Exploring Social Media Sharing: The Nature of Private Information Shared by *Instagram* Users

Melanie H. Modrall

MacEwan University

Author Note

Correspondence can be addressed to Melanie H. Modrall (modrallm@mymail.macewan.ca).

Abstract

This research explores the nature of social media sharing of personally identifiable information on Instagram by conducting a content analysis on a total of ten profiles and five posts per profile, selected randomly. Main themes reflected in data were categorized by three methods of information sharing: profile, captions, and images and videos. Fourteen themes were discovered for sharing in profiles and captions, while fifteen were found for images and videos. Results showed users were most likely to share information in their profile (64.1%) and most commonly shared their gender presentation, face, and location. This correlates with previous research concerning privacy and security risks from social media sharing, indicating that these concepts need to be re-evaluated to remain relevant.

Keywords: social media, privacy, information sharing

Exploring Social Media Sharing: The Nature of Private Information Shared by *Instagram* Users

Social media is everywhere in our lives, with access readily available on our cellphones, computers, laptops, tablets, etc. People cannot stop talking about the videos and photos friends are posting – thus continuing the chain of sharing social information for entertainment. Instead of visiting with friends to tell them about an event, users of social media can upload videos or pictures with captions and hashtags to save time in showing off their experiences. While people are busy sharing aspects of their lives with others online, they do not always consider who has access to this information and how it is being used (Hogan, 2010). Privacy and security settings exist, yet this does not necessarily ensure protection from the risks involved by engaging in revealing aspects of our personal lives to known and unknown online audiences (O'Neil, 2001).

Comprehensive research has uncovered some of the concerns people have about online privacy and their willingness to share personal information (Joinson, 2001; Zemmels & Khey, 2015). There is, however, little research addressing the links between privacy concerns, the types of personal information people share on social media sites, and the actions that people and businesses take to ensure their accounts' security (O'Neil, 2001).

Social Media: Background and Definitions

According to Boyd and Ellison (2007), the concept of social networking began with the launch of *SixDegrees.com* in 1997. Although many separate programs already offered the features networking sites provided, such as instant messaging and blogging, they were slowly phased out due to the convenience of combining these aspects into one platform. Early public online communities, such as Usenet, were organized around topics or shared interests. However, social networking sites eventually shifted this focus onto the individual as the creator

of their own communities. With this unlimited potential to make and join multiple groups, it is easy to conclude that these networks incentivized users to engage socially on the Internet.

For the purposes of this content analysis, I will be using Boyd and Ellison's (2007) concept of “social networking sites” in tandem with “social media,” since these two concepts are used interchangeably. Social networking sites are defined as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site” (p. 211). The ways in which connections are made and maintained with other users vary for each social media platform, since each site has its own approach to the capabilities and features available for use online or through cellphone applications (Boyd & Ellison, 2007). Typical features found on a social media website include: personal profile, friends list, photo and video sharing, private messaging, and status-posting – often presented in a news-feed style (Lange, 2008).

A “profile” on social media is a personal place where the creator can present a social image or a “digitally mediated identity” of themselves. Access to viewing profiles can be set to public, private, or semi-public (Lange, 2008; Zemmels & Khey, 2015). Boyd and Ellison (2007) state that the main feature of social media is the act of “networking”, where creating and maintaining social relationships is important, but public displays of these connections is where emphasis is placed (Lange, 2008). Social media is also a place where learning, entertainment, and expression of oneself in front of an online audience is encouraged (Boyd & Ellison, 2007; Hogan, 2010; Zemmels & Khey, 2015). However, with the prevalence of 'opt-in' privacy settings,

such as those used on Facebook, many social media profiles are public by default and require users to manually enable these settings. This approach leads to a greater number of online users with publicly available data, often without their knowledge.

Aspects of Privacy

According to O'Neil (2001), the concept of privacy is a subjective construct in that concerns, definitions, and how it is valued differs for each person. However, there are four central themes of privacy which have been prevalent in past research. Informational or acquisition privacy is defined by an individual's desire to control the types of access to personal information that others have. The physical dimension of privacy is where physical access or possession of information is a concern. Expressive privacy protects the individual from social control over choices, lifestyles, and self-expression. The social or communicative aspect is concerned with an individual's control over social contacts and how those connections are maintained (Buchanan et. al, 2007; O'Neil, 2001).

