over and above what can be gained from each approach in isolation. As the four pieces of evidence act as proofs in the context of a proof-by-demonstration approach, it is clear that more demonstrators would further advance this claim. A large number of reported personal attributes and attitudes can replace Study 1's uses and gratifications as objects of study, for example personality, socioeconomic status, depression, anxiety, loneliness, empathy, civic participation. Similarly, any number of dyadic relationship characteristics can replace the ones used in Study 2, such as influence, workplace hierarchy, respect, or perceived attractiveness. Future studies can also put their attention on the study of interaction at different levels. Many examples of such studies come to mind that present specific characteristics of interest when studied in the context of social media; teams and small groups of people that may or may not play different roles in the group (e.g., co-workers), organizations (which follow a certain hierarchy and may be dispersed geographically), communities (e.g., around influential people, products, or ideas), societies (e.g., governments, political campaigns, international issues). Methodologically, there is room to study SNSs with different SNA approaches. While the studies in this dissertation employed egocentric network metrics, future work may also examine complete networks and even bipartite networks (e.g., networks that include both people and posts as nodes). Approaches like text or sentiment analysis can also provide alternative or complementary ways to get insights from such studies.

The studies I presented in this dissertation focused on Facebook for the computational data collection, and looked only at Facebook and Twitter to study media selection. Clearly, any number of other SNSs can become the object of studies combining traditional social science and computational approaches. Furthermore, inclusion of more SNSs can paint a more complete picture of media selection in the SNS ecosystem, although that would bring new challenges with regard to participant recruiting and sampling. Future work may also decide to drill down to the study of specific features of SNSs, such Facebook groups and fan pages.

Study 2A identified tie strength and interpersonal trust as factors that affect the goal-oriented rational-actor approach towards disclosure on SNSs. Further research may seek more factors that affect this approach and, thus, provide a more holistic view with regard to the rationality of SNS disclosure.

In our studies, the Facebook API presented an excellent platform both for data collection and for developing a survey instrument in the form of an appealing interactive application. However, recent changes to the Facebook API and the APIs of other SNSs (see Hogan, 2018 for further discussion), lack of control of APIs, and several other concerns that researchers have expressed about API data (Lomborg & Bechmann, 2014; Snodgrass & Soon, 2019) may lead future studies to utilize alternative computational methods of data collection, such as other forms of data mining, web crawling, or log data where they may be available. It is worth noting, however, that such alternative methods of data collection bring new challenges in terms of ethics, privacy and legal compliance; for instance, Freelon (2018) refers to a post-API age of computational research, where adherence to a platform's Terms of Service (ToS) constitutes a major challenge for researchers.

BIBLIOGRAPHY

Abramova, O., Wagner, A., Krasnova, H., & Buxmann, P. (2017). Understanding Self-Disclosure on Social Networking Sites - A Literature Review. In *23rd Americas Conference on Information Systems - AMCIS 2017*. Boston, MA, USA.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Alexandrov, A. (2010). Characteristics of single-item measures in likert scale format. *Electronic Journal of Business Research Methods*, *8*(1), 1–12.

Alhabash, S., & Ma, M. (2017). A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? *Social Media + Society*, *3*(1). https://doi.org/10.1177/2056305117691544

Alhazmi, H., & Gokhale, S. S. (2016). Mining Social Capital on Online Social Networks with Strong and Weak Ties. In *2nd International Conference on Open and Big Data (OBD)* (pp. 9–16). IEEE. https://doi.org/10.1109/OBD.2016.9

Altman, I. (1973). Reciprocity of Interpersonal Exchange. *Journal for the Theory of Social Behaviour*, *3*(2), 249–261. https://doi.org/10.1111/j.1468-5914.1973.tb00325.x

Archambault, A., & Grudin, J. (2012). A longitudinal study of facebook, linkedin, & twitter use. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (p. 2741). New York, New York, USA: ACM Press. https://doi.org/10.1145/2207676.2208671

Arnaboldi, V., Conti, M., Passarella, A., & Pezzoni, F. (2012). Analysis of Ego Network Structure in Online Social Networks. In *SocialCom*.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web - WWW '12* (p. 519). New York, New York, USA: ACM Press. https://doi.org/10.1145/2187836.2187907

Barak, A., & Gluck-Ofri, O. (2007). Degree and Reciprocity of Self-Disclosure in Online Forums. *CyberPsychology & Behavior*, *10*(3), 407–417. https://doi.org/10.1089/cpb.2006.9938

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Basak, E., & Calisir, F. (2015). An empirical study on factors affecting continuance intention of using Facebook. *Computers in Human Behavior*, *48*, 181–189. https://doi.org/10.1016/j.chb.2015.01.055

Baumer, E. P. S., Adams, P., Khovanskaya, V. D., Liao, T. C., Smith, M. E., Schwanda Sosik, V.,

& Williams, K. (2013). Limiting, leaving, and (re)lapsing: : An Exploration of Facebook Non-Use Practices and Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (p. 3257). New York, New York, USA: ACM Press. https://doi.org/10.1145/2470654.2466446

Baumer, E. P. S., Ames, M. G., Burrell, J., Brubaker, J. R., & Dourish, P. (2015). Why study technology non-use? *First Monday*, *20*(11).

Bergkvist, L., & Rossiter, J. R. (2007). The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, *44*(2), 175–184. https://doi.org/10.1509/jmkr.44.2.175

Bevan, J. L., Cummings, M. B., Kubiniec, A., Mogannam, M., Price, M., & Todd, R. (2015). How Are Important Life Events Disclosed on Facebook? Relationships with Likelihood of Sharing and Privacy. *Cyberpsychology, Behavior, and Social Networking*, *18*(1), 8–12. https://doi.org/10.1089/cyber.2014.0373

Binder, J., Howes, A., & Sutcliffe, A. (2009). The problem of conflicting social spheres: effects of network structure on experienced tension in social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 965–974). https://doi.org/10.1145/1518701.1518849

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

boyd, d., & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230.

boyd, d., & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, *15*(8).

Brewer, D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, *22*, 29–43.

Bright, L. F., Kleiser, S. B., & Grau, S. L. (2015). Too much Facebook? An exploratory examination of social media fatigue. *Computers in Human Behavior*, *44*, 148–155. https://doi.org/10.1016/j.chb.2014.11.048

Brooker, P., Barnett, J., & Cribbin, T. (2016). Doing social media analytics. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716658060

Brooks, B., Welser, H. T., Hogan, B., & Titsworth, S. (2011). Socioeconomic Status Updates: Family SES and emergent social capital in college student Facebook networks. *Information, Communication & Society*, *14*(4), 529–549. https://doi.org/10.1080/1369118X.2011.562221

Buccafurri, F., Lax, G., Nicolazzo, S., & Nocera, A. (2015). Comparing Twitter and Facebook user behavior: Privacy and other aspects. *Computers in Human Behavior*, *52*, 87–95. https://doi.org/10.1016/j.chb.2015.05.045

Burke, M., & Kraut, R. (2013). Using Facebook after Losing a Job : Differential Benefits of Strong and Weak Ties. In *Proceedings of CSCW 2013*.

