The Utility of Using Social Media Networks for Data Collection in Survey Research

Research-in-Progress

Mirabeau, Laurent
Telfer School of Management
University of Ottawa
55 Laurier East
Ottawa, Ontario, K1N 6N5
Canada
mirabeau@telfer.uottawa.ca

Mignerat, Muriel
Telfer School of Management
University of Ottawa
55 Laurier East
Ottawa, Ontario, K1N 6N5
Canada
mignerat@telfer.uottawa.ca

Grangé, Camille
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal, Québec, H3T 2A7
Canada
camille.grange@hec.ca

Abstract

Social media networks (SMNs) such as Facebook, LinkedIn or Twitter seem appealing tools for matters of reaching potential candidates for survey or case study research. Yet scholars remain cautious about leveraging these platforms. This research in progress paper compares and discusses the benefits of six generic strategies for reaching survey candidates on SMNs, and argues that while their use has potential pitfalls, the upside for explanatory type research may outweigh its risks. Furthermore, the paper outlines the empirical setting of a study that has been conducted to assess our propositions, and in which LinkedIn was used to identify and solicit survey candidates.

Keywords: Social media networks, LinkedIn, case study, methodology
Introduction

Research in Management in general and Management Information Systems in particular often requires recruiting top managers and executives, i.e., knowledgeable sources and key informants on many organizational processes (Cycyota and Harrison 2006), for filling-up surveys or conducting interviews. Researchers are thus faced with two recurring problems: finding relevant respondents and convincing them to dedicate time and energy to answer questions. This step in the research process can be challenging, often requiring considerable time and or money and sometimes constituting a limiting factor in reaching data saturation (Miles and Huberman 1994). Social media networks (SMNs) could be useful for alleviating such issues and enhancing traditional data collection methods (Fielding et al. 2008). SMNs are digital platforms via which users (i) build an online profile, (ii) produce and search for content within the platform, (iii) connect with others, and (iv), traverse their and others’ list of connections (Kane et al. Forthcoming). Given these affordances, SMNs could be useful for researchers to select and contact appropriate candidates, gather survey data, as well as leverage participants’ social network for referral.

Thus, this research-in-progress study investigates the following questions: 1) How may SMNs facilitate or challenge survey research carried out in organizational contexts? 2) Which strategies for data collection may be used in these contexts and how do they compare in terms of their ability to deliver higher response rates? Our research context is that of a larger study that aims to analyze the demise of a global organization and our data collected using a survey of past employees and other stakeholders administered on LinkedIn, a SMN dedicated to building work profiles and connecting with co-workers or people met in the context of work. LinkedIn is a particularly relevant SMN for our purpose because it is primarily dedicated to professional networking and is now well established (with more than 200 million members in more than 200 countries and territories according to recent statistics1).

This paper first presents an overview of the literature on the challenges of collecting data in Management research. Next, it explains how and why SMNs might be useful or challenging for response rates and sampling validity by building on recent research and analyzing the unique features of SMNs. Subsequently, the paper describes six different types of data collection strategies that researchers might employ to engage survey candidates using SMNs and develops a set of hypotheses about their effectiveness in generating responses. The methodology for collecting data and the expected contributions for this research are then discussed.

Literature Review

Problems and Remedies Associated with Traditional Data Collection Methods

Top managers and executives are not only good informants, they are also often the only source of information for capturing key research variables (Cycyota and Harrison 2006). In a study looking at research in top management journals featuring executives as respondents, Cycyota and Harrison report response rates of 24% over the period between 1992 and 2003. Furthermore they identify two fundamental issues for researchers drawing on these hard to reach populations: first the response rates amongst managers and executives are traditionally significantly lower than in consumer and employee populations, and second this rate is steadily declining year after year. This study also projects that if researchers keep on with the same data collection methodologies, we could expect response rates as low as 4% by 2050. While care must be taken in validating a simple linear projection, the persisting downward trend would suggest the need for new data collection methods that could enhance participation rates and minimize nonresponse bias.

Interestingly, using the social network (here used in the usual sociological sense of the term, i.e., the network of relations one has with other individuals) of executives in order to contact potential respondents (mailed surveys) was found to lead to a higher response rates (Cycyota and Harrison 2006). It is shown that, by drawing on existing social ties in the industry, professional groups, university contacts to name a few, researchers are able to access people with greater motivation in spending the time required in surveys and interviews than if they are contacted directly through cold call mailings. In the field of

1 Source: www.linkedin.com last consulted April 29, 2013
sociology, two techniques are used in order to reach people in a given social network: snowball sampling and its variation, respondent driven sampling (Goodman 2011).

