iTweet about #Privacy
Mapping Privacy Frames in Twitter Conversation

Federica Fornaciari
Department of Arts and Humanities
National University
La Jolla, CA, USA
e-mail: ffornaciari@nu.edu

***Abstract*— Adopting a pragmatic bottom-up approach, the current study applies semantic network analysis and discourse analysis to unfold individual frames of privacy emerging in Twitter. To do so, the author collected and analyzed 100,000 publicly available Tweets selected using the word "privacy." The following two overarching questions guided the study: What are the frames that emerge in relation to privacy on Twitter? How are these frames discussed? Through a mixed method approach, the author identified the following nine frames of privacy: Privacy and Technology, Personal Privacy, Legal Privacy, Fundamental Privacy, Privacy Concerns, Spatial Privacy, Gossip, Trading Privacy, and Expected Flow of Information. The author also developed robust dictionaries to automate frame detection. In a future step, the author plans to use these dictionaries of privacy to analyze larger corpora of text and reach a meaningful understanding of how individuals frame privacy in everyday conversation.**

Keywords-network analysis; discourse analysis; framing theory; privacy; Twitter

I. Introduction

In a technologically driven communication environment, privacy is undoubtedly a major concern influencing how we share or withhold information – and how we think about personal data. Perhaps paradoxically, many voice their privacy concerns in rather public venues, such as social media. These public platforms facilitate researchers who wish to explore, unobtrusively, the textures and patterns of user's casual discussions – and thereby observe how individuals understand and frame reality [1]. Based on the assumption that Twitter discussion mimics an online word-of-mouth [2], the author suggests combining semantic network analysis and discourse analysis to explore the frames of privacy emerging on Twitter, and to develop dictionaries that facilitate frame detection. The results of such study begin to shed light upon how individuals discuss privacy in everyday conversation.

Privacy is an increasingly relevant issue in today's computing era. Currently, many individuals store personal data in the Cloud, a virtual data storage where users can archive and remotely access information [3]. For many, accessing the Internet and sharing information online has become a routine activity. Yet – partly due to the gained popularity of online-networked platforms – privacy increasingly becomes a concern for users who desire to protect their data. Belonging in this category of networked environments, social media too are platforms where users share information becoming potential victims of privacy loss. However, social media are also possible vehicles for discussing concerns and solutions related to privacy.

Section two provides a short review of relevant literature to contextualize framing theory, semantic network analysis, and discourse analysis. Section three presents the research questions and describes the methodological approach adopted in the current study. Section four introduces some preliminary findings. Finally, section five discusses findings and limitations.

II. Literature Review

A. Framing theory

Goffman [4] explained that individuals approach the complexity of reality developing or borrowing primary frames, or "schemata of interpretations," based on abstract principles that organize, untangle, and simplify reality. Frames emerge through symbolic forms of expression and provide structures that enforce preferred interpretations of the social world. Frames may be individual or collective [5], and emerge within different occurrences of the communication process: the communicator, the text, the receiver, and the culture itself [6]. Available frames are either consciously recognized or unconsciously processed, often influencing how people understand, assess, remember and discuss issues [7].

B. Semantic Network Analysis and Discourse Analysis

Semantic network analysis is a specific type of automated content analysis that investigates text to explore the networks that emerge from the occurrences and co-occurrences of concepts [8]. In such a way, semantic network analysis allows drawing conceptual maps as they emerge in text.

Discourse analysis, on the other hand, is a qualitative process seeking to provide deeper explanation of meaning through the analysis of themes and patterns that emerge from texts [9] [10]. It also takes into account the role of context in developing the semantic networks of "privacy." Such a qualitative approach may be used to strengthen the findings obtained in the quantitative steps of this research project.

The current study combined quantitative semantic network analysis [11]–[13] with qualitative discourse analysis [9] [10]. Such mixed method approach enabled the author to validate, contextualize, and strengthen the results obtained through each method of analysis [14].

III. Method And Research Questions

The current study is twofold. First, the author combined semantic network analysis and discourse analysis to map and explore the frames of privacy emerged in Twitter using a bottom-up approach [15]. Then, the author developed robust dictionaries to automate frame detection in large corpora of text. In a future step, the author plans to use these dictionaries to analyze larger samples of tweets and thereby develop a more robust understanding of existing *individual frames* of privacy, which refer to our cognitive understanding of privacy [4].

In the current study, the author analyzed 100,000 publicly available tweets collected using the keyword "privacy" between July 1st 2016 and July 25th 2016 (the software HootSuite facilitated the collection of tweets). Considering the nature of the current study, the author did not distinguish between tweets and re-tweets, as both were considered equally useful and meaningful in frame implementation. After collection, the author used the software Automap [11] to generate frequency lists and begin the analysis.

The following two overarching questions guided the study:

RQ1 – *What are the frames that emerge in relation to privacy on Twitter?*

RQ2 – How are these frames discussed?

The quantitative analysis, implemented to address RQ1 included three steps.