Burke, M., Kraut, R., & Marlow, C. (2011). Social capital on Facebook: Differentiating uses and users. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (pp. 571–580). ACM.

Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (p. 1909). New York, New York, USA: ACM Press. https://doi.org/10.1145/1753326.1753613

Butts, C. T. (2009). Revisiting the foundations of network analysis. *Science (New York, N.Y.)*, *325*(5939), 414–416. https://doi.org/10.1126/science.1171022

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in twitter: The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM).*

Chen, G. M. (2011). Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior*, *27*(2), 755–762. https://doi.org/10.1016/j.chb.2010.10.023

Chen, X., Pan, Y., & Guo, B. (2016). The influence of personality traits and social networks on the self-disclosure behavior of social network site users. *Internet Research*, *26*(3), 566–586. https://doi.org/10.1108/IntR-05-2014-0145

Chen, Y.-C., Shang, R.-A., & Kao, C.-Y. (2009). The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. *Electronic Commerce Research and Applications*, *8*(1), 48–58. https://doi.org/10.1016/j.elerap.2008.09.001

Cheshire, C., & Cook, K. S. (2004). The Emergence of Trust Networks under Uncertainty – Implications for Internet Interactions. *Analyse & Kritik*, *26*(1), 220–240. https://doi.org/10.1515/auk-2004-0112

Chiu, C., Ip, C., & Silverman, A. (2012). Understanding Social Media in China. *McKinsey Quarterly*, (March), 78–81.

Conrad, F. G., Keusch, F., & Schober, M. F. (2020). Call for Papers: New Data in Social and Behavioral Research. *Public Opinion Quarterly*.

Contena, B., Loscalzo, Y., & Taddei, S. (2015). Surfing on Social Network Sites: A comprehensive instrument to evaluate online self-disclosure and related attitudes. *Computers in Human Behavior*, *49*, 30–37. https://doi.org/10.1016/j.chb.2015.02.042

Cook, K. S., Cheshire, C., Rice, E. R. W., & Nakagawa, S. (2013). Social Exchange Theory. In *Handbook of Social Psychology* (pp. 61–88). https://doi.org/10.1007/978-94-007-6772-0_3

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks*, *25*(4), 283–307. https://doi.org/10.1016/S0378-8733(03)00012-1

Coursaris, C. K., Osch, W. Van, Sung, J., & Yun, Y. (2013). Disentangling Twitter's Adoption and Use (Dis)Continuance: A Theoretical and Empirical Amalgamation of Uses and Gratifications and Diffusion of Innovations. *AIS Transactions on Human-Computer Interaction*, *5*(1), 57–83.

Coursaris, C. K., Yun, Y., & Sung, J. (2010). Twitter Users vs. Quitters: A Uses and Gratifications and Diffusion of Innovations Approach in Understanding the Role of Mobility in Microblogging. In *2010 Ninth International Conference on Mobile Business and 2010 Ninth Global Mobility Roundtable (ICMB-GMR)* (pp. 481–486). IEEE. https://doi.org/10.1109/ICMB-GMR.2010.44

Croom, C., Gross, B., Rosen, L. D., & Rosen, B. (2016). What's Her Face(book)? How many of their Facebook "friends" can college students actually identify? *Computers in Human Behavior*, *56*, 135–141. https://doi.org/10.1016/j.chb.2015.11.015

Day, S. (2013). Self-disclosure on Facebook: how much do we really reveal? *Journal of Applied Computing and Information Technology*, *17*(1), 1–6.

Derlega, V. J., Wilson, M., & Chaikin, A. L. (1976). Friendship and disclosure reciprocity. *Journal of Personality and Social Psychology*, *34*(4), 578–582. https://doi.org/10.1037/0022-3514.34.4.578

DeVito, M. A., Birnholtz, J., & Hancock, J. T. (2017). Platforms, People, and Perception: Using Affordances to Understand Self-Presentation on Social Media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17* (pp. 740–754). New York, New York, USA: ACM Press. https://doi.org/10.1145/2998181.2998192

Dhir, A., Chen, G. M., & Chen, S. (2017). Why do we tag photographs on Facebook? Proposing a new gratifications scale. *New Media & Society*, *19*(4), 502–521. https://doi.org/10.1177/1461444815611062

Dienlin, T., & Metzger, M. J. (2016). An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample. *Journal of Computer-Mediated Communication*, *21*(5), 368–383. https://doi.org/10.1111/jcc4.12163

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Durbin, J., & Watson, G. S. (1971). Testing for serial correlation in least squares regression. III. *Biometrika*, *58*(1), 1–19. https://doi.org/10.2307/2332325

Dwyer, C., Hiltz, S., & Passerini, K. (2007). Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace. In *AMCIS*.

Eftekhar, A., Fullwood, C., & Morris, N. (2014). Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Computers in Human Behavior*, *37*, 162–170. https://doi.org/10.1016/j.chb.2014.04.048

Ellison, N., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social

Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

Ellison, N., Steinfield, C., & Lampe, C. (2011). Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media & Society*, *13*(6), 873–892. https://doi.org/10.1177/1461444810385389

Evans, J. M., Hendron, M. G., & Oldroyd, J. B. (2015). Withholding the Ace: The Individual- and Unit-Level Performance Effects of Self-Reported and Perceived Knowledge Hoarding. *Organization Science*, *26*(2), 494–510. https://doi.org/10.1287/orsc.2014.0945

Facebook. (2019). Facebook Newsroom. Retrieved June 11, 2019, from https://newsroom.fb.com/company-info/

Fairfield, J., & Shtein, H. (2014). Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism. *Journal of Mass Media Ethics*, *29*(1), 38–51. https://doi.org/10.1080/08900523.2014.863126

Farnham, S. D., & Churchill, E. F. (2011). Faceted identity, faceted lives: Social and Technical Issues with Being Yourself Online. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11* (p. 359). New York, New York, USA: ACM Press. https://doi.org/10.1145/1958824.1958880

Fast Company. (2018). Why Google defined a new discipline to help humans make decisions. Retrieved July 27, 2018, from https://www.fastcompany.com/90203073/why-google-defined-a-new-discipline-to-help-humans-make-decisions

Feld, S. L. (1991). Why Your Friends Have More Friends Than You Do. *Am.J.Sociol.*, *96*(6), 1464–1477.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). Sage Publications.

