Social Media Visualization & Privacy

Danyel Fisher
Microsoft Research
1 Microsoft Way
danyelf@microsoft.com
+1-425-722-5098

ABSTRACT
Data visualization can help make privacy issues more visible within a social media space, both by creating explicit visualizations of privacy policies, and by implicitly or accidentally showing what information is available to a user. In this paper, we provide several vivid examples from past literature of how visualization and aggregation of network information can provide knowledge about a user that they might not have chosen to provide themselves. While these visualizations might be seen as threats, we prefer to view them as a warning that the information *is* available.

Author Keywords
Social media, information visualization, privacy

INTRODUCTION
Privacy threats in the social media space can be hard to get our heads around. When a user voluntarily shares information about themselves, they may not realize how powerful that information is in aggregate; one participant on a social media system might implicate others through their behavior.

Data visualization can be of tremendous benefit to understanding these complex spaces. By its nature, visualization can put large amounts of information on screen and available. Projects that are not intended to illustrate privacy issues may do so unintentionally, but vividly. Providing this sort of information can have a dual effect: it can make them highly aware of the privacy implications of their own actions, and it can help system designers understand the implications of their settings and configurations.

In this position paper, I review several visualizations of social media that provide a broad picture of activity within social media systems. While these visualizations were originally designed with goals such as "allowing users to reflect on the behavior of a network," they are presented here instead for their privacy implications.

Submitted to Workshop on Collaborative Privacy Practices in Social Media
CSCW 2011, March 19–23, 2011, Hangzhou, China.

VISUALIZATION OF SOCIAL INTERACTION
In this review, I have selected three different visualizations that show radically different views, but reveal substantial information about individual behavior (Figure 1). **Vizster** [2] is a social-network representation of a social networking system. It shows all the social network information that a single user was able to extract from the system: not only their friends, but all of their friends' friends, and their interconnections.

It is unclear what the privacy expectations of a user using a social networking site are. A Facebook user today might now expect that a friend can see our friends—but they might not expect that this implies the degree of connection. This graph, however, makes it vividly clear who those connections are: I can know that you are connected to Bob. We can see that the author has some friends who are well-connected to other of their friends, and some friends who are poorly connected. (Vizster also has features that specifically show communities and connected groups.)

The **Newsgroup Crowd** image is a bubble chart, showing information about the frequency of communication. Unlike the Vizster diagram, it does not show connections between people; however, each person does have unique data associated with their own past interactions. The Newsgroup Crowd is based on public data, extracted from archives. While the user's usual view is disaggregated data, the Crowd brings it together, showing a person's activity history in one burst. The diagram was originally intended to show how groups differ from each other (see [6] for examples), but, as a side effect, shows individual user behavior. The participant who is represented by the rather large bubble might be casually known as a frequent user—but the fact that they were the *most* frequent poster might be less well-known. (Indeed, turning to some users and asking them "who shows up on 300 of 365 days" might be more than enough to identify that individual.) The visualization allows them to reflect on the footprint they are leaving on the group.

The **Pairgram** is a time line showing when a pair of users search for and read documents. This is different from the other visualizations presented, as the data is based on the