To address RQ1, the author implemented several steps. First, the author imputed the tweets in the software Automap to generate frequency lists. This resulted in almost 10,000 item recorded in a frequency list. Using Automap, the frequency list was refined by deleting non-content bearing elements such as articles, conjunctions, and other noise from the text [11].

Second, the author manually processed the frequency list to qualitatively assess the contexts of use of each word. To undertake this second step, the author read and reread carefully the frequency list. During each reading, and informed by existing literature, the author added new themes as they emerged from the words in the list. For instance, keywords referring to legislations – such as the Health Insurance Portability and Accountability Act (HIPAA) and the Family Educational Rights and Privacy Act (FERPA) were included in a "legal privacy" dictionary.

As a result, the author developed lists of recurring terms and expressions, and clustered these into overarching sub-themes and groups that co-occurred with the word "privacy" in Twitter conversation. The words in the frequency list were sorted into 40 sub-themes. These themes were then combined into nine overarching frames including the following: privacy and technology, personal privacy, legal privacy, fundamental privacy, privacy concerns, spatial privacy, gossip, trading privacy, and expected flow of information. Each frame was subsequently analyzed through qualitative discourse analysis to allow a deeper, qualitative understanding of how privacy was discussed within each category.

Third, the author and a coder manually processed the list of the frequencies and placed each word in the corresponding category. Agreement between the two researchers was then calculated to gauge the reliability of the third step. Intercoder reliability scored between .88 and .95 [16]. These three phases enabled the author to map the semantic networks of privacy as they emerged from the 100,000 tweets collected. It also allowed the author to start developing robust dictionaries of the individual frames of privacy.

To address RQ2, the author used the keyword in the dictionaries to select sub-samples of tweets belonging in each theme emerged from the semantic network analysis. For example, the theme "privacy is a fundamental human right" emerged from words such as: human right, sacred, freedom, liberty, and universal. These keywords were used to retrieve a sub-sample of tweets from the original sample. Each sub-sample consisted of 20 tweets randomly selected. The author further analyzed each sub-sample through discourse analysis to understand and clarify how each theme was discussed in the tweets.

IV. Preliminary Findings

After a preliminary analysis, eight frames emerged. The frames were labeled as follow: privacy and technology, personal privacy, privacy concerns, legal privacy, fundamental privacy, spatial privacy, gossip, and trading privacy.

The frame "Privacy and Technology" implies that when new technology is introduced, new privacy concerns develop.

"Personal Privacy" suggests that privacy is related to sociality, social roles, relationships, and personal feelings.

"Privacy Concerns" emerges from tweets suggesting that privacy infringements generate problems and that the tradeoff is often unfair.

"Legal Privacy" emphasizes that the government, regulations, contextual norms, permission, and transparency are fundamentally related to privacy.

"Fundamental Privacy" emerges when users frame privacy as a fundamental human right suggesting that, as such, it should be protected.

"Spatial Privacy" emerges in tweets that describe privacy in terms of access or boundary control.

"Gossip" is implemented when users describe gossip as an invasion of someone's privacy.

Finally, "Trading Privacy" emerges when tweets focus upon the economic value of information, implying that personal data are commodities that can be stolen or sold.

Table 1, in the next page, summarizes the eight frames identified providing examples of the dictionaries used for frame detection. It also delivers data on the cumulative frequencies to provide an overview of frame implementation in the sample analyzed.

TABLE I. CUMULATIVE FREQUENCIES OF FRAMES

Frame	Example of Keywords	Cumul. Freq.
Privacy and Technology	Cellphone, Pokemon Go, Google, Facebook, cameras…	81,812
Personal Privacy	Boyfriend, relationship, girlfriend, angry, annoying…	51,697
Privacy Concerns	Data, concerns, dossiers, spy, cookie, surveillance, war, security	35,232
Legal Privacy	Laws, setting, bill, banned, transparency, setting, court, Obama, permission, health, education…	34,408
Fundamental Privacy	Right, need, important, respect, essential, hope, human...	32,028
Spatial Privacy	Border, gates, space, location, bedroom, bar, wall, cars…	12,719
Gossip	Paparazzi, famous, popstar, vanityfair, popularity…	8,104
Trading Privacy	Business, consumers, property, buy, marketing, commercial	5,822

V. DISCUSSION

In Twitter discussion, privacy surfaces as multifaceted and complex. Users discuss privacy as a social construct that entails a variety of components and perspectives.

Not surprisingly "privacy and technology" was the most frequent frame adopted in Twitter discussion. People often express their concerns about personal data stored in networked environments, such as Facebook and Google. When voicing these concerns, users also criticize the obscurity and scarce usability of existing privacy policies. Strong privacy concerns are frequently channeled to new technologies such as face recognition software, drones, and Pokemon Go. These concerns develop a very typical pattern of reactions to the introduction of new technological devices that emerge as powerful, unexpected, and often intimidating due to their potential for information collection, processing, and shareability.

Current research on social capital emphasizes that privacy is fundamentally related to publicity and sociality. In fact, needs for connection and sociality often encourage individuals to share personal information [17]. As a social media, Twitter could be considered a preferred platform for discussing the importance of relationships and sociality, and the risks that privacy infringement may cause in this respect. The high frequency of tweets referring to "personal privacy" reflects that privacy, sociality, and publicity meaningfully intersect in individuals' frames of privacy as well.