Frederic, S., & Woodrow, H. (2012). Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*. New York, New York, USA: ACM Press. https://doi.org/10.1145/2145204.2145320

Freelon, D. (2014). On the Interpretation of Digital Trace Data in Communication and Social Computing Research. *Journal of Broadcasting & Electronic Media*, *58*(1), 59–75. https://doi.org/10.1080/08838151.2013.875018

Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, *35*(4), 665–668. https://doi.org/10.1080/10584609.2018.1477506

Friedkin, N. (1980). A test of structural features of granovetter's strength of weak ties theory. *Social Networks*, *2*(4), 411–422. https://doi.org/10.1016/0378-8733(80)90006-4

Gambino, A., Kim, J., Sundar, S. S., Ge, J., & Rosson, M. B. (2016). User Disbelief in Privacy Paradox: Heuristics that Determine Disclosure. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16* (pp. 2837–2843). ACM Press. https://doi.org/10.1145/2851581.2892413

García-Martín, J., & García-Sánchez, J.-N. (2015). Use of Facebook, Tuenti, Twitter and Myspace among young Spanish people. *Behaviour & Information Technology*, *34*(7), 685–703. https://doi.org/10.1080/0144929X.2014.993428

Gerlitz, C., & Rieder, B. (2018). Tweets Are Not Created Equal: Investigating Twitter's Client Ecosystem. *International Journal of Communication*, *12*, 528–547. https://doi.org/1932–8036/20180005

Giannakos, M. N., Chorianopoulos, K., Giotopoulos, K., & Vlamos, P. (2013). Using Facebook out of habit. *Behaviour & Information Technology*, *32*(6), 594–602. https://doi.org/10.1080/0144929X.2012.659218

Gibson, J. J. (1979). The Theory of Affordances. In *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI '09* (p. 211). New York, New York, USA: ACM Press. https://doi.org/10.1145/1518701.1518736

Golbeck, J. (2005). *Computing and applying trust in web-based social networks. PhD Dissertation*. University of Maryland (College Park, Md.).

Golbeck, J., & Hendler, J. (2006). Inferring binary trust relationships in Web-based social networks. *ACM Transactions on Internet Technology*, *6*(4), 497–529. https://doi.org/10.1145/1183463.1183470

Golder, S. A., & Yardi, S. (2010). Structural predictors of tie formation in twitter: Transitivity and mutuality. In *2nd IEEE International Conference on Social Computing* (pp. 88–95). Minneapolis, MN: IEEE.

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The Dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, *1*(1), 197. https://doi.org/10.1038/srep00197

Grandhi, S. A., Plotnick, L., & Hiltz, S. R. (2019). Do I Stay or Do I Go?: Motivations and Decision Making in Social Media Non-use and Reversion. *Proceedings of the ACM on Human-Computer Interaction*, *3*(GROUP), 1–27. https://doi.org/10.1145/3361116

Granovetter, M. S. (1973). The Strength of Weak Ties. *The Journal of Applied Psychology*, *78*(6), 1360–1380.

Greene, K., Derlega, V. J., & Mathews, A. (2006). Self-Disclosure in Personal Relationships. In A. L. Vangelisti & D. Perlman (Eds.), *The Cambridge Handbook of Personal Relationships* (pp. 409–428). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511606632.023

Greitemeyer, T., Mügge, D. O., & Bollermann, I. (2014). Having Responsive Facebook Friends Affects the Satisfaction of Psychological Needs More Than Having Many Facebook Friends. *Basic and Applied Social Psychology*, *36*(3), 252–258. https://doi.org/10.1080/01973533.2014.900619

Hall, M., Mazarakis, A., Chorley, M., & Caton, S. (2018). Editorial of the Special Issue on Following User Pathways: Key Contributions and Future Directions in Cross-Platform Social Media Research. *International Journal of Human–Computer Interaction*, *34*(10), 895–912.

https://doi.org/10.1080/10447318.2018.1471575

Hann, I. H., Hui, K. L., Lee, S. Y. T., & Png, I. P. L. (2007). Analyzing online information privacy concerns: An information processing theory approach. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *24*(2), 13–42. https://doi.org/10.1109/HICSS.2007.81

Harrison, S., Tatar, D., & Sengers, P. (2007). The three paradigms of HCI. In *Extended abstracts CHI 2007*. https://doi.org/10.1234/12345678

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach* (2nd ed.). New York: Guilford Press.

Haythornthwaite, C. (2002). Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society*, *18*(5), 385–401. https://doi.org/10.1080/01972240290108195

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., … Verplank, W. (1992). *ACM SIGCHI Curricula for Human-Computer Interaction*. Retrieved from http://portal.acm.org/citation.cfm?id=133967

Hogan, B. (2007). Using Information Networks to Study Social Behavior: An Appraisal. *IEEE Data Engineering Bulletin*, *30*(2), 6–14.

Hogan, B. (2010a). The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. *Bulletin of Science, Technology & Society*, *30*(6), 377–386. https://doi.org/10.1177/0270467610385893

Hogan, B. (2010b). Visualizing and Interpreting Facebook Networks. In D. L. Hansen, B. Shneiderman, & M. A. Smith (Eds.), *Analyzing Social Media Networks with NodeXL* (pp. 165–180). Burlington, MA: Morgan Kaufmann.

Hogan, B. (2018). Social Media Giveth, Social Media Taketh Away: Facebook, Friendships, and APIs. *International Journal of Communication*, *12*(20), 592–611.

Homans, G. (1958). Social Behavior as Exchange. *American Journal of Sociology, 63*(6), 597–606. https://doi.org/10.1086/222355

Horrigan, J. B. (2016). *Information overload. Pew Research Center.* Retrieved from http://www.pewinternet.org/2016/12/07/information-overload/

Houghton, D., Joinson, A., Caldwell, N., Marder, B., & Collins, E. (2018). Photographic Disclosure in Facebook and Relational Closeness with Others. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (Vol. 9, pp. 2078–2087).

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems*, *12*(12), 767–797.

Huang, H. Y. (2016). Examining the beneficial effects of individual's self-disclosure on the social network site. *Computers in Human Behavior*, *57*, 122–132. https://doi.org/10.1016/j.chb.2015.12.030

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy

Protection. *Computer*, *51*(8), 56–59. https://doi.org/10.1109/MC.2018.3191268

Islam, A. K. M. N., & Patil, S. (2015). Engagement and Well-being on Social Network Sites. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15* (pp. 375–382). New York, New York, USA: ACM Press. https://doi.org/10.1145/2675133.2675299

Jiang, L. C., Bazarova, N. N., & Hancock, J. T. (2011). The Disclosure-Intimacy Link in Computer-Mediated Communication: An Attributional Extension of the Hyperpersonal Model. *Human Communication Research*, *37*(1), 58–77. https://doi.org/10.1111/j.1468-2958.2010.01393.x

John, L. K., Acquisti, A., & Loewenstein, G. (2010). Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. *Journal of Consumer Research*, *37*(5), 858–873. https://doi.org/10.1086/656423

Johnson, P., & Yang, S. (2009). Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use. In *Communication Technology Division of the annual convention of the Association for Education in Journalism and Mass Communication in Boston, MA* (Vol. 54). MIT Press.