Lange (2008) suggests that there are many ways in which users of social media engage online audiences while also maintaining a level of privacy and security, such as the use of privacy settings that restrict access to one's profile and posted content. The two types of access granted to other users are defined as “publicly private”, where personal information is available, but content visibility is restricted to a select few, or as “privately public”, when access to content is unrestricted but personal information is not shared. This suggests that information would remain either public or private until a user changes their privacy settings.

However, as the distinctions between online and offline events have become blurrier, the experiences that users have offline may become permanent due to online sharing and the

data-traces it leaves behind (Lange, 2008; Robards, 2012; Zemmels & Khey, 2015). Data-traces are any information that a user reveals about themselves or others, as well as online activities they engage in, that are retained by online third-party services, such as Facebook (Lange, 2008; Robards, 2012).

Purpose

The purpose of this content analysis is to analyze profiles and uploaded visual content (images and videos) that depict personal information on Instagram, with various criteria being considered in determining and exploring the nature and frequency in which information is presented. Instagram is a social media platform launched in October 2010 that can be used online or through a cellphone application. The ways in which users can share information on this platform include user profiles, captions, and images and videos – where any kind of information can be made available to other users. Exploring what people are comfortable sharing on social media sites provides insight into what people consider "private." This is especially helpful in improving future services, where private information is used in maintaining online security and preventing identity theft.

Research Question

The following research questions guide this study:

1. What is the nature of the personal information people are willing to share via Instagram?
2. What type of personal information is frequently shared in Instagram user profiles and posts?

Method

Design

A qualitative research paradigm along with an explorative approach was used in this content analysis to discover the nature of personally identifiable information being shared via Instagram. This design provided an opportunity to study the types of personal information that was made accessible in user profiles, captions in photo or video posts, and the ways this content within these areas were shared on Instagram.

Sample

This systematic random sampling consisted of personal information gathered from 10 user profiles and five posts from each of these profiles that were created between January 1st, 2017 and December 31st, 2017.

Sample Selection

On March 30th, 2018, the principal researcher analyzed a total of 10 user profiles and five posts from each of these profiles on Instagram, using the hashtag filter *#trending*. Due to the large quantity of posts using this filter, a systematic random sampling approach was used to select the initial post from which the corresponding profile could then be found. This same approach was then used to select five posts from each of these profiles. The principal researcher used a random number generating tool Stat Trek (https://stattrek.com) to limit the number of user profiles and posts that were available to be selected. User profiles were chosen from the first 100 listed while the posts within these profiles were selected from the 20 most recent entries within the specified timeframe (January 1st, 2017 to December 31st, 2017). Both profiles and posts were identified by counting from left-to-right, top-to-bottom. If a profile had

fewer than five posts or was identified as an advertising profile, a new number was generated to select a different profile.

Inclusion and Exclusion Criteria

Data was collected using information in posts made from January 1st, 2017 to December 31st, 2017, using the filter *#trending* to ensure that only active accounts were included in the study. Profiles that did not have posts within this timeframe were not used. The posts listed under the subheading "Most Recent" were used to find profiles and subsequent posts from the same user-account. Posts under the subheading "Top Posts" were not used in this process, because this study was not focused on the popularity of sharing or "liking" a particular post. If a post contained a series of images, only the first was analyzed. Profiles and posts made or used for advertising and selling purposes were not included, as these types of accounts are more likely to include identifying information, such as contact numbers and location.

To ensure greater accuracy of capturing what was divulged by users and to avoid making assumptions on latent meanings and themes, only manifest content was used in this study. This method was chosen due to the nature of content sharing, where posts are not always of, or created by, the profile owner. Information that included personality features, opinions, likes and dislikes, etc. was not included, as it is more subjective and not a major feature in identifying a person online or offline.

Units of Analysis

The units of analysis for this study were the 10 user profiles, out of the 31 randomly selected, that met the study's requirements. This also included the analysis of a total of 50 posts, five from each of the 10 profiles, posted within the timeframe of January 1st, 2017 to

December 31st, 2017. Each post was assessed based on user-created captions, and the content within the images and videos that the users shared.