Snowball sampling may be one of the most commonly used techniques in social science. This approach is used to locate information-rich respondents (Patton 2002). In Sociology, Coleman (1958) and Goodman (1961) introduced snowball sampling as a means for studying the structure of social networks starting from a random sample of participants. With time, snowball sampling became more commonly used to refer to a nonprobability approach to sampling design in hard to reach populations. A second technique used to reach such populations is respondent-driven sampling (RDS), which employs a variant sampling strategy to collect data from hard-to-reach populations (Gile and Handcock 2010). Chain-referral sampling can be used as an umbrella term for the class of sampling methods in hard-to-reach populations in which a first set of study participants (i.e., the “seeds”) is selected from a convenience sample and then used for recruiting relevant others, e.g., others who have experienced a similar phenomenon of interest, share the same beliefs, or engage in the same practices (Biernacki and Waldorf 1981; Penrod et al. 2003).

**Using Social Media for Research**

While the recent emergence of social media platforms would suggest fertile new context for data collection for researchers in all fields and in particular in Management, our literature review has uncovered a paucity of studies either considering SMNs as novel methods or making explicit mention of it as a means for generating positive responses to surveys or interviews. Our initial scan of seminal journals in the field (such as MIS Quarterly, Information Systems Research, Journal of Information Technology, Academy of Management Journal, Academy of Management Review etc.), over the last ten years in fact did not identify a single article in either category. Accordingly we have broadened our review to include the fields of Sociology and Psychology that seem more permeable to such data collection methods to date.

The means traditionally used to support the difficult task of surveying hard-to-reach managers and executives (e.g., use of social relations, use of referrals chains) might be augmented thanks to social media. Social media typically refers to the set of Internet-based tools that support interpersonal connections, communication, and collaboration. It is described by the American Psychological Association (Phillips 2011) as a means to find research participants, a great alternative to the traditional undergraduates as respondents: social media can help researchers in many different ways such as collect data for interviews or surveys, perform content analysis, conduct data mining, analyze social networks via the Internet, and performing virtual ethnographies (Fielding et al. 2008) (note that, importantly, survey research administered through the Internet does not seem to yield different conclusions from traditional, paper-based, techniques (e.g., Howell et al. 2010)). For instance, the emergence of virtual communities has provided researchers with the possibility to adapt ethnographic enquiries and to study networks and social interactions in this context (Hine 2008).

**Hypotheses Development**

In the Information Systems field, Kane et al. (forthcoming) have recently provided a definition and thorough description of these new social media tools and their capabilities, referring to them as social media network in order to make a clear distinction with offline social networks (i.e., emphasize the fact that they are Internet based) and other digital tools that do not include the ability to formally connect users together. Hence, SMNs, such as Facebook, LinkedIn, Google+, Twitter, Youtube, and Flickr, are unique in that they are built around four key classes of features: those for users to build a profile, create and search for content, connect with others, and view and traverse these connections. All SMNs share these core capabilities even if each may go its own way in terms of design and implementation (e.g., choose to enable users to view and traverse the full network vs. only part of it).

The unique features of SMNs can be mapped to benefits they might provide to survey researchers. First of all, users’ profile information (which may be produced by both users and the platform provider) together with search features could enable researchers to identify potential respondents and create a reasonably comprehensive sampling frame. Second, the social connections and traversing functions could enable the use of referral chains, a way to increase individuals’ motivation to answer a survey. Tables 1 and 2 synthesize the potential pros and cons of using SMNs for collecting data in survey research.
<table>
<thead>
<tr>
<th>Benefits</th>
<th>Description</th>
</tr>
</thead>
</table>
| Non-response bias            | - Given that SMNs are open and non-discriminatory platforms (anyone can participate), researchers should be able to collect data from a variety of stakeholders, decreasing the extent of bias that non-responses might trigger.  
- Non-response bias can be assessed more easily when using SMNs as the demographic details and platform behaviors of SMNs users tend to be recorded and made visible (often part of the user ‘profile’). |
| Sampling frame               | - Well known SMNs have millions of members and should provide a good basis for a sampling frame (i.e., a list of the focal organization’s stakeholders) that would be difficult to construct otherwise.                                                                                                                                                                                                                                                                            |
| Tracking, pace and direction of sampling | - Using SMNs to collect survey data enables researchers to monitor responses and non-responses, and to adjust the pace of the data collection accordingly.                                                                                                                                                                                                                                                                                                                                 |

