
Title	Methodological issues in using data from social networking sites
Author(s)	Ho Keat Leng
Source	<i>Cyberpsychology Behavior and Social Networking</i> , 16(9), 686-689. doi: 10.1089/cyber.2012.0355
Published by	Mary Ann Liebert, Inc.

This document may be used for private study or research purpose only. This document or any part of it may not be duplicated and/or distributed without permission of the copyright owner.

The Singapore Copyright Act applies to the use of this document.

This is the author's accepted manuscript (post-print) of a work that was accepted for publication in the following source:

Leng, H. K. (2013). Methodological issues in using data from social networking sites. *Cyberpsychology Behavior and Social Networking*, 16(9), 686-689. doi: 10.1089/cyber.2012.0355

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document.

Final publication is available from Mary Ann Liebert, Inc., publishers <http://dx.doi.org/10.1089/cyber.2012.0355>

© 2013 Mary Ann Liebert, Inc.

Abstract

Social network sites have become increasingly popular in recent years. With the amount of data available on social network sites, the potential exists for researchers to use this data for their research. However, like any research method, there are limitations in using data from social network sites. Firstly, as members of social network sites are not representative of the population, there is the limitation in generalizing the findings to the population. Secondly, in social network sites with a low level of activity, there is also the issue of whether the data is sufficient for analysis. Thirdly, the validity of the postings by members of social network sites should be considered as members of social network sites may not be truthful in their responses. In addition, as the environment for social network site favors a quick emotive response as opposed to a cognitive response, the review suggests that the researcher will need to be aware of possibly different behavior when members of a social network site are faced with a high involvement decision. This paper concludes that while there is potential in analyzing data from social network sites, researchers should be aware of the limitations in using this data.

Keywords: Social network site, social media, research method, data validity

Methodological Issues in Using Data from Social Network Sites

boyd & Ellison¹ define social network sites as web-based services that allow individuals to construct a public profile within a bounded system and articulate a list of other users with whom they share a connection. People join social network sites to stay in touch with their friends and maintain existing social relations. As the members know each other, there is a high level of trust and regular communication between members of social network sites^{2, 3}.

The growing popularity of social network sites has attracted the interest of researchers. With an increasing number of members, social network sites allow researchers to access a large number of people efficiently. Researchers can invite members of social network sites to participate in their research or even distribute survey instruments via the social network sites.

In particular, social network sites allow researchers to access segments of the population which are difficult to reach or are empirically underrepresented. For example, homosexuals form a small proportion of the general population and to reach this group using random sampling is generally difficult. As such, research with empirically underrepresented groups has largely used a snowballing sampling design where respondents refer the researcher to other prospective participants⁴. This process of obtaining access to prospective participants can be made more efficient through social network sites. Members of social network sites share similar interests and are known to each other. By selecting social network sites where members share specific characteristics, a researcher can quickly gain access to prospective participants.

Other researchers see the regular communication between members of social network sites as a rich source of naturalistic behavioral data that lends itself well to cyber-ethnographic research^{1, 4}. Similar to research using diary entries as a means of collecting data, entries on social network sites can potentially offer deep insights as there is a regular stream of undirected data. This allows for temporal analysis of the data by examining how members of social network sites construct

meaning over time and how attitudes are changed through interaction with other members⁵⁻⁷. As entries in social network sites are entered close to the respondent's experience, it is also subject to less distortion and hence, regarded as more valid^{8,9}.

However, the use of such data from social network sites for research is not without its criticisms. As a medium leveraging on the internet, social network site research shares some of the benefits associated with internet-based research including lower research costs and greater access to respondents¹⁰⁻¹². At the same time, some of the limitations with internet-based research also apply to research conducted on social network sites. These include the criticism of whether findings from internet users can be generalized to non-internet users, the validity of responses due to anonymity and low accountability of internet users, and the potential for unethical research practices¹⁰.

The aim of this paper is to review the methodological issues in using data from social network sites for research and to address the criticisms in using such data for research purposes. Both the issue of generalizability of findings and the validity of responses as limitations will be discussed in this paper.