Johnson, T. J., & Kaye, B. K. (2015). Reasons to believe: Influence of credibility on motivations for using social networks. *Computers in Human Behavior*, *50*, 544–555. https://doi.org/10.1016/j.chb.2015.04.002

Joinson, A. (2001). Knowing Me, Knowing You: Reciprocal Self-Disclosure in Internet-Based Surveys. *CyberPsychology & Behavior*, *4*(5), 587–591. https://doi.org/10.1089/109493101753235179

Joinson, A. (2008). Looking at, looking up or keeping up with people? In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* (p. 1027). New York, New York, USA: ACM Press. https://doi.org/10.1145/1357054.1357213

Joinson, A., & Paine, C. B. (2007). Self-disclosure, Privacy and the Internet. In A. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *Oxford Handbook of Internet Psychology* (pp. 235–250). Oxford University Press.

Joinson, A., Reips, U.-D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, Trust, and Self-Disclosure Online. *Human-Computer Interaction*, *25*(1), 1–24. https://doi.org/10.1080/07370020903586662

Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C., & Fowler, J. H. (2013). Inferring Tie Strength from Online Directed Behavior. *PLoS ONE*, *8*(1), e52168. https://doi.org/10.1371/journal.pone.0052168

Jordan, K. (2018). Validity, Reliability, and the Case for Participant-Centered Research: Reflections on a Multi-Platform Social Media Study. *International Journal of Human–Computer Interaction*, 1–9. https://doi.org/10.1080/10447318.2018.1471570

Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, *29*(3), 626–631. https://doi.org/10.1016/j.chb.2012.11.007

Kairam, S., Brzozowski, M., Huffaker, D., & Chi, E. (2012). Talking in circles: selective sharing in

Google+. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 1065–1074). https://doi.org/10.1145/2207676.2208552

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36.

Karnik, M., Oakley, I., Venkatanathan, J., Spiliotopoulos, T., & Nisi, V. (2013). Uses & gratifications of a facebook media sharing group. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13* (p. 821). New York, New York, USA: ACM Press. https://doi.org/10.1145/2441776.2441868

Kashian, N., Jang, J., Shin, S. Y., Dai, Y., & Walther, J. B. (2017). Self-disclosure and liking in computer-mediated communication. *Computers in Human Behavior*, *71*, 275–283. https://doi.org/10.1016/j.chb.2017.01.041

Katz, E., Gurevitch, M., & Haas, H. (1973). On the use of the mass media for important things. *American Sociological Review*, *38*, 164–181.

Kaur, H., Johnson, I., Miller, H. J., Terveen, L. G., Lampe, C., Hecht, B., & Lasecki, W. S. (2018). Oh The Places You'll Share: An Affordances-Based Model of Social Media Posting Behaviors. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (pp. 1–6). New York, New York, USA: ACM Press. https://doi.org/10.1145/3170427.3188601

Kelley, P. G., Brewer, R., Mayer, Y., Cranor, L. F., & Sadeh, N. (2011). An investigation into facebook friend grouping. In *IFIP Conference on Human-Computer Interaction* (pp. 216–233). Springer Berlin Heidelberg.

Kim, B., & Kim, Y. (2019). Facebook versus Instagram: How perceived gratifications and technological attributes are related to the change in social media usage. *The Social Science Journal*, *56*(2), 156–167. https://doi.org/10.1016/j.soscij.2018.10.002

Kim, J., & Lee, J.-E. R. (2011). The Facebook Paths to Happiness: Effects of the Number of Facebook Friends and Self-Presentation on Subjective Well-Being. *Cyberpsychology, Behavior, and Social Networking*, *14*(6), 359–364. https://doi.org/10.1089/cyber.2010.0374

Kizilcec, R. F., Bakshy, E., Eckles, D., & Burke, M. (2018). Social Influence and Reciprocity in Online Gift Giving. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (pp. 1–11). New York, New York, USA: ACM Press. https://doi.org/10.1145/3173574.3173700

Kneidinger-Müller, B. (2018). Self-Tracking Data as Digital Traces of Identity: A Theoretical Analysis of Contextual Factors of Self-Observation Practices. *International Journal of Communication*, *12*, 629–646.

Knijnenburg, B. P., Kobsa, A., & Jin, H. (2013). Dimensionality of information disclosure behavior. *International Journal of Human Computer Studies*, *71*(12), 1144–1162. https://doi.org/10.1016/j.ijhcs.2013.06.003

Knobloch-Westerwick, S. (2014). *Choice and preference in media use: Advances in selective exposure theory and research*. Routledge.

Knoke, D., & Yang, S. (2008). *Social Network Analysis* (2nd ed.). California: Sage Publications, Inc.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers and Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. In M. Smith & P. Kollock (Eds.), *Communities in Cyberspace* (pp. 220–239). London: Routledge.

Koroleva, K., Krasnova, H., & Günther, O. (2010). "Stop Spamming me!" - Exploring Information Overload on Facebook. In *Proceedings of the Sixteenth Americas Conference on Information Systems (AMCIS)* (pp. 1–8).

Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks*, *28*(3), 247–268. https://doi.org/10.1016/j.socnet.2005.07.002

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: why we disclose. *Journal of Information Technology*, *25*(2), 109–125. https://doi.org/10.1057/jit.2010.6

Krasnova, H., Veltri, N. F., Eling, N., & Buxmann, P. (2017). Why men and women continue to use social networking sites: The role of gender differences. *The Journal of Strategic Information Systems*. https://doi.org/10.1016/j.jsis.2017.01.004

Kraut, R., Maher, M. L., Olson, J., Malone, T. W., Pirolli, P., & Thomas, J. C. (2010). Scientific Foundations: A Case for Technology- Mediated Social- Participation Theory. *Computer*, *43*(11), 22–28. https://doi.org/10.1109/MC.2010.324

Krcmar, M., & Strizhakova, Y. (2009). Uses and Gratifications as Media Choice. In T. Hartmann (Ed.), *Media Choice: A Theoretical and Empirical Overview* (pp. 53–69).