Coding Procedures

The principal researcher recorded each instance of personally identifiable information that users included in their profiles as well as the captions of and content within selected photos and videos. Information was categorized into broader contexts based on the common themes found and how it was obtained once the data was retrieved and analyzed.

Setting and Materials

Data collection and coding took place at the principal researcher's residence. Necessary materials included access to a computer, connection to the Internet, an Instagram compatible device (iPod Touch, iPad, iPhone, Android, or computer), and a statistical number generator such as *Stat Trek* (https://stattrek.com). Registering for an Instagram account was also required in order to access profiles and view posts on the site.

Results

From the hashtag filter, *#trending*, a total of 31 user profiles and 50 posts were randomly selected and analyzed for content. Out of the 31 profiles, the first 10 that met the research criteria were chosen. From these profiles, five posts were randomly selected and assessed based on user-created captions and the content within the images and videos that the users shared. Due to the unknown type of information that would be collected, categories were created after examining the data. Instances of personal information were organized by the three methods of sharing: (1) user profile content, (2) captions, and (3) image and video content.

Categories

Information gathered from the data was classified into 14 distinct categories for each method of sharing, with an additional category (voice) included for images and videos. Name (1) was recorded when all or part of a user's real name was present in their username, profile, captions, images, or videos. Face (2), or facial recognition, was counted for user profile images and posts that depicted the user. If a user presented themselves in images or videos or stated their presentation within a caption it was recorded as gender (3). Location (4) was scored when a user stated their location in their profile or captions or if the user's location was clearly present within images or videos. A user's relationship (5) information was recorded based on indications of familial ties, such as mentions of being in a relationship or having children. Instances of other social media (6) were recorded when a user provided a link to another Instagram or social media account, such as Facebook or YouTube. Mentions of religion (7), sports (8), and ethnicity (9) were coded when a user indicated an affiliation with these groups in their profile or captions. Occupation (10) was counted when a user discussed their job and contact information (11) for when methods to reach out to the user, such as a website or email address, were provided. Ownership of user content (12) was recorded when a user used an identifying image, by using a watermark or a hashtag, such as *#me*, to claim something as their own. Recordings for features (13) were made when a user described their physical features in captions (e.g., *#shorthair*) and age (14) for when they stated their age in some fashion. An additional category was included for images and videos for when a user's voice (15) could be heard.

Findings

As shown in Figures 1 and 2, user profiles were the most common location for personally identifiable information to be shared. Captions were the least common method of sharing utilized by the sample profiles. Across all methods of sharing, gender presentation (52 instances), face (41), and location (25) were the most common pieces of information that were shared (see Figure 3). Other categories, such as religion, sports, and occupation were rarely shared.

Figure 2 depicts the percentage of where different types of information was shared, dependent on category occurrences (yes/no) for each profile. To aggregate this data, the total possible instances of information being shared within the relevant category was multiplied by the total number of profiles and the range of categories. This provided percentages reflecting whether data was present for a particular category, across all profiles, and for each of the three methods of sharing information.

Personal identifiable information that tended to be shared in user profiles the most was real names used as a username or listed in the profile (15 instances), gender presentation in profile pictures that depicted the user (11), and facial recognition through profile pictures that depicted the user (8). Captions that tended to display information through the use of hashtag filters were about the user's ethnicity (11), physical features (11), location (10), ownership of their posts' content (10), and gender presentation (8). In the images and videos portion of the analysis, facial recognition and gender representation scored the highest (33 each), followed by location (11).

Due to analyzing five posts per profile, data was more skewed towards information being

represented in captions, images and videos. To account for the varying number of possible instances in each of the three profile areas (user profiles, captions, and images and videos), the data shown in Figure 1 has been normalized to show the number of occurrences per 100 possible instances.

Discussion

The purpose of this study was to determine whether user profiles and their respective posts contained personally identifiable information and to explore the nature of this information. Ten active profiles on the social media site, Instagram, were randomly selected by using the hashtag filter, *#trending*, and a random number generator. Additionally, five posts within each of these profiles were randomly selected and a content analysis was then performed on both the profiles and their respective posts.