The predominance of the frame "legal privacy" emphasizes that Twitter users are often adopting a legal or ethical framework to understand and discuss privacy. They emphasize the role of government in the protection of privacy, while highlighting the role of contextual norms of information flow [18].

As shown in the frequency of "fundamental privacy", Twitter users understand privacy as a sacred and fundamental human right that should always be protected. Moving forward, the author believes that the frame "Gossip" should be included as a sub-theme of "fundamental privacy." In fact, the frame "gossip" suggests that celebrities are human beings and – as such – deserve privacy.

Due to the nature of the current contribution (i.e., short paper), the author emphasizes the need to further refine the methodology and to more closely interpret the findings. Moving forward, the author intends to use software for Natural Language Processing such as Nooj [19] in order to automatically distinguish between orthographical sequences of letters and relevant linguistic units. Moreover, the author will use the dictionaries of privacy developed during the current study to facilitate extracting semantic information from text [as suggested in 20]. The dictionaries will prove particularly useful in analyzing tweets collected over a longer and therefore more representative timeframe. Despite its limitation, the author believes that the current study provides valuable toolkits to automate the detection of individual frames of privacy in Twitter conversation.

REFERENCES

[1] A. Jalal, and M. Zaidieh, "The use of social networking in education: Challenges and opportunities," World of Computer Science and Information Technology Journal, vol. 2, pp. 18-21, 2012.

[2] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the American Society for Information Science and Technology,* vol. 60,pp. 2169–2188, 2009.

[3] E. A. Aldakheel, "A Cloud Computing Framework for Computer Science," Education. Master's Thesis, Bowling Green State University, United State of America, 2011.

[4] E. Goffman, "Frame analysis: An essay in the organization of experience," Northeastern University Press, Boston, 1974.

[5] S. D. Reese, "Prologue: Framing public life, a bridging model for media research," in Reese, S.D, Gandy, O.H., and Grant, A.E. (Eds.), Framing public life: Perspectives on media and our understanding of the social world, Erlbaum, Mahwah, NJ, 2001.

[6] Z. Papacharissi and M. de Fatima Oliveira, "News Frames Terrorism: A Comparative Analysis of Frames Employed in Terrorism Coverage in U.S. and U.K. Newspapers," *The International Journal of Press/Politics*, vol. 13, pp. 52–74, January 2008.

[7] W. Gamson, and A. Modigliani, "Media discourse and public opinion on nuclear power," *American Journal of Sociology*, vol. 95, pp. 1-37, 1989.

[8] W. van Atteveldt, J. Kleinnijenhuis, and N. Ruigrok, "Parsing, semantic networks, and political authority using syntactic analysis to extract semantic relations from Dutch newspaper articles," *Political Analysis*, vol. 16, pp. 428–446, October 2008.

 ISBN: 978-1-61208-552-4

[9] N. Fairclough, “Critical analysis of media discourse,” in: Marris, P., Thornham, S., (Eds.), Media studies. New York University Press, New York, pp. 308–325, 2000.

[10] T. vanDijk, “The study of discourse,” in: Treun A., van Dijk (Eds.), Discourse as structure and process. Sage, London, pp. 1-34, 1997.

[11] K. M. Carley, D. Columbus, and A. Azoulay, “Automap user’s guide 2012,” CASOS technical report Carnegie Mellon University, Pittsburgh, PA, 2012.

[12] A. E. Smith and M. S. Humphreys, “Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping,” *Behavior Research Methods*, vol. 38, pp. 262–279, 2006.

[13] M. Stubbs, “Text and Corpus Analysis,” Computer-Assisted Studies of Language and Culture. Cambridge, MA, Blackwell, 1996.

[14] A. Tashakkori, and C. Teddlie, “Handbook of mixed methods in social and behavioral research.” Thousand Oaks, CA: Sage, 2003.

[15] D. J. Solove, “Conceptualizing privacy,” *California Law Review*, pp. 1087–1155, 2002.

[16] W. Perreault, and L. Leigh, “Reliability of nominal data based on qualitative judgments,” *Journal of Marketing Research,* vol. 26, pp. 135–148, 1989.

[17] N. B. Ellison, J. Vitak, C. Steinfield, R. Gray, and C. Lampe, “Negotiating privacy concerns and social capital needs in a social media environment,” in Trepte, S. and Reinecke, L. (eds*.) Privacy Online,* Berlin: Heidelberg, 2011.

[18] H. Nissenbaum, “Privacy in context. technology, policy, and the integrity of social life,” Stanford: Stanford Law Books, 2010.

[19] M. Monteleone, and S. Vietri, “The NooJ English dictionary. Formalising natural languages with NooJ 2013,” Selected Papers from the NooJ 2013 International Conference, pp. 69-84, 2014.

[20] M. P. di Buono, A. Maisto, and D. Pelosi, “From linguistic resources to medical entity recognition: a supervised morpho-syntactic approach,” The First International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2015), IARIA, pp. 81-86, Apr. 2015.