Ku, Y.-C., Chen, R., & Zhang, H. (2013). Why do users continue using social networking sites? An exploratory study of members in the United States and Taiwan. *Information & Management*, *50*(7), 571–581. https://doi.org/10.1016/j.im.2013.07.011

Lampe, C., Ellison, N., & Steinfield, C. (2006). A face(book) in the crowd: Social Searching vs. Social Browsing. In *Proceedings of the 2006 conference on Computer supported cooperative work - CSCW '06* (p. 167). New York, New York, USA: ACM Press. https://doi.org/10.1145/1180875.1180901

Lampe, C., Vitak, J., & Ellison, N. (2013). Users and Nonusers: Interactions between Levels of Facebook Adoption and Social Capital. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13*. New York, New York, USA: ACM Press. https://doi.org/10.1145/2441776.2441867

Lampe, C., Wash, R., Velasquez, A., & Ozkaya, E. (2010). Motivations to participate in online communities. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (p. 1927). New York, New York, USA: ACM Press. https://doi.org/10.1145/1753326.1753616

Lampinen, A. (2016). Why we need to examine multiple social network sites. *Communication and the Public*, *1*(4), 489–493. https://doi.org/10.1177/2057047316681171

Laufer, R. S., & Wolfe, M. (1977). Privacy as a Concept and a Social Issue: A Multidimensional

Developmental Theory. *Journal of Social Issues*, *33*(3), 22–42. https://doi.org/10.1111/j.1540-4560.1977.tb01880.x

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science (New York, N.Y.)*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., … Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science (New York, N.Y.)*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lee, A. R., Son, S.-M., & Kim, K. K. (2016). Information and communication technology overload and social networking service fatigue: A stress perspective. *Computers in Human Behavior*, *55*, 51–61. https://doi.org/10.1016/j.chb.2015.08.011

Lee, D.-H. (2009). Mobile Snapshots and Private/Public Boundaries. *Knowledge, Technology & Policy*, *22*(3), 161–171. https://doi.org/10.1007/s12130-009-9081-0

Levordashka, A., Utz, S., & Ambros, R. (2016). What's in a like? Motivations for pressing the like button. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*, 623–626.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, *30*(4), 330–342.

Lin, C.-P., & Bhattacherjee, A. (2008). Elucidating Individual Intention to Use Interactive Information Technologies: The Role of Network Externalities. *International Journal of Electronic Commerce*, *13*(1), 85–108. https://doi.org/10.2753/JEC1086-4415130103

Lin, H., & Qiu, L. (2013). Two sites, two voices: Linguistic differences between facebook status updates and tweets. *Lecture Notes in Computer Science*, *8024 LNCS*(PART 2), 432–440. https://doi.org/10.1007/978-3-642-39137-8-48

Lin, K.-M. (2016). Understanding undergraduates' problems from determinants of Facebook continuance intention. *Behaviour & Information Technology*, *35*(9), 693–705. https://doi.org/10.1080/0144929X.2016.1177114

Lin, R., & Utz, S. (2015). The emotional responses of browsing Facebook: Happiness, envy, and the role of tie strength. *Computers in Human Behavior*, *52*, 29–38. https://doi.org/10.1016/j.chb.2015.04.064

Lin, Y., Margolin, D., Keegan, B., Baronchelli, A., & Lazer, D. (2013). #Bigbirds Never Die: Understanding Social Dynamics of Emergent Hashtags. In *ICWSM* (pp. 370–379).

Liu, I. L. B., Cheung, C. M. K., & Lee, M. K. O. (2010). Understanding Twitter usage : What drive people continue to tweet. *Proceedings of the Pacific Asia Conference on Information Systems*, 927–939.

Lomborg, S., & Bechmann, A. (2014). Using APIs for Data Collection on Social Media. *The Information Society*, *30*(4), 256–265. https://doi.org/10.1080/01972243.2014.915276

Luarn, P., & Chiu, Y.-P. (2015). Key variables to predict tie strength on social network sites. *Internet Research*, *25*(2), 218–238. https://doi.org/10.1108/IntR-11-2013-0231

Malik, A., Dhir, A., & Nieminen, M. (2016). Uses and Gratifications of digital photo sharing on Facebook. *Telematics and Informatics*, *33*(1), 129–138. https://doi.org/http://dx.doi.org/10.1016/j.tele.2015.06.009

Malik, A., Hiekkanen, K., Dhir, A., & Nieminen, M. (2016). Impact of privacy, trust and user activity on intentions to share Facebook photos. *Journal of Information, Communication and Ethics in Society*, *14*(4), 364–382. https://doi.org/10.1108/JICES-06-2015-0022

Marder, B., Joinson, A., & Shankar, A. (2011). Every post you make, every pic you take, I'll be watching you: Behind social spheres on facebook. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 859–868). https://doi.org/10.1109/HICSS.2012.12

Marder, B., Joinson, A., Shankar, A., & Thirlaway, K. (2016). Strength matters: Self-presentation to the strongest audience rather than lowest common denominator when faced with multiple audiences in social network sites. *Computers in Human Behavior*, *61*, 56–62. https://doi.org/10.1016/j.chb.2016.03.005

Margolin, D., Lin, Y.-R., Brewer, D., & Lazer, D. (2013). 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* (pp. 9–10).

Markus, M. L. (1987). Toward a "Critical Mass" Theory of Interactive Media: Universal Access, Interdependence and Diffusion. *Communication Research*, *14*(5), 491–511. https://doi.org/10.1177/009365087014005003

Marsden, P. V., & Campbell, K. E. (1984). Measuring Tie Strength. *Social Forces*, *63*(2), 482–501.

Marwick, A. E., & boyd, d. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

Mayol, A., & Pénard, T. (2017). Facebook use and individual well-being: Like me to make me happier!. In *Revue d'économie industrielle* (pp. 101–127). De Boeck Supérieur.

Mendelson, A. L., & Papacharissi, Z. (2010). Look at us: Collective Narcissism in College Student Facebook Photo Galleries. In Z. Papacharissi (Ed.), *The Networked Self: Identity, Community and Culture on Social Network Sites* (pp. 251–273). Routledge.

Mesch, G. S. (2012). Is online trust and trust in social institutions associated with online disclosure of identifiable information online? *Computers in Human Behavior*, *28*(4), 1471–1477. https://doi.org/10.1016/j.chb.2012.03.010

Millham, M. H., & Atkin, D. (2016). Managing the virtual boundaries: Online social networks, disclosure, and privacy behaviors. *New Media & Society*. https://doi.org/10.1177/1461444816654465

Moll, R., Pieschl, S., & Bromme, R. (2017). Whoever will read it – The overload heuristic in collective privacy expectations. *Computers in Human Behavior*, *75*, 484–493. https://doi.org/10.1016/j.chb.2017.05.035

Morris, M. R., Teevan, J., & Panovich, K. (2010a). A comparison of information seeking using search engines and social networks. In *Proceedings of the Fourth International AAAI Conference* *on Weblogs and Social Media (ICWSM)* (pp. 291–294).

Morris, M. R., Teevan, J., & Panovich, K. (2010b). What Do People Ask their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (p. 1739). New York, New York, USA: ACM Press. https://doi.org/10.1145/1753326.1753587

Neubaum, G., & Krämer, N. C. (2015). My Friends Right Next to Me: A Laboratory Investigation on Predictors and Consequences of Experiencing Social Closeness on Social Networking Sites. *Cyberpsychology, Behavior, and Social Networking*, *18*(8), 443–449. https://doi.org/10.1089/cyber.2014.0613

Nielsen, F. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big things come in small packages* (pp. 93–98).

