Identity and Behavior in Online Health Communities

Cheng Guo
Clemson University
Clemson, SC 29630, USA
chengg@clemson.edu

Abstract

Online communities and forums are becoming an important source of health information. People share and discuss health information at all levels of topic using different identities ranging from anonymity to real names. The overall amount of user engagement and the quality of that engagement may differ depending on the different levels of identity. Users may also perform differently using different identities. Yet little is known how different levels of identity affect user behavior in online health communities. My dissertation research focuses on understanding the relationship between identities and Q&A behavior in online health communities. My work also focuses on designing processes and tools to better support online health Q&A behavior. My dissertation research will inform the building of safer and more useful online health communities.

Author Keywords

HCI; health informatics; online communities; identity; anonymity; health information

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction

People are increasingly seeking answers to questions about health via non-traditional channels [15]. Whereas many people used to rely primarily on their healthcare-provider for answers to heath questions, people now regularly seek information online. One in three U.S. adults have been looking for answers to health questions online [5]. Using the Internet, people consult health information posted to websites, blogs, online reviews, social network sites and social Q&A sites [6]. Compared to asking questions via a search engine, Q&A sites provide unique benefits for users such as social fulfillment [9], expert opinions [14] and an increase in perceived speed and answer quality [11]. However, similar to other ways people may learn about health information online, the quality of information on social Q&A sites varies from excellent to poor [1].

One factor that may be related to Q&As quality on Q&A sites is the identity policy of the site. Different Q&A sites have different identity policies ranging from real name, where a contributor's name (and subsequently the contributor's identity) is associated with his or her question or answer, to anonymity, where no identifying information such as name is given and therefore no identity is associated with specific questions or answers. For example, Zhihu (https://zhihu.com), a Chinese Q&A site, encourages users to register using their real names, but this is not mandatory. As a result, some users register using real names while others registered using pseudonyms. On the other hand, Quora (https://quora.com), has a very strict "real name" policy [12] where they force all users to register using their real names. Another popular Q&A site - Yahoo Answers (https://answers.yahoo.com) has a pseudonym policy that is even less strict than Zhihu [16]. Yahoo asks users to provide their Yahoo ID or a nickname of their preference. Identity presents implications for firms seeking to develop their

sites and engage users. One of the most important implication is privacy. Zhihu, Quora and Yahoo all provide a privacy feature for users to ask or answer questions anonymously. As a result, on these social Q&A sites, there are three different levels of identity: real name, pseudonymity and anonymity. In social Q&A sites, I believe that pseudonymity is a different form of identity than anonymity because social Q&A sites are community driven. On these sites, each user has a profile page which display basic user information and all the questions and answers the user posted before. Even the user can make these information private, given the user name, other users could still retrieve the user in other questions and answers. Being anonymous online can bring certain benefits such as making users feel relaxed, feeling free to express views and helping control personal information disclosure [8]. Anonymity can also bring negative effects such as undermining personal contribution in group communications [13]. Being identified online also bring positive effects (e.g., less offensive words [7]) and negative effects (e.g., discrimination [4]). Prior work suggests that privacy of sensitive information such as health information is a big concern for individuals [2, 10]. However, there is little research that guides health Q&A sites and other online health communities about identity policy. Users also need implications to help them make decisions on online identity disclosure. In my dissertation, I seek to deepen our understanding of how different levels of identity affect a set of Q&A metrics and user behavior such as user engagement [7], Q&A quality [1] and trolling behavior [3] on health Q&A sites.

Research Agenda

To address my research goal, my dissertation will combine both qualitative and quantitative research methods in a series of studies: *Study 1* will be a quantitative analysis of identity and content sensitivity on three Q&A sites (Yahoo

Answers, Quora and Zhihu); *Study 2* will be a controlled lab experiment examining and influential factors ; *Study 3* will focus on exploring design approaches on Q&A sites to elicit more and higher quality questions and answers.