<table>
<thead>
<tr>
<th>Threats</th>
<th>Comment or suggestion for dealing with threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self selection bias: users of a certain SMN platform might be systematically different from those who don’t use it</td>
<td>- Researchers should ensure that the SMN platform they use for data collection is comprehensive enough. While Linked In, for example, is the de facto standard of work-related SMN, it is possible that certain organizations or activities are less present on it. Researchers should be aware of this threat and make sure that a focal SMN includes a wide diversity of their stakeholders of interest. Consistency could also be checked by collecting a sample from outside the SMN.</td>
</tr>
<tr>
<td>Sampling frame: the SMN may not include the list of ALL the population of interest and this list will depend on the SMN platform being used.</td>
<td>- In most cases, it is virtually impossible to build a complete sampling frame. Researchers should strive for high numbers. For example, a sampling frame of 100,000 individuals for our focal study is a strong indicator that there has been a reasonable coverage of the population of interest.</td>
</tr>
</tbody>
</table>
| Convenience samples of initial subjects (i.e., seeding bias) | - If a researcher is used as seed in sending the survey to potential respondents in his or her personal network, the sample might be biased in that only individuals closely related to the researcher would participate.  
- The research team could use multiple seeding points to distribute the survey or randomly select individuals within the sampling frame to act as seeds. Also, the research team could also use Ads features often available on such platforms (e.g., Facebook, LinkedIn) to target the whole sampling frame at random (e.g., Ramo and Prochaska 2012). |

**Generic Strategies**

In this section, we propose and describe six generic strategies (not mutually exclusive) that researchers could follow to collect data using SMNs. From those, we develop a set of hypotheses about the their comparative effectiveness in generating responses from potential participants.

**Direct Contact**

- **Strategy 1: reach 1st degree contacts**: Given that SMNs users are embedded in a network of relationships, one strategy would consist in using a research confederate or member of the research team to contact his/her relevant first-degree connections.
Strategy 2: reach 2nd degree contacted: Researchers could use the capability provided by SMNs platforms to identify and connect with 'friends of friends', i.e., to the connections of one’s connections.

Strategy 3: reach 3rd degree contacts: Researchers could use another capability of these platforms to traverse the network further away and get in touch with relevant other users who are not in one's close network (i.e., those located further away than two steps away from one’s network position).

H1: Response rates will increase for candidates that are situated closer in the network to the person sending the request. (i.e., RR strategy 1 > strategy 2 > strategy 3)

Referrals

Strategy 4: active referral: This would involve asking respondents to actively refer the requester to his/her relevant connections as identified by the researcher.

Strategy 5: passive referral: This strategy would consist in asking respondents to name relevant others in their network that they think would be knowledgeable for filling the survey as well.

In a study of network sampling using the AddHealth and Facebook networks, Mouw and Verdery (2012) define a sample from a social network as one in which current respondents help recruit new respondents through either active (when respondents drive recruitment by bringing new respondents) or passive (when they facilitate contact between researchers and would-be subjects) means. For them, even hard-to-reach respondents are more likely to participate and respond honestly when they are referred to by acquaintances or friends. This view is consistent with Cycyota and Harrison (2006) that find that “prior consent screening of executive samples will generate higher reporter response rates” (p.137) and that “sponsorship (and therefore access) by a member of the executive's or firm’s social network will increase the likelihood of response to mailed surveys” (p.138). Hence, we propose H2:

H2: Contacting people via a referral yields better response rates than contacting them directly (i.e., RR strategy 4 and 5 > strategies 1, 2, 3)

H3: Response rates will be higher for active referrals compared to passive ones (i.e., RR strategy 4 > strategy 5)

Affiliations

Strategy 6: reach participants holding common affiliation: Users on SMNs are able to connect to each other directly (i.e., become 'friends' or 'colleagues) but also indirectly via their common affiliation to certain interests, groups, or events (Breiger 1974). These patterns of connections are commonly referred to as affiliations and can enable researchers to get in touch with a set of individuals that gather in groups relevant to the topic or organization being studied.

While topical salience has been shown to increase response rates (Cycyota and Harrison 2006) (i.e. people more familiar with the research question are more likely to respond to a survey investigating such a topic), it seems plausible that social media platform salience could have a similar favorable impact on response rates.

H4: Response rates will be higher for people holding common affiliations.

Method

In this section we begin by presenting the case study, second we discuss our data collection procedures, and third we explain how the generic strategies were operationalized in the study context.

Case Study

This research study is part of a larger study that aims to analyze the demise of a large global technology company, which at its height featured more than 120,000 employees, generating revenues in excess of one billion dollars and operating in more than forty countries. The larger study investigates the reasons for failure of the once iconic company from a dominant player of the industry throughout most of the 1990s to its eventual demise and bankruptcy. One of the methods used in this larger study, is a survey of past
employees, customers, competitors, suppliers and industry experts such as journalists and academics, which was administered on LinkedIn, a social media platform. The survey, which was conducted over a period of six months by sending requests at regular by-monthly intervals and collecting and tracking response rates as surveys were either completed or declined, serves as the basis for the data set that we present as part of this study.