After normalizing the collected data to account for the varying number of possible instances, the results showed that user profiles contained nearly three times as much information than what was included in captions and almost five times that which was found within post images and videos (see Figure 1). Profiles also included a broader range of categories of personally identifiable information (28.57%) when compared to post images and videos (20.67%) and captions (13.57%; see Figure 2).

This may be explained by the idea that the textual information within a user's profile is directly controlled by the user, whereas the information shared within a post can be re-shared unbeknownst to the user (Boyd & Ellison, 2007; Zemmels & Khey, 2015). Zemmels and Khey (2015) argue that losing control of access to images versus textual information have drastically different consequences and ways of being attributed to individuals. They state that images and

videos can be denied, unless persons and activities are easily identified within them. However, textual information tends to be more easily controlled, less embarrassing, and not as intimately tied to the individual when compared to images. This is especially pertinent when the nature of personally identifiable information that was shared most was gender presentation (52 instances) and being able to recognize user faces (41).

Concerns over privacy and controlling access to personal information influence sharing behaviours online, such as making online purchases, sending emails, and impression management (Buchanan et al., 2007; O'Neil, 2001). This is largely due to the fact that control over private information about oneself is a growing issue in terms of public attitudes, Internet activities, and legal actions (O'Neil, 2001). For example, a video posted by User E showed tropical plants in the background and a person wearing a hat with the logo “Hollywood bets.” The principal researcher was able to search for this logo and discovered that it was for a sports gambling company located in Southern Africa (https://www.hollywoodbets.net).

Conclusion

The ease in which the principal researcher was able to obtain and locate information regarding where the User E resided suggests that the context and value of personally identifiable information is socially created. Thus, the ways in which this information is used needs to be explored and re-evaluated in order to remain relevant (Lange, 2008). This would allow for the contrasting of different types of personal information being shared through social media and the context of what people consider private or personal information relative to Internet security measures and behaviours.

Boyd (as cited in Zemmels and Khey, 2015) states that there are four traits of digital

media that create risk: persistence, searchability, replicability, and scalability. Due to these traits, "[…] digital media do not fade or deteriorate . . . copies can be infinitely reproduced and re-circulated" (p. 289). This suggests that when people post or share something, they leave a data-trace that is lasting, searchable, recordable, copyable, and shareable with others. Due to the nature of easily reproducing information online and the ease in which content is posted using social media sites as a third-party, users are not capable of overseeing or fully knowing their entire audience (Hogan, 2010; Livingston, 2004).

However, people continue to provide personal information online in order to be a part of the "new participatory culture," where sharing such information is vital in exchange for the services others are enjoying (O'Neil, 2001; Rauhofer, 2012; Robards, 2012). Buchanan et. al. (2007) refer to this as privacy fundamentalism, where users recognize the risks, yet are willing to trade personal information for further personalized online benefits and services, such as online banking and shopping. In this way, personal information becomes a currency in order to use "free" Internet services, where online activities are monitored by third parties and sold to other businesses (O'Neil, 2001).

Research around privacy and security has been primarily focused on concerns about adolescent use of social media, especially since personal information contained in profiles, such as use of one's real name or the town they were born in, could be used for malicious purposes (Boyd & Ellison, 2007). O'Neil (2001) describes the invasion of privacy as when information is collected about, or from, an individual for one purpose, but used for another, including when it occurs without the individual's knowledge or consent. Internet sites and businesses have been known to retain various forms of user information for purposes unknown to those individuals

(Buchanan et al., 2007; O'Neil, 2001). However, while there is a known possibility of third-party services retaining personal information, there are more insidious threats and risks.

Identity theft can occur online and offline, by using another individual's personal information to impersonate and exploit them. The types of information that is typically used may include date and place of birth, social security and credit card numbers, and a person's full name (Hille et. al., 2015). Using a *femtocell* – a device that intervenes and records phone conversations, text messages, and images sent through cellphones – is one of the many ways that this information may be obtained (Zemmels & Khey, 2015).

