This is a peer-reviewed, accepted author manuscript of the following article: Nicol, E., Htait, A., Azzopardi, L., & Moncur, W. (Accepted/In press). Towards identifying, understanding and controlling cumulative revelations in social media. Poster session presented at 84th Annual Meeting of the Association for Information Science & Technology (ASIS&T), Salt Lake City, United States.

Towards Identifying, Understanding and Controlling Cumulative Revelations in Social Media

Emma Nicol
University of Strathclyde, UK
emma.nicol@strath.ac.uk

Amal Htait
University of Strathclyde, UK
amal.htait@strath.ac.uk

Leif Azzopardi
University of Strathclyde, UK
leif.azzopardi@strath.ac.uk

Wendy Moncur
University of Strathclyde, UK
wendy.moncur@strath.ac.uk

ABSTRACT

Every day, people post information about themselves and others on online social networks, making such information accessible within their social circles but also, potentially, way beyond. While the information posted may seem benign or innocuous, small pieces of information, when tied together, can potentially reveal much more about the person than intended. Such cumulative revelations could expose them to risks such as identity theft, fraud or loss of employment. This paper describes findings from interviews about people's online interactions, focusing on their requirements and desires for improved ways to identify, understand and control cumulative revelations arising from their social media profiles that could put them at risk of grave consequences. These findings motivate our future work on how to better raise awareness among social media users of the risks and consequences to which combinations of posts may lead.

KEYWORDS

Social Media, Online Social Networks, Personal Information Sharing, Cumulative Revelations

INTRODUCTION

Online Social Networks (OSNs) allow people to construct personalised profiles, post content and build connections with others. Such actions enable people to introduce themselves online while expressing their personality, thoughts, feelings, and sharing personal information (e.g., birthday, location, school, etc.) (Krasnova et al., 2010; Xia et al., 2013; Haimson et al., 2016). While small pieces of information shared online by or about a person may seem harmless, over time they may reveal cumulatively more than the person intends, as identifiable traces can be linked then exploited. Such linking leads to "cumulative revelations", for instance, geo-tagged posts could reveal a home address, while posts from the airport about a trip away, could give sufficient clues to thieves to target a home. Meanwhile, a social media user may lament online about living alone, while sharing check-ins at cafes and running routes. Taken together, a stalker may infer that the person is vulnerable and follow or otherwise target them. Further, the combination of many posts may reveal aspects of a person's personality, preferences (e.g. politics) and even their mental health, such traits being increasingly detectable via machine learning tools. When coupled with machine learning capabilities, the abundance of information online about social media users constitutes a major privacy exposure risk and could, potentially, have serious negative consequences for an individual e.g., via identity theft (Acquisti and Gross, 2009), financial loss, damage to reputation (Chen et al., 2016), cyberattacks or reputational damage for their employer. There could even be risks to national security (Dressler et al., 2015). Previous studies have highlighted risks and consequences of revealing information online, and e.g. the trade-offs of doing so e.g. (Min et al., 2014; Vishwanath et al., 2018) but little research has investigated social media users' understanding of risks and consequences from combinations of traces, and how people conceive of, and control social media usage given such risks. This work contributes to our understanding of these phenomena and outlines some requirements for tools to manage such risks.

OVERVIEW OF STUDY

To investigate cumulative revelations in social media, we conducted semi-structured interviews to inquire about information sharing habits, social media use and digital traces, with additional exploration of effects of the pandemic lockdown and homeworking. We asked participants about the private and public aspects of online information sharing, explored the personal and professional contexts of information sharing, and their experiences of cumulative revelation, whether direct or observed. We interviewed 26 people aged 20-59 years, (13m, 12f, 1nb), in employment in the UK, May-July 2020. Interviews were conducted via videoconferencing, each lasting 60-90 minutes. Participants received a shopping voucher (approx. $30 USD). Interviews were audio recorded, transcribed, then coded and analysed

84th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 3, 2021 | Salt Lake City, UT.

thematically (Braun and Clarke, 2006), using NVivo software. Analysis drew out similarities and variations between participants' practices, with coding reducing the data into themes. We summarise key findings organised as follows:

(i) concerns and fears about social media usage
(ii) control over social media streams
(iii) awareness of risks and consequences of using social media
(iv) reflections on risks of own usage of social media

Concerns: Fears about Social Media Usage

Participants feared that very detailed information about them (e.g. from location tracking) could be shared without permission by (or be stolen from) "big tech firms". They feared this information could be used to target them with advertising or content in a way they could not understand or control. An additional concern was that information collected about them (e.g., date of birth, address, credit score, images) could be used to cause them harm (e.g. via identity or financial fraud). For those <25 years old, there were concerns that childhood online activities would be visible to new peers or potential employers, causing embarrassment or compromising professionality. For participants from or wishing to visit authoritarian countries, there were worries that governments could collect their information and use it against them e.g. to deny visas or right to remain, or to sanction them for inciting political unrest etc.

