

Description

The Mixed Blessing of Social Media Data

The phenomenal growth in utilization rates of popular social networking sites such as Facebook, Twitter, and Instagram, offer academic researchers an absolute treasure trove of consumer-generated content for analysis. However, access to that data isn't without challenges—all providers have unique policies and procedures on data collection, sharing and publication.

The Potential of Social Media Data

Assuming the challenges to access data can be overcome, social media data offers several advantages over [data gathered by more traditional research methods](#):

- Topical information that is potentially available in real time
- Information that is unfiltered by survey tools and other potential biases
- Naturally occurring data collected from an actively engaged population
- Clearly identified, albeit narrow user demographics

Ethical Challenges

Beyond the default privacy settings established by the respective social networking sites (which offer varied customization options to users), critics of social media data mining raise concerns about the issue of consent. Does the fact that the posts are made to public sites with the potential for re-posting constitute a de facto grant of consent? Should consent be sought from each user? Is that even feasible when software programs are scraping tens of thousand of posts per minute?

Is there some middle ground where the anonymity of your username can be seen to allay fears about identity protection? Is your Twitter user name sufficiently anonymous, or should that name be kept anonymous too?

Facebook has recently begun running television commercials advising users not to post vacation photos since their declaration of absence from home is an open invitation to burglars. Given such an apparently cavalier attitude to privacy, should the protections of identity for academic research be adjusted accordingly?

Research Challenges

Given the relatively recent prominence of social networking (Facebook was founded as far back as 2004, Twitter in 2006, and Instagram is a 2010 venture), researchers are still struggling with the extent to which [traditional research methodologies](#) should be applied to this data:

- There is increasing evidence of propaganda as each venture has accepted advertising to limit

their cash burn and generate some profits for investors, the arrival of new posts, tweets, and photos that the respective company “thought you might find interesting” has opened the floodgates for propaganda.

- The veracity of the information being posted and the identities of the posters continue to be questionable. In December 2014, Instagram purged all the suspected fake and spam accounts from its site. After a loss of 18.8 million accounts (nearly 30% of its total membership), the event became known as the “Instagram Rapture.”
- Perceived anonymity on the Internet has been shown to promote exaggeration and falsehood. Advocates might argue that this is no different than the hyperbole of survey responses where enthusiastic participants give you the answers they think you want, but the sheer volume of social media data has a greater potential to skew the results.

Too Rich to Ignore?

While ethical questions and practical challenges remain, the research potential held within this data is too rich to ignore. It may require the development of new rigorous research methodologies and some adjustment on the part of traditional researchers, but the data will find its way into broader academic research strategies eventually.

Consider that much of the astronomical market valuations of these companies (over \$220 billion for Facebook, \$35 billion for Instagram, and over \$23 billion for [Twitter](#)) are tied-up in the market research potential of their users. As advertisers demand better access to those users, academic researchers should be able to benefit from those improvements.

Category

1. Promoting Research
2. Using Social Networks

Date Created

2015/10/29

Author

editor