Along with the possibility of financial losses of identity theft, reputational damages may also occur, such as an individual's name being associated with illegal or embarrassing purchases. The effects of a damaged reputation are less studied; however, the possibilities are subtler and harmful to quality of life (Hille et al., 2015).

Limitations and Future Considerations

User posts were limited to those made within the timeframe of January 1st, 2017 to December 31st, 2017. Another aspect of this limitation that was unaccounted for was the way Instagram sorts posts, therefore the data was skewed towards the more recent posts being selected. This was further impacted by the random selection being limited to the 20 most recent posts. This study did not capture the prevalence of sharing personal information in all types of accounts, since access to viewing private accounts was not available. Some examples of accounts that were not considered and excluded from the study included accounts not in the principal researcher's native language, news reporting, and celebrity fan accounts. These types of user accounts were not appropriate to this study, because the accounts were not used for

personal reasons, therefore personal information could not be gathered.

While trying to obtain a profile during the third stage of random numbers, for the Profile D, an Internet connection error occurred. This caused a refresh of the Instagram website, which changed the pool of sample profiles being selected. The principal researcher concluded that while this changed the sample pool, it did not change the quality or quantity of information being collected. The connection error only further randomized which profiles would be selected. Instead, a new random number was generated, and the error was treated the same as a duplicate number or an account that did not meet the study's requirements.

There are issues surrounding the use of inductive approaches to measuring variables, leading to possible biases and invalidity. This is due to the nature of creating coding rules and measurements while analysis of content is being conducted, rather than having defined variables before collecting data (Macnamara, 2005). The principal researcher was the only coder present, which also reduced reliability. A way to avoid these issues in the future would be to use a coding list, multiple coders, and a multivariate approach where data is collected from other social media platforms and can be compared to the findings reported here.

References

Boyd, D., & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Buchanan, T., Paine, C., Joinson, A. N., & Reips, U. D. (2007). Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology, 58*(2), 157–165. https://doi.org/10.1002/asi.20459

Hille, P., Walsh, G., & Cleveland, M. (2015). Consumer fear of online identity theft: Scale development and validation. *Journal of Interactive Marketing, 30*, 1-19. https://doi.org/10.1016/j.intmar.2014.10.001

Hogan, B. (2010). The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society, 30*(6), 377-386. https://doi.org/10.1177/0270467610385893

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology, 31*(2), 177–192. https://doi.org/10.1002/ejsp.36

Lange, P. (2008). Publicly private and privately public: Social networking on YouTube. *Journal of Computer-Mediated Communication, 13*(1), 361-380. https://doi.org/10.1111/j.1083-6101.2007.00400.x

Livingston, S. (2004). Media literacy and the challenge of new information and communication technologies. *The Communication Review, 7*(1), 3-14.

https://doi.org/10.1080/10714420490280152

Macnamara, J. (2005). Media content analysis: Its uses, benefits and best practice methodologies. *Asia Pacific Public Relations Journal, 6*(1), 1-34.

O'Neil, D. (2001). Analysis of Internet users' level of online privacy concerns. *Social Science Computer Review, 19*(1), 17-31. https://doi.org/10.1177/089443930101900103

Rauhofer, J. (2012). Future-proofing privacy: Time for an ethical introspection? *Surveillance & Society, 10* (3/4), 356-361. https://doi.org/10.24908/ss.v10i3/4.4531

Robards, B. (2012). Leaving MySpace, joining Facebook: 'Growing up' on social network sites. *Continuum: Journal of Media & Cultural Studies*, *26*(3), 385-398. https://doi.org/10.1080/10304312.2012.665836

Zemmels, D., & Khey, D. (2015). Sharing of digital visual media: Privacy concerns and trust among young people. *American Journal of Criminal Justice, 40*(2), 285-302. https://doi.org/10.1007/s12103-014-9245-7