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox : Personal Information Disclosure Intentions versus Behaviors. *The Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1083-6101.2009.01494.x

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, *6*(3), 38–43. https://doi.org/10.1145/301153.301168

Nunamaker, J. F. J., Chen, M., & Purdin, T. D. M. (1990). Systems development in information systems research. *Journal of Management Information Systems*, *7*(3), 89–106.

Ozenc, F. K., & Farnham, S. D. (2011). Life "modes" in social media. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11* (p. 561). New York, New York, USA: ACM Press. https://doi.org/10.1145/1978942.1979022

Pai, P., & Arnott, D. C. (2013). User adoption of social networking sites: Eliciting uses and gratifications through a means–end approach. *Computers in Human Behavior*, *29*(3), 1039–1053. https://doi.org/10.1016/j.chb.2012.06.025

Paik, A., & Sanchagrin, K. (2013). Social Isolation in America: An Artifact. *American Sociological Review*, *78*(3), 339–360. https://doi.org/10.1177/0003122413482919

Panovich, K., Miller, R., & Karger, D. (2012). Tie strength in question & answer on social network sites. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, 1057. https://doi.org/10.1145/2145204.2145361

Papacharissi, Z. (2008). Uses and Gratifications. In M. Salwen & D. Stacks (Eds.), *An Integrated Approach to Communication Theory and Research* (pp. 137–152). Lawrence Erlbaum.

Papacharissi, Z., & Mendelson, A. (2011). Toward a New(er) Sociability: Uses, Gratifications and Social Capital on Facebook. In S. Papathanassopoulos (Ed.), *Media Perspectives for the 21st Century*. Routledge.

Park, C. S. (2013). Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Computers in Human Behavior*, *29*(4), 1641–1648.

https://doi.org/10.1016/j.chb.2013.01.044

Park, N., Jin, B., & Annie Jin, S.-A. (2011). Effects of self-disclosure on relational intimacy in Facebook. *Computers in Human Behavior*, *27*(5), 1974–1983. https://doi.org/10.1016/j.chb.2011.05.004

Pelaprat, E., & Brown, B. (2012). Reciprocity: Understanding online social relations. *First Monday*, *17*(10). https://doi.org/10.5210/fm.v17i10.3324

Pereira, D. (2014). *A Feasibility Study for Modelling Tie Strength with the Facebook API*. University of Madeira. https://doi.org/http://hdl.handle.net/10400.13/521

Pew Research Center. (2013). *Photo and Video Sharing Grow Online*. Retrieved from http://www.pewinternet.org/2013/10/28/photo-and-video-sharing-grow-online/

Pew Research Center. (2015). *The Evolving Role of News on Twitter and Facebook*.

Pew Research Center. (2018). *Social Media Use in 2018*. *Pew Research Center*. Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

Phan, K. L., Sripada, C. S., Angstadt, M., & McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(29), 13099–13104. https://doi.org/10.1073/pnas.1008137107

Phua, J., Jin, S. V., & Kim, J. J. (2017a). Gratifications of using Facebook, Twitter, Instagram, or Snapchat to follow brands: The moderating effect of social comparison, trust, tie strength, and network homophily on brand identification, brand engagement, brand commitment, and membership intention. *Telematics and Informatics*, *34*(1), 412–424. https://doi.org/10.1016/j.tele.2016.06.004

Phua, J., Jin, S. V., & Kim, J. J. (2017b). Uses and Gratifications of Social Networking Sites for Bridging and Bonding Social Capital: A Comparison of Facebook, Twitter, Instagram, and Snapchat. *Computers in Human Behavior*, *72*, 115–122. https://doi.org/10.1016/j.chb.2017.02.041

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155–167. https://doi.org/10.1016/j.chb.2016.03.084

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Quan-Haase, A., Martin, K., & McCay-Peet, L. (2015). Networks of digital humanities scholars: The informational and social uses and gratifications of Twitter. *Big Data & Society*, *2*(1), 205395171558941. https://doi.org/10.1177/2053951715589417

Quan-Haase, A., & Young, A. L. (2010). Uses and Gratifications of Social Media: A Comparison of Facebook and Instant Messaging. *Bulletin of Science, Technology & Society*, *30*(5), 350–361. https://doi.org/10.1177/0270467610380009

Quan-Haase, A., & Young, A. L. (2014). The Uses and Gratifications (U&G) Approach as a Lens for Studying Social Media Practice. In *The Handbook of Media and Mass Communication Theory* (pp. 269–286). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118591178.ch15

Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., & Crowcroft, J. (2012). The personality of popular facebook users. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, 955. https://doi.org/10.1145/2145204.2145346

Quinn, K. (2016). Why We Share: A Uses and Gratifications Approach to Privacy Regulation in Social Media Use. *Journal of Broadcasting & Electronic Media*, *60*(1), 61–86. https://doi.org/10.1080/08838151.2015.1127245

Quinn, K., & Papacharissi, Z. (2018). The contextual accomplishment of privacy. *International Journal of Communication*, *12*(0), 45–67.

Raacke, J., & Bonds-Raacke, J. (2008). MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *11*(2), 169–174. https://doi.org/10.1089/cpb.2007.0056

Rae, J. R., & Lonborg, S. D. (2015). Do motivations for using Facebook moderate the association between Facebook use and psychological well-being? *Frontiers in Psychology*, *6*(June), 771. https://doi.org/10.3389/fpsyg.2015.00771

Rains, S. A., & Ruppel, E. K. (2016). Channel Complementarity Theory and the Health Information-Seeking Process. *Communication Research*, *43*(2), 232–252. https://doi.org/10.1177/0093650213510939

Resnick, P. (2002). Beyond Bowling Together: SocioTechnical Capital. In J. M. Carroll (Ed.), *HCI in the New Millenium* (pp. 247–272). Addison-Wesley.

Resnick, P., Adar, E., & Lampe, C. (2015). What Social Media Data We Are Missing and How to Get It. *Annals of the American Academy of Political and Social Science*, *659*(1), 192–206. https://doi.org/10.1177/0002716215570006

Richardson, K., & Hessey, S. (2009). Archiving the self? Facebook as biography of social and relational memory. *Journal of Information, Communication and Ethics in Society*, *7*(1), 25–38. https://doi.org/10.1108/14779960910938070

Rieder, B. (2013). Studying Facebook via Data Extraction: The Netvizz Application. In *ACM WebScience 2013*.