Although it is not within the scope of this paper to discuss the ethical issues in using data from social network sites, researchers should be aware that conducting research on social network sites can be potentially unethical. Online postings are generally considered information available in the public domain. As such, there is no need to obtain informed consent in using the data according to existing standards of most institutional ethics review boards. It is also debatable whether it is practically feasible to obtain informed consent from members of social networks sites in the first place. However, as information from social network sites can be traced back to the individual, such information can be perceived to be personal. Hence, the use of information from social network sites without the informed consent of individuals can be considered a violation of privacy^{10, 12, 13}.

This is not an issue that can be resolved easily. Researchers should review their research practice for potential violation of ethical norms when using data from social network sites. Moreno, Fost and Christakis¹³ has suggested an ethical framework which may be useful in guiding researchers in this aspect.

Generalizability of research findings

One of the main criticisms of research conducted on social network sites is whether the findings can be generalizable to the general population. At issue is whether there is any difference between members and non-members of social network sites.

Some studies have suggested that there are demographic differences between members and non-members of social network sites. Members of social network sites tend to be younger than the general population^{4, 10, 14-17}. As such, they are not representative of the population and the findings cannot be generalized to the larger population¹⁰. For example, a politician who is popular on social network sites may not see his popularity translate into a victory at the polls as the online community is made up of the younger segment of the population, some of which may not be even eligible to vote¹⁸.

Another difference between members and non-members of social network sites lies in members being more likely to have access to internet and are more comfortable with technology^{15, 19, 20}. With an increasing number of people with access to internet, it can be argued that this difference can be eroded with time. However, in countries where access to the internet and social network sites is limited to segments of the population with higher income or specific geographic locations, research findings are necessarily still limited in generalizability.

Studies on psychographic differences between members and non-members of social network sites have not been conclusive. A study found that members of social network sites have

a higher need to belong to social groups¹⁵. However, this has not been replicated in other studies and a recent review of the literature suggests that there is insufficient evidence to conclude that there are differences in personality traits between the two groups¹⁹.

Low levels of activity

One of the attractions in analyzing data from social network sites is the availability of a high level of regular communication between members. However, especially on social network sites of commercial companies, the level of activity may be too low to allow for meaningful analysis. More importantly, when there is an absence of a large group of people actively communicating with one another in the account, the views cannot be representative².

Some commercial companies see social network sites as an extension of their traditional marketing communication tool and use social network sites to advertise their products. As such, the social network site for such companies is essentially a profile of the company, similar to a corporate web page. Social network sites of this type do not encourage a conversation between members and more importantly, fail to engage consumers to determine the cultural meaning of the brand^{19, 21}. The conversation is dominated by the corporation and as such, the views on such social network sites cannot be representative.

Corporations who see their social network site as part of their traditional one-directional marketing tool will also be more likely to be uncomfortable with the interactive element in social network sites. This is especially so when negative comments on their products or company are posted on their social network site. As more attention is given to negative reviews made online, such corporations may delete negative comments on their social network sites²². While there have been few studies to examine the extent of this practice, it suggests that researchers should be aware of this possibility when analysing data on social network sites given that it can affect

the validity of research findings. Hence, it is important to consider this issue especially when the level of activity is low and the conversation is dominated by the corporation.

It is not only the number of entries that matter, but also the quality of the posting on the social network site. Where members are unable to express an experience to sufficient depth, it will be difficult to provide a meaningful analysis of the data, especially when social network site entries tend to be short⁶. The value of social network site data lies in analyzing the postings over a period of time. This temporal analysis allows researchers to form a theory as to how attitudes or behavior may change with time or in response to other members in the social network. Hence, when members are unable to express themselves clearly over a period of time, researchers will not be able to analyze how consumers make cultural meaning of a brand or issue from the data on social network sites.

Validity of Postings

As members of social network sites know each other, there is a high level of trust between members, and the information shared is likely to be accurate and truthful. To a researcher, this data becomes valid as it is a reflection of the true experience of the members in the social network site. However, there exists the potential for respondents to be less than honest about socially undesirable behaviors as there is pressure to conform to group expectations and norms^{5, 10, 23}. Members of social network sites will be less willing to be honest about behaviors that may embarrass them.

The literature suggests that when a group is formed due to attachment between members, the group is defined as a common bond group. Social network sites with members who are friends of each other can be considered as common bond groups. In contrast, when a group is formed due to the attachment to the group as a whole, this group is defined as a common identity group. Such social network sites include those formed by corporations and other organizations like schools, communities and non-profit organizations. This distinction is important as there is

greater pressure for members of common identity groups to adhere to group norms²⁴. Hence, researchers should be aware of the greater pressure among members of common identity groups to exhibit behaviors that are consistent with group expectations or norms.