**Data Collection**

The first step in this process was to build a survey. To accomplish this task we used SurveyMonkey, an online surveying tool, which allowed us to send a link to the people surveyed. A unique identifier was generated for each candidate in order to track response rates and respondent demographics.

The second step in administering the survey was to identify potential candidates that would have first hand knowledge of the organization and that could provide their opinions on the reason for its eventual demise (i.e., the sampling frame). The search criteria from LinkedIn was therefore simple as we merely used a former employee’s LinkedIn account (a member of the greater study team) and searched for people that had listed our organization of interest in their profile. We used several keywords with the basic search tool, all of them variations of the company’s name throughout its history. The vast majority of the people that fit our search criteria had the organization listed as a past employer, while the remainder had made an explicit reference to the company name in their summary information profile, either as industry experts or as a consequence of working for a supplier or a customer. This first search produced over 100,000 candidates in the network of our colleague at the 1st, 2nd or 3rd degree, which represents sizable coverage of the entire population (given the size of the organization at its peak and its average turnover rate, we estimate the population having first hand knowledge of the organization of interest to be comprised between 200,000 and 500,000 individuals).

The number of degrees represents the network distance between the LinkedIn survey researcher and the profile of the person searched. People connected in the 1st degree have agreed to be connections of one another and share all their profile information with one another. Connections of the 2nd degree nature are people that have at least one person in common in their 1st degree connections, while 3rd degree connections are made of people that are connected to someone that in turn is connected to someone in their 1st degree group of connections. Thus the number of degrees between two people represents the social network distance between them.

The third step was to contact candidates to fill the survey. To do so, we instantiated each generic strategy to apply to the LinkedIn context, as described in the next section.

The fourth and final step in collecting our data was to systematically track our response rates. Given the unique identifier that was created for each survey request we were able to track surveys as filled or not and track this back to the various types of strategies employed in generating requests.

**Operationalizing the Strategies in the LinkedIn Context**

**Strategy 1: Direct Messaging of 1st Degree Contacts**

We asked 1st degree connections of the LinkedIn survey researcher for their participation via a LinkedIn message and sent them a link to the survey along with their unique identifier. This strategy was largely limited by the fact that our LinkedIn survey researcher had few first-degree connections that fit our search criteria.

**Strategy 2: Contact Requests to 2nd Degree Contacts**

Another interesting feature in LinkedIn is the ability to contact 2nd degree connections directly with a short paragraph in order to ask them to be LinkedIn connections. This does not require the requester to have the connection target’s email but asked to state the nature of his/her relationship with the person targeted. One of the options is to state you are or have been colleagues and worked for the same company. This option was selected for all people contacted with this 3rd strategy. One limitation in using this strategy on LinkedIn is the fact that the initial message asking to connect does not allow the requester to include a weblink which meant we had to work a two step process, first asking the target person to connect
stating our purpose to administer a survey, and second, once the connection was made to send the weblink to our SurveyMonkey page. A unique identifier allowed us to track respondent’s acceptance.

**Strategy 3: Direct Messaging of 3rd Degree Contacts**

In order to contact 3rd degree connections on LinkedIn, we used InMail, a messaging tool that allows users to contact anyone directly. However, given the restrictions on volume (30 InMails only are allowed for premium users per month), our data collection is limited in numbers.

**Strategy 4: Active Referral Using an “Introduction”**

A unique feature of LinkedIn is the ability to contact a 2nd degree connection via a 1st degree connection by asking for an ‘introduction’. The requester includes a message for the person making the introduction and one for the person being sought in the 2nd degree. The 1st degree connection may then forward the introduction or refuse to do so. In the case the introduction is made, the 2nd degree person can in turn also choose to accept the request or deny it. This strategy meant a two-step method with acceptance from two people, the person making the introduction and the candidate for the survey. A unique identifier allowed us to track respondent’s acceptance.

**Strategy 5: Passive Referral**

This strategy consisted in following up with people that were specifically named by some of the people that filled the survey and then suggested names of some of their contacts on LinkedIn. In such cases we followed the steps of strategy 2 but named the person that had suggested we contact them.

**Strategy 6: Reach Participants Holding Common Affiliations to a Group**

LinkedIn like other social platform allows users to register with interest groups moderated by the group’s creator. Our organization of interest, as a large company, has several groups of interest on LinkedIn. Our LinkedIn survey researcher applied for group membership to several groups and then created a generic message about the survey, posting the link to the survey along with a generic identifier to track group respondents.