Figure 1

Proportion of Each Method Shared

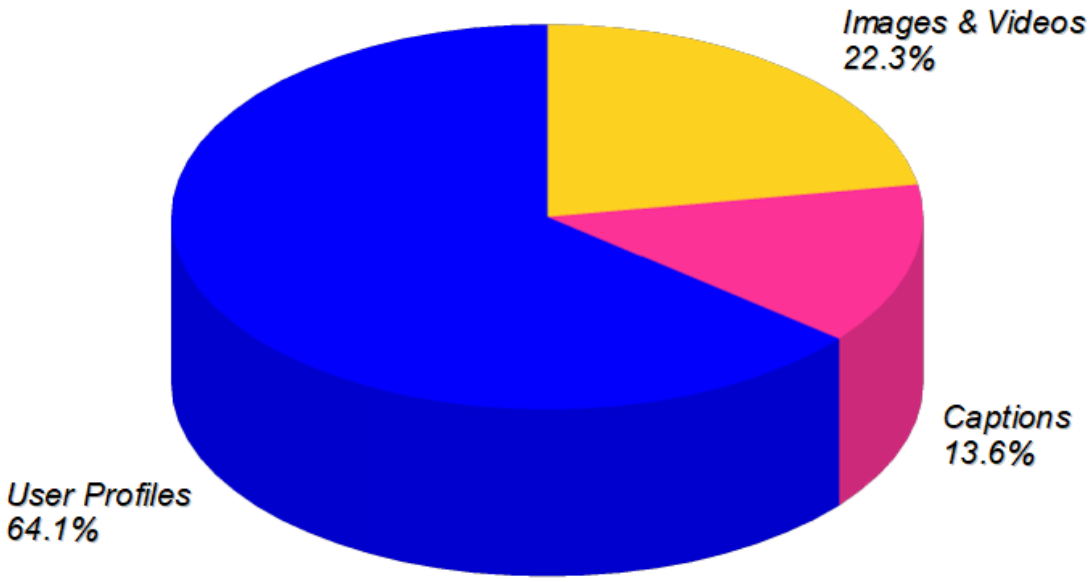

Figure 2

Range of Information Shared

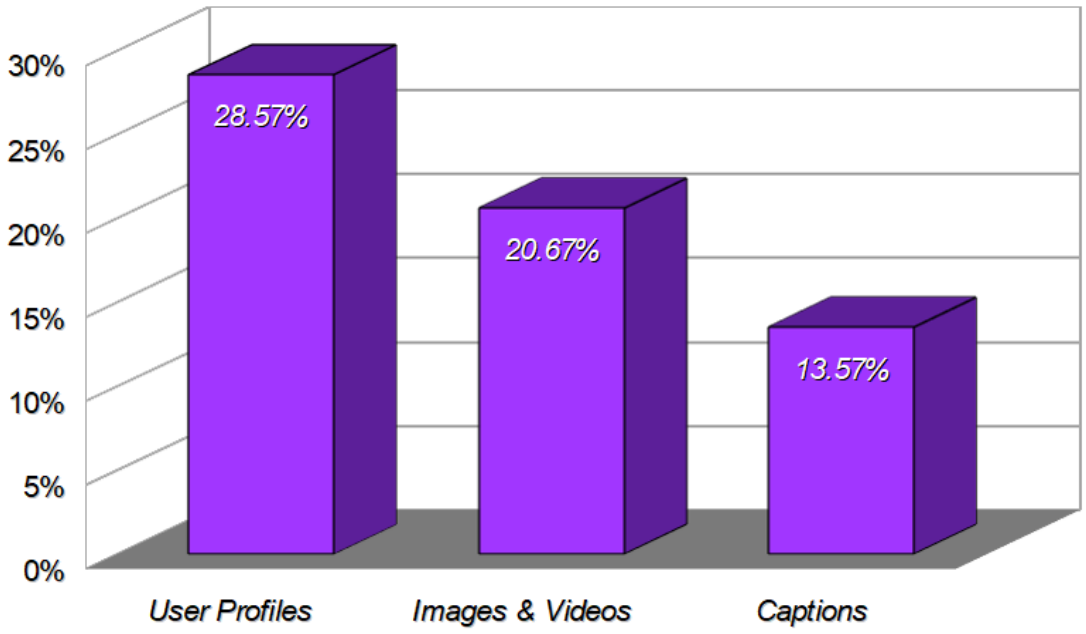

Figure 3

Total Times Each Category was Shared