In addition, where their security or safety is at risk, members may choose a response that is less than truthful when they can be identified. For example, an incumbent politician or a politician who is expected to have a high probability of winning in the elections may receive more positive feedback on social network sites. This is because the members of the social network sites may choose behaviors to avoid repercussions if the politician does get voted in²⁵.

While most social network sites encourage users to construct accurate representations of themselves, members of social network sites do this to varying degrees¹. As such, in analyzing data from social network sites where most members do not know each other or do not have a high level of trust, there is some anonymity provided to members. This is especially so on corporate social network sites where the number of members tends to be large. In such cases, using the internet as a medium can encourage respondents to be more expressive of their thoughts and feelings because it does not involve face-to-face communications²⁶.

Also, to preserve their anonymity, members of the online community may create fake accounts in order to post negative comments without repercussions. While the anonymity may encourage the responses to be more valid, the researcher will also need to be aware that such anonymity may invalidate the research findings, especially when fake accounts are used to specifically create a negative perception of the company, person or issue without basis or out of malice.

More importantly, the researcher needs to be aware that the discussions on social network sites do not necessarily translate into behaviors. This is because the nature of social network sites favors a quick emotive response to postings rather than a prolonged, conscious and cognitive

response. Members of social network sites can click on a “like” button or react to a posting with quick comments without much thought. However, when faced with the consumption opportunity, especially in high involvement decisions, the rational or cognitive process takes precedence and may result in a different behavior.

For example, social network sites facilitate affective political alliance as they allow the public to easily contribute comments and engage with the candidate. However, this affective political alliance may not translate into votes as voting behavior remains a rational choice that require a deeper level of information processing than the requirements of clicking a “like” button on Facebook^{18, 27, 28}.

Conclusion

With the popularity of social network sites, the volume of data on social network sites on how people form opinions and intended behaviors on products, companies and issues has increased over the last few years. This presents an opportunity for researchers to mine the data and analyze it. However, as reviewed in this paper, the researcher needs to be aware that analyzing data from social network site is not without its limitations.

Firstly, as the members on social network sites are self-selected, they may not be representative of the population. Those who are members of social network sites tend to be younger and more at ease with Internet technology. Other variables may also work towards the bias in encouraging certain segments of the population in being members of social network sites. As such, the researcher will need to be aware of the potential bias in findings resulting from the analysis of data from social network sites.

Secondly, a researcher should determine if there is sufficient data in the social network site for analysis. Besides reviewing the quantity of postings, the researcher should also review

the number of people posting on the site and the proportion of the posting by each individual to ensure representativeness of the responses.

Thirdly, the researcher should consider whether there is a risk for members not to be truthful with their responses and hence invalidate the findings. The risks will be higher where socially undesirable behaviors are investigated amongst members who know each other or where there is motivation to post untruthful responses. In addition, noting that the environment for social network site favors a quick emotive response as opposed to a cognitive response, the researcher will need to be aware of a possibly different behavior when members of a social network site are faced with a high involvement decision.

The review suggests that while there is potential in analyzing data from social network sites, the researcher will need to be aware of the limitations. Perhaps it is not so much as to which method is superior but rather the researcher needs to be aware of the strengths and weaknesses in the different methods^{9, 11, 12}.

References

1. boyd, d.m. and N.B. Ellison, *Social Network Sites: Definition, History, and Scholarship*. Journal of Computer-Mediated Communication, 2008. 13: p. 210-230.
2. Ridings, C.M., D. Gefen, and B. Arinze, *Some antecedents and effects of trust in virtual communities*. The Journal of Strategic Information Systems, 2002. 11(3-4): p. 271-295.
3. Ridings, C.M. and D. Gefen, *Virtual community attraction: Why people hang out online*. Journal of Computer-Mediated Communication, 2004. 10(1).
4. Mathy, R.M., M. Schillace, S.M. Coleman, and B.E. Berquist, *Methodological rigor with Internet samples: New ways to reach underrepresented populations*. CyberPsychology & Behavior, 2002. 5(3): p. 253-266.
5. Leigh, B.C., *Alcohol consumption and sexual activity as reported with a diary technique*. Journal of Abnormal Psychology, 1993. 102(3): p. 490.
6. Zillinger, M., *Germans' tourist behaviour in Sweden*. Tourism, 2008. 56(2): p. 143-158.
7. Golder, S.A., D.M. Wilkinson, and B.A. Huberman, *Rhythms of social interaction: Messaging within a massive online network*. Communities and Technologies 2007, 2007: p. 41-66.
8. Alaszewski, A., *Using diaries for social research* 2006: Sage Publications Ltd.
9. Broderick, J.E., *Electronic diaries: Appraisal and current status*. Pharmaceutical medicine, 2008. 22(2): p. 69.
10. Skitka, L.J. and E.G. Sargis, *The Internet as psychological laboratory*. Annu. Rev. Psychol., 2006. 57: p. 529-555.