**Potential Contributions and Limits**

Our study aims to make several important contributions to a topic that has received little attention in the field of Management.

First it provides empirical validation that social media platforms provide promising opportunities for augmenting traditional methods for collecting survey data in Management research. Our initial response rate of 23% is comparable with projected rates for mail surveys (Cycyota and Harrison 2006).

Second we make a methodological contribution by suggesting six alternative strategies for generating higher response rates via the leading professional social networking platform, LinkedIn. This is important as social media platforms may provide greater reach and richness for data collection than traditional methods such as snowballing and respondent-driven sampling. While our survey researcher, a former employee from the technology company, used his account to contact people and thus already had a strong network (twenty-five 1st degree connections), the reach achieved by the very first search, an endeavor that took only a few minutes and yielding over 100,000 candidates from 1st, 2nd and 3rd degrees is simply unimaginable in a non social media platform context. In addition to reach, the richness of the information is also unparalleled and practically free as compared to traditional research methods. Indeed, 1st degree profiles provide complete information about the individual’s career including companies worked at, years, positions, interests, group affiliations, skills, age, level attained, gender and sometimes email and phone numbers. Further distance profiles (2nd and 3rd degrees) may also provide a sub-set of that information based on what the individual has chosen to share on his/her public profile. Thus researchers have access to rich information that is volunteered by the candidates and that allows researcher to filter and search for candidates with greater accuracy.
Third our study contributes to a novel literature on emergent practices in using social media networks for data collection. The scarcity of studies admitting to using these new sources may be attributed to several factors that should be investigated further. On the one hand, the newness of the platform combined with the long lead times required for publishing in top management journals may explain such paucity. On the other hand, the ethical requirements for these novel methods in social sciences are at best embryonic in the fields of Sociology and Psychology, and at worse seem entirely absent in the field of Management. Our intuition is that these sources are widely used but not yet reported in any systematic fashion. While this presents an opportunity it also suggests limitations.

**Reluctance to Using Social Media Networks as a Data Collection Tool**

While several types of non-random errors (Hufnagel and Conca 1994) (i.e., biases) can be assumed to be present both in traditional method survey and SMNs surveys (question wording, order of items), the most important limitation in this study has to do with method bias (King et al. 2007) specific to the use of SMNs. Our paper does not address this issue specifically as our primary interest is to find how to facilitate subject identification and participation. We discuss biases in the hypotheses section however further research remains in systematically identifying and testing method biases with the use of SMNs.

While Masters and PhD students have limited time and resources to collect data, and are avid users of web 2.0 and social media, they are often resistant to the idea of collecting data through web 2.0 tools, which might be explained by their advisors preaching extreme caution (Rich 2011). Indeed, other fields than Management have identified limitations to the use of these tools. For instance, the American Psychological Association (Phillips 2011) highlights the following issues: privacy and confidentiality (even if the information provided is public, some institutional review boards will ask researchers to make sure that participants provide explicit permission to collect information), data accessibility (the terms of service of a given social media platform, such as Facebook, may prohibit automated data collection, thus imposing for example that the research asks participants to copy their own data and send it to the researcher), demographics (social media users may not be representative of a larger group, but this will highly differ depending on the topic under study and the platform used).

Recommendations that are given to the researchers include: keep professional and personal networks separates, do not accept ‘friends’ invitations from participants, do not share personal information with participants, do not use the data collected for other purposes than originally indicated. Another important limitation to using SMNs to collect data has to do with potential new biases inherent to their use. For instance, rating biases (Burton-Jones 2009) require care when accepting first-degree respondents that may intentionally or not provide a faithful answer by virtue of their network proximity to the researcher. While there are caveats to using social media platforms in data collection activities, it seems that the advantages of using these new web-based tools outweigh the disadvantages (Birnbaum 2004), as long as the researchers follow basic rules. In particular, we feel that there is a distinct advantage in using these novel methods for surveys that primarily seek explanation and description, that requires a more stringent criterion on both the representativeness of the sample frame and the sample, as opposed to exploration type studies (Pinsonneault and Kraemer 1993). Indeed, while the use of SMNs in research causes a blurring in boundaries regarding privacy, the gains cannot be ignored for the field.

**Final Words**

We feel our study contributes to a field in great need of empirical validation and also offers a starting point for a wider discussion on issues relating to this emerging practice. With this paper we call for more transparency in the use of social media in collecting research data and we encourage researchers in the field to engage with a topic that has to date resisted treatment from what is considered a field dominated by traditional methods.

---

References