Control: Over Social Media

Participants had desires regarding social media and controls they wished were available. Specifically, the ability to easily:

- Correct false online information about themselves
- Remove/delete online shared information permanently
- Prevent data aggregation of their information across sites and usage
- Filter the audience for their shared information

Participants also stipulated that social media sites should present to users (in an easy to understand and usable manner) the information held about them. In addition, any tools developed to help manage their online postings should also assure data protection i.e. would not collect and store their information in the manner of many social media companies.

Awareness: Understanding the Risks and Consequences of Using Social Media

Given that social media users often do not know the consequences of sharing certain information, participants suggested that training or education should be available to help raise awareness about the potential for cumulative revelation from social media usage by: (i) giving advice on how to protect one's personal information, and (ii) showing how to avoid revealing details which can be combined together, increasing the risks of harmful consequences.

Reflections: Identifying Risks in One's Social Media Profiles

Most participants did not have a clear sense of the image they projected to others online or how it might be interpreted. Increased and novel use of social media due to the lockdown and homeworking meant that for many it had become even more of a challenge keeping track of their digital traces. They desired to see more clearly the picture they present online via visualisation of their online information where they could: identify and flag false, misleading or outdated information about themselves, remove, delete or hide any past shared online information and be able to compare how much information they shared about themselves relative to other social media users. Many participants indicated the importance of simplicity and ease of use of a tool that would help them control, identify and understand their information, emphasising that it should be usable by people at particular risk online e.g. older people and children.

SUMMARY AND FUTURE WORK

This paper presented themes emerging from a qualitative study of cumulative revelations in personal information shared online. Participants expressed fears about their online activity, discussed the types of information they tended to share, and outlined how their desires and needs for safer online activity might translate to requirements for tools to manage this. Our next steps will be to further investigate these fears, before developing appropriate technological and other interventions to mitigate risky social media usage that can lead to such revelations and associated consequences.

ACKNOWLEDGMENTS

This work was supported by grant EP/R033889/1 and was subject to approval by the ethics committee of DJCAD, University of Dundee. We acknowledge the input and effort of our interview participants.

REFERENCES

Acquisti, A., & Gross, R. (2009). Predicting Social Security numbers from public data. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(27), 10975–10980.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Chen, H., Beaudoin, C. E., & Hong, T. (2016). Protecting oneself online: The effects of negative privacy experiences on privacy protective behaviors. *Journalism and Mass Communication Quarterly*, *93*(2), 409–429.

Dressler, J. C., Bronk, C., & Wallach, D. S. (2015). Exploiting military OpSec through open-source vulnerabilities. In *Proceedings - IEEE Military Communications Conference MILCOM* (Vol. 2015-December).

Haimson, O. L., Brubaker, J. R., Dombrowski, L., & Hayes, G. R. (2016). Digital footprints and changing networks during online identity transitions. *Conference on Human Factors in Computing Systems - Proceedings*, 2895–2907.

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online Social Networks: Why We Disclose Author Version. *Journal of Information Technology*, *25*(2), 109–125.

Min, J. and Kim, B. (2015), How Are People Enticed to Disclose Personal Information Despite the Privacy Concerns in Social Network Sites? The Calculus Between Benefit and Cost. J Assn Inf Sci Tec, 66: 839-857.

Vishwanath, A., Xu, W. and Ngoh, Z. (2018), How people protect their privacy on facebook: A cost-benefit view. Journal of the Association for Information Science and Technology, 69: 700-709.

Xia, N., Song, H. H., Liao, Y., Iliofotou, M., Nucci, A., Zhang, Z. L., & Kuzmanovic, A. (2013). Mosaic: Quantifying privacy leakage in mobile networks. *Computer Communication Review*, *43*(4), 279–290.