Riegelsberger, J., Sasse, M. A., & Mccarthy, J. D. (2005). The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies*, *62*(3), 381–422. https://doi.org/10.1016/j.ijhcs.2005.01.001

Rubin, A. M., & Windahl, S. (1986). The Uses and dependency model of mass communication. *Critical Studies in Mass Communication*, *3*(2), 184–199. https://doi.org/10.1080/15295039609366643

Ruehl, C. H., & Ingenhoff, D. (2015). Communication management on social networking sites. *Journal of Communication Management*, *19*(3), 288–302. https://doi.org/10.1108/JCOM-04-2015-0025

Ruggiero, T. E. (2000). Uses and Gratifications Theory in the 21st Century. *Mass Communication and Society*, *3*(1), 3–37. https://doi.org/10.1207/S15327825MCS0301_02

Salminen, J., Şengün, S., Kwak, H., Jansen, B. J., An, J., Jung, S., … Harrell, D. F. (2018). From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas. *First Monday*, *23*(6). https://doi.org/10.5210/fm.v23i6.8415

Satchell, C., & Dourish, P. (2009). Beyond the user: Use And Non-Use in HCI. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group on Design: Open 24/7 - OZCHI '09* (p. 9). New York, New York, USA: ACM Press. https://doi.org/10.1145/1738826.1738829

Sato, Y. (2013). Rational choice theory. *Sociopedia.Isa*. https://doi.org/0.1177/205684601372

Schensul, J. J., & Burkholder, G. J. (2005). Vulnerability, Social Networks, Sites, and Selling as Predictors of Drug use among Urban African American and Puerto Rican Emerging Adults. *Journal of Drug Issues*, *35*(2), 379–408. https://doi.org/10.1177/002204260503500208

Schoenebeck, S. Y. (2014). Giving up Twitter for Lent: How and Why We Take Breaks from Social Media. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 773–782). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556288.2556983

Scott, J. (2000). Rational choice theory. In G. Brawning, A. Halcli, & F. Webster (Eds.), *Understanding Contemporary Society: Theories of the Present* (pp. 126–138). Sage Publications.

Sen, A. (1997). Maximization and the Act of Choice. *Econometrica*, *65*(4), 745–779.

Shankar, V., Smith, A. K., & Rangaswamy, A. (2003). Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, *20*(2), 153–175. https://doi.org/10.1016/S0167-8116(03)00016-8

Sheldon, P. (2009). "I'll poke you. You'll poke me!" Self-disclosure, social attraction, predictability and trust as important predictors of Facebook relationships. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *3*(2).

Sleeper, M., Balebako, R., Das, S., McConahy, A. L., Wiese, J., & Cranor, L. F. (2013). The post that wasn't: exploring self-censorship on facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13* (p. 793). New York, New York, USA: ACM Press. https://doi.org/10.1145/2441776.2441865

Sleeper, M., Melicher, W., Habib, H., Bauer, L., Cranor, L. F., & Mazurek, M. L. (2016). Sharing Personal Content Online: Exploring Channel Choice and Multi-Channel Behaviors. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 101–112). New York, New York, USA: ACM Press. https://doi.org/10.1145/2858036.2858170

Smith, H. J., Dinev, T., & Xu, H. (2011). Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly*, *35*(4), 989–1015.

Smock, A. D., Ellison, N., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior*, *27*(6), 2322–

2329. https://doi.org/10.1016/j.chb.2011.07.011

Snodgrass, E., & Soon, W. (2019). API practices and paradigms: Exploring the protocological parameters of APIs as key facilitators of sociotechnical forms of exchange. *First Monday*, *24*(2). https://doi.org/10.5210/fm.v24i2.9553

Spiliotopoulos, T., Karnik, M., Oakley, I., Venkatanathan, J., & Nisi, V. (2013). Towards Understanding Social Media: Two Studies Exploring the Uses and Gratifications of Facebook. In *HCI Korea 2013* (pp. 1405–1408).

Spiliotopoulos, T., & Oakley, I. (2012). Applications of Social Network Analysis for User Modeling. In *International Workshop on User Modeling from Social Media, IUI 2012*. Lisbon, Portugal.

Spiliotopoulos, T., & Oakley, I. (2013a). Replicating and Extending a Facebook Uses & Gratifications Study: Five Years Later. In *Proceedings of the CHI2013 Workshop on the Replication of HCI Research*. Paris, France.

Spiliotopoulos, T., & Oakley, I. (2013b). Understanding Motivations for Facebook Use: Usage Metrics, Network Structure, and Privacy. In *Proceedings of the 2013 ACM annual conference on Human Factors in Computing Systems - CHI '13* (pp. 3287–3296). ACM. https://doi.org/10.1145/2470654.2466449

Spiliotopoulos, T., & Oakley, I. (2015). An exploratory study on the use of Twitter and Facebook in tandem. In *Proceedings of the 2015 British HCI Conference* (pp. 299–300). New York, New York, USA: ACM Press. https://doi.org/10.1145/2783446.2783620

Spiliotopoulos, T., & Oakley, I. (2016). Post or Tweet: Lessons from a Study of Facebook and Twitter Usage. In *Following User Pathways: Using Cross Platform and Mixed Methods Analysis in Social Media Studies Workshop at ACM CHI 2016*. https://arxiv.org/abs/2011.13802v1

Spiliotopoulos, T., & Oakley, I. (2019). Altruistic and selfish communication on social media: the moderating effects of tie strength and interpersonal trust. *Behaviour & Information Technology*. https://doi.org/10.1080/0144929X.2019.1688392

Spiliotopoulos, T., & Oakley, I. (2020). An exploration of motives and behavior across Facebook and Twitter. *Journal of Systems and Information Technology*, *22*(2), 201–222. https://doi.org/10.1108/JSIT-12-2019-0258

Spiliotopoulos, T., Pereira, D., & Oakley, I. (2014). Predicting Tie Strength with the Facebook API. In *Proceedings of the 18th Panhellenic Conference on Informatics - PCI '14*. ACM. https://doi.org/10.1145/2645791.2645817

Stafford, T. F., Stafford, M. R., & Schkade, L. L. (2004). Determining uses and gratifications for the internet. *Decision Sciences*, *35*(2), 259–288.