Study 1
Yahoo, Zhihu and Quora have different identity policies. Yahoo asks users to register using any nickname they desire. Zhihu recommends users to use real names but not mandatory. Quora has a strict real name policy. All three sites have in common a privacy feature allowing users to be anonymous when ask or answer questions. I will leverage these differences in identity policy to examine user engagement, answer quality and trolling behavior. I will collect questions and answers manually under both health and non-health topics. Then, these Q&As will be parsed via a Python script. A set of Q&A metrics will be analyzed such as ask identity, answerer identity, number of answers per question, number of question comments, number of answer comments, number of upvotes and downvotes per answer. Study 1 is expected to explore in a naturalistic environment, what is the relationship between different levels of identity and Q&A behavior. User engagement will be quantified by those Q&A metrics listed above. Answer quality and trolling behavior will be measured by human raters. I have already collected 300 health and non-health Q&As on Yahoo and found that anonymous answers have more upvotes than pesudonymous answers but the number of downvotes is similar. These findings provide fundamental support for my dissertation work to further investigate identities and other metrics on other Q&A sites.

Study 2
Study 2 aims to reveal the underlying mechanisms behind relationships observed in study 1 through conducting controlled lab experiment and studies of various influential fac-tors (e.g., user characteristics, different site design features, etc.) in isolation. I plan to design experiments to manipulate asker and answer identity (not only name, but also other social cues such as title, occupation) and measure users' ratings on the quality of the Q&As I collected in Study 1. Study 2 will be conducted in the summer and fall of 2018.

Study 3
Study 2 should provide design guidelines and implications about identity for health Q&A sites. In study 3, I will design and conduct experimental field studies on real Q&A sites to assess the effectiveness of those approaches. I plan write browser extensions to manipulate the content on the page. Study 3 will be conducted in the spring of 2019.

Expected Contributions
My dissertation contributes to the field of pervasive health privacy, social computing and health informatics by extending our knowledge about online identities and how those relate to quality and quantity of health information available via health Q&A communities. My work will benefit both online health community administers and users with practical guidelines to design sustainable online health communities which are more safe and more useful.

Acknowledgments

REFERENCES
1. Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. 2008. Finding high-quality content in social media. In *Proceedings of*

the 2008 international conference on web search and data mining. ACM, 183–194.

2. Gaurav Bansal, David Gefen, and others. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision support systems* 49, 2 (2010), 138–150.

3. Sally Jo Cunningham and Annika Hinze. 2014. Social, religious information behavior: An analysis of Yahoo! Answers queries about belief. *Advances in the Study of Information and Religion* 4, 1 (2014), 3.

4. Jennifer L Doleac and Luke CD Stein. 2013. The visible hand: Race and online market outcomes. *The Economic Journal* 123, 572 (2013).

5. Susannah Fox and Maeve Duggan. 2013. Health online 2013. *Washington, DC: Pew Internet & American Life Project* (2013).

6. Susannah Fox and S Jones. 2010. The social life of health information: AmericansâĂŹ pursuit of health takes place within a widening network of both online and offline sources. 2009. *Washington, DC: Pew Internet & American Life Project Google Scholar* (2010).

7. Rolf Fredheim, Alfred Moore, and John Naughton. 2015. Anonymity and Online Commenting: The Broken Windows Effect and the End of Drive-by Commenting. In *Proceedings of the ACM Web Science Conference*. ACM, 11.

8. Ruogu Kang, Stephanie Brown, and Sara Kiesler. 2013. Why do people seek anonymity on the internet?: informing policy and design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2657–2666.

9. Soojung Kim, Jung Sun Oh, and Sanghee Oh. 2007. Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective. *Proceedings of the American Society for Information Science and Technology* 44, 1 (2007), 1–15.

10. Daniel Masys, Dixie Baker, Amy Butros, and Kevin E Cowles. 2002. Giving patients access to their medical records via the Internet. *Journal of the American Medical Informatics Association* 9, 2 (2002), 181–191.

11. Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1739–1748.

12. Quora. 2016. Terms of Service. (Jul. 2016). `https://www.quora.com/about/tos`

13. Stephen A Rains. 2007. The impact of anonymity on perceptions of source credibility and influence in computer-mediated group communication: A test of two competing hypotheses. *Communication Research* 34, 1 (2007), 100–125.

14. Chirag Shah, Vanessa Kitzie, and Erik Choi. 2014. Modalities, motivations, and materials–investigating traditional and social online Q&A services. *Journal of Information Science* 40, 5 (2014), 669–687.

15. C Lee Ventola. 2014. Social media and health care professionals: benefits, risks, and best practices. *Pharmacy and Therapeutics* 39, 7 (2014), 491.

16. Yahoo. n.d. Yahoo Answers Community Guidelines. (n.d.). `https://answers.yahoo.com/info/community_guidelines`