11. Riva, G., T. Teruzzi, and L. Anolli, *The use of the internet in psychological research: comparison of online and offline questionnaires*. *CyberPsychology & Behavior*, 2003. 6(1): p. 73-80.
12. Bainbridge, W.S., *The Scientific Research Potential of Virtual Worlds*. *Science*, 2007. 317(5837): p. 472-476.
13. Moreno, M.A., N.C. Fost, and D.A. Christakis, *Research ethics in the MySpace era*. *Pediatrics*, 2008. 121(1): p. 157-161.
14. Cha, J., *Shopping on social networking Web sites: Attitudes toward real versus virtual items*. *Journal of Interactive Advertising*, 2009. 10(1): p. 77-93.
15. Gangadharbatla, H., *Facebook Me: Collective Self-Esteem, Need to Belong, and Internet Self-Efficacy as Predictors of the iGeneration's Attitudes toward Social Networking Sites*. *Journal of Interactive Advertising*, 2008. 8(2): p. 1-28.
16. Kelly, L., G. Kerr, and J. Drennan, *Avoidance of Advertising in Social Networking Sites: The Teenage Perspective*. *Journal of Interactive Advertising*, 2010. 10(2): p. 16-27.
17. Peluchette, J. and K. Karl, *Social networking profiles: An examination of student attitudes regarding use and appropriateness of content*. *CyberPsychology & Behavior*, 2008. 11(1): p. 95-97.
18. Erikson, E., *Hillary is my friend!: MySpace and political fandom*. *Rocky Mountain Communication Review*, 2008. 4(2): p. 3-16.
19. Anderson, B., P. Fagan, T. Woodnutt, and T. Chamorro-Premuzic, *Facebook psychology: Popular questions answered by research*. *Psychology of Popular Media Culture*, 2012. 1(1): p. 23.

20. Hargittai, E., *Whose space? Differences among users and non-users of social network sites*. *Journal of Computer-Mediated Communication*, 2008. 13(1): p. 276-297.
21. Deighton, J. and L. Kornfeld, *Interactivity's Unanticipated Consequences for Marketers and Marketing*. *Journal of Interactive Marketing*, 2009. 23(1): p. 4-10.
22. Sen, S. and D. Lerman, *Why are you telling me this? An examination into negative consumer reviews on the Web*. *Journal of Interactive Marketing*, 2007. 21(4): p. 76-94.
23. Klein, O., R. Spears, and S. Reicher, *Social identity performance: Extending the strategic side of SIDE*. *Personality and Social Psychology Review*, 2007. 11(1): p. 28-45.
24. Sassenberg, K., *Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats*. *Group Dynamics: Theory, Research, and Practice*, 2002. 6(1): p. 27.
25. Dalsgaard, S., *Facework on Facebook: The Presentation of Self in Virtual Life and its Role in the US Elections*. *Anthropology Today*, 2008. 24(6): p. 8-12.
26. Esterberg, K.G., *Qualitative methods in social research 2002*: McGraw-Hill Boston.
27. Vitak, J., P. Zube, A. Smock, C.T. Carr, N. Ellison, and C. Lampe, *It's Complicated: Facebook Users' Political Participation in the 2008 Election*. *Cyberpsychology, Behavior, and Social Networking*, 2011. 14(3): p. 107-114.
28. Sweetser, K.D. and R.W. Lariscy, *Candidates Make Good Friends: An Analysis of Candidates' Uses of Facebook*. *International Journal of Strategic Communication*, 2008. 2(3): p. 175-198.