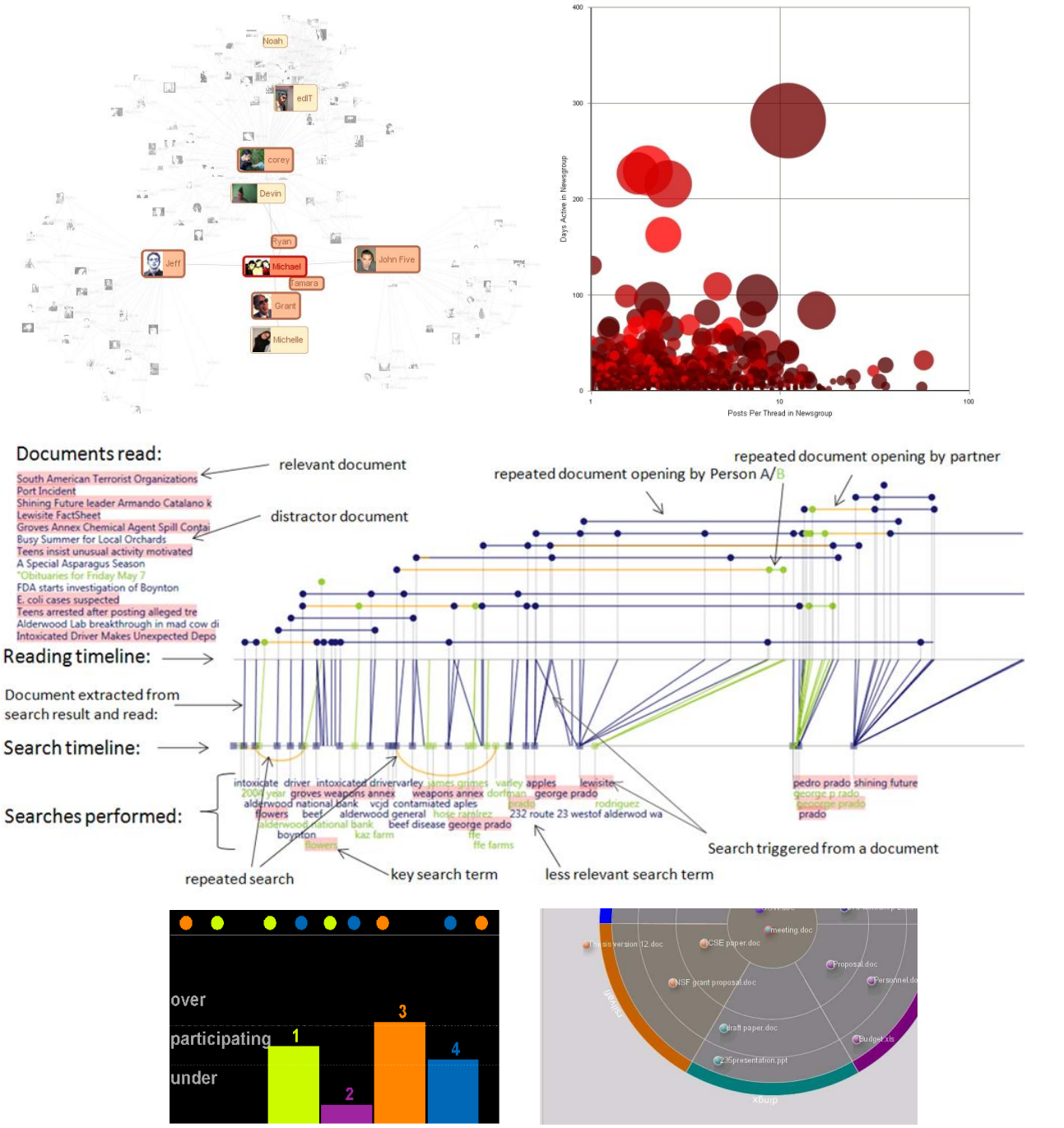

Figure 1, left to right, top to bottom: Vizster [2] visualizes a friend network on Friendster, an early social networking site. Note that network links between friends are shown—and, indeed, many nodes are not connected to the central focus. This helps understand that a user on Friendster can see their friends-of-friends, and their interconnections. This **Newsgroup Crowd** [6] shows posters to a Usenet forum, "alt.politics.clinton". Each bubble represents a single person; the y-axis shows the number of distinct days that the person posted. (The x-axis shows the average length of threads that the person participated in.) The user who contributed on 300 distinct days might be worried to realize how *visible* their aggregate behavior is. The **Pairgram** (unpublished) was a visualization developed to help analyze log files of a collaborative analysis project [3]. While this was not a social media system, the collaborative aspects illustrate the ways that user information can be made visible. DiMiccio *et al's* Shared Display [1] shows a visualization of participation in a small chat. **Impromptu** [5] explicitly shows privacy settings by placing documents: a document at a particular place indicates its degree of being shared.

None the less, it richly exposes information about their methods for using the system. As a result, it is a useful tool for both understanding what information is collected, and for allowing users to reflect on their own participation.

Last, DiMiccio *et al* [1] created a visualization to reflect behavior of speakers in a conversation—speaking more would grow a bar, while speaking less would cause it to shrink. This information again reflected data about users' personal behavior.

Sharing Information Collaboratively

Other researchers have experimented with displays that help users explicitly share data. These systems are meant to help show what information is being shared, help users see each others' activity, and understand how to control privacy settings. One example is Impromptu [5], which allows users to drag documents further in toward the center of a space (more shared) or further outward (less shared). One can imagine extending this metaphor outward to other types of data.

PRIVACY-PRESERVING VISUALIZATION

These visualizations we have shown here have different needs for specificity of data. The Newsgroup Crowd is based on statistical data, while the Vizster and the Pairgram are based on detailed, individual data. An analyst who wanted to fulfill the main goal of the Newsgroup Crowd (learning about the group) without wanting to expose this degree of aggregate, private information might use statistical methods to blur the data.

This field of research is known as *differential privacy* (e.g. [4]). Differential privacy essentially avoids showing information narrow enough to identify a single person. For example, it produces data with intentional errors (within bounds) in response to aggregation questions. Several toolkits have begun to support differential queries [4], which allow users to easily request aggregated data with certain, strict privacy guarantees. It might be possible to modify visualizations like the Newsgroup Crowd—where the individual is not critical—to show a probabilistic distribution rather than a detailed set of bubbles.

CONCLUSION

Data visualization can allow us to reflect on social media behavior and to control data access. This can be a valuable warning to users and a useful control panel. We should continue to investigate ways that we can create visualizations that are non-intrusive but informative, just as we continue to build systems with more secure controls over our data. Systems that allow us to implement formally-defined privacy controls may make a good back-end to these visualizations.

BIOGRAPHY

Danyel Fisher is a researcher at Microsoft Research. His research centers on online collaborative systems, and on information visualization. His recent research has shown He received his PhD from the University of California, Irvine in 2004.

ACKNOWLEDGEMENTS

My thanks to Petra Isenberg, who created the first versions of the Pairgram, and Fernanda Viegas, who created the first version of the Newsgroup Crowds.

REFERENCES

1. J.M. DiMiccio, A. Pandolfo, W. Bender. Influencing Group Participation with a Shared Display. CSCW 2004.
2. J. Heer, d. boyd. Vizster: Visualizing Online Social Networks. IEEE Information Visualization (InfoVis), 32–39, 2005
3. P. Isenberg, D. Fisher, M. R. Morris, K. Inkpen, and M. Czerwinski. An Exploratory Study of Co-