Statista. (2018). Device usage of Facebook users worldwide as of January 2018. Retrieved January 7, 2019, from https://www.statista.com/statistics/377808/distribution-of-facebook-users-by-device/

Steinfield, C., Ellison, N., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, *29*(6), 434–445. https://doi.org/10.1016/j.appdev.2008.07.002

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2019). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, 089443931984366. https://doi.org/10.1177/0894439319843669

Sundar, S. S., & Limperos, A. M. (2013). Uses and Grats 2.0: New Gratifications for New Media. *Journal of Broadcasting & Electronic Media*, *57*(4), 504–525. https://doi.org/10.1080/08838151.2013.845827

Taddei, S., & Contena, B. (2013). Privacy, trust and control: Which relationships with online self-disclosure? *Computers in Human Behavior*, *29*(3), 821–826. https://doi.org/10.1016/j.chb.2012.11.022

Taddicken, M. (2014). The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure. *Journal of Computer-Mediated Communication*, *19*(2), 248–273. https://doi.org/10.1111/jcc4.12052

Tidwell, L. C., & Walther, J. B. (2002). Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time. *Human Communication Research*, *28*(3), 317–348. https://doi.org/10.1093/hcr/28.3.317

Tosun, L. P. (2012). Motives for Facebook use and expressing "true self" on the Internet. *Computers in Human Behavior*, *28*(4), 1510–1517. https://doi.org/10.1016/j.chb.2012.03.018

Trepte, S., Scharkow, M., & Dienlin, T. (2020). The privacy calculus contextualized: The influence of affordances. *Computers in Human Behavior*, *104*. https://doi.org/10.1016/j.chb.2019.08.022

Twitter. (2019). Selected company metrics and financials - Q1 2019. Retrieved June 11, 2019, from https://s22.q4cdn.com/826641620/files/doc_financials/2019/q1/Q1-2019-Selected-Company-Metrics-and-Financials.pdf

Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. (2012). Structural diversity in social contagion. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 5962–5966. https://doi.org/10.1073/pnas.1116502109

Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The Anatomy of the Facebook Social Graph. *ArXiv:1111.4503v1*.

Urista, M. A., Dong, Q., & Day, K. D. (2009). Explaining Why Young Adults Use MySpace and Facebook Through Uses and Gratifications Theory. *Human Communication*, *12*(2), 215–229.

Utz, S., Muscanell, N., & Khalid, C. (2015). Snapchat Elicits More Jealousy than Facebook: A Comparison of Snapchat and Facebook Use. *Cyberpsychology, Behavior, and Social Networking*, *18*(3), 141–146. https://doi.org/10.1089/cyber.2014.0479

Valenzuela, S., Correa, T., & Gil de Zúñiga, H. (2018). Ties, Likes, and Tweets: Using Strong and Weak Ties to Explain Differences in Protest Participation Across Facebook and Twitter Use. *Political Communication*, *35*(1), 117–134. https://doi.org/10.1080/10584609.2017.1334726

Vasalou, A., Joinson, A., & Courvoisier, D. (2010). Cultural differences, experience with social networks and the nature of "true commitment" in Facebook. *International Journal of Human-Computer Studies*, *68*(10), 719–728. https://doi.org/10.1016/j.ijhcs.2010.06.002

Vitak, J. (2012). The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting & Electronic Media*, *56*(4), 451–470. https://doi.org/10.1080/08838151.2012.732140

Vitak, J., & Kim, J. (2014). "You can't block people offline": Examining how Facebook's affordances shape the disclosure process. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14* (pp. 461–474). New York, New York, USA: ACM Press. https://doi.org/10.1145/2531602.2531672

Wang, Y.-C., Burke, M., & Kraut, R. (2016). Modeling Self-Disclosure in Social Networking Sites. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16* (Vol. 25, pp. 74–85). New York, New York, USA: ACM Press. https://doi.org/10.1145/2818048.2820010

Wang, Y.-C., Hinsberger, H., & Kraut, R. E. (2016). Does Saying This Make Me Look Good? How Posters and Outsiders Evaluate Facebook Updates. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 125–129). New York, New York, USA: ACM Press. https://doi.org/10.1145/2858036.2858502

Wang, Yanbo, Min, Q., & Han, S. (2016). Understanding the effects of trust and risk on individual behavior toward social media platforms: A meta-analysis of the empirical evidence. *Computers in Human Behavior*, *56*, 34–44. https://doi.org/10.1016/j.chb.2015.11.011

Wang, Yang, Leon, P. G., Acquisti, A., Cranor, L. F., Forget, A., & Sadeh, N. (2014). A field trial of privacy nudges for facebook. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2367–2376). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556288.2557413

Wanous, J. P., & Hudy, M. J. (2001). Single-Item Reliability: A Replication and Extension. *Organizational Research Methods*, *4*(4), 361–375. https://doi.org/10.1177/109442810144003

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Weiler, M. (2018). Measuring Real-World Tie Strength with Digital Footprint Data: An Assessment of Convergent Validity. In *International Conference on Information Systems (ICIS)*. San Francisco, USA: Association for Information Systems.

Wellman, B., Carrington, P. J., & Hall, A. (1988). Networks as Personal Communities. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Analysis* (pp. 130–184). Cambridge, UK: Cambridge University Press.

Wellman, B., & Wortley, S. (1990). Different Strokes from Different Folks: Community Ties and Social Support. *American Journal of Sociology*, *96*(3), 558–588.

Wheeless, L. R., & Grotz, J. (1976). Conceptualization and Measurement of Reported Self-Disclosure. *Human Communication Research*, *2*(4), 338–346. https://doi.org/10.1111/j.1468-2958.1976.tb00494.x

Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research*, *3*(3), 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x

Wilken, R. (2015). Mobile media and ecologies of location. *Communication Research and Practice*, *1*(1), 42–57. https://doi.org/10.1080/22041451.2015.1042423

Yang, C., & Brown, B. B. (2015). Factors involved in associations between Facebook use and college adjustment: Social competence, perceived usefulness, and use patterns. *Computers in Human Behavior*, *46*, 245–253. https://doi.org/10.1016/j.chb.2015.01.015

Yao, X., Phang, C. W., & Ling, H. (2015). Understanding the Influences of Trend and Fatigue in Individuals' SNS Switching Intention. In *2015 48th Hawaii International Conference on System Sciences* (Vol. 2015-March, pp. 324–334). IEEE. https://doi.org/10.1109/HICSS.2015.46

Yu, R. P. (2016). The relationship between passive and active non-political social media use and political expression on Facebook and Twitter. *Computers in Human Behavior*, *58*, 413–420. https://doi.org/10.1016/j.chb.2016.01.019

Zafarani, R., & Liu, H. (2014). Users Joining Multiple Sites: Distributions and Patterns. In *ICWSM 2014*.

Zhao, J., Wu, J., & Feng, X. (2011). Information propagation in online social networks : a tie-strength perspective. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-011-0445-x

Zhao, X., Lampe, C., & Ellison, N. (2016). The Social Media Ecology: User Perceptions, Strategies and Challenges. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 89–100). New York, New York, USA: ACM Press. https://doi.org/10.1145/2858036.2858333

Zhao, X., & Lindley, S. E. (2014). Curation through use: Understanding the Personal Value of Social Media. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2431–2440). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556288.2557291

Zhong, C., Chan, H., Karamshuk, D., Lee, D., & Sastry, N. (2017). Wearing Many (Social) Hats: How Different are Your Different Social Network Personae? In *ICWSM 2017*.

Zimmer, J. C., Arsal, R. E., Al-Marzouq, M., & Grover, V. (2010). Investigating online information disclosure: Effects of information relevance, trust and risk. *Information and Management*, *47*(2), 115–123. https://doi.org/10.1016/j.im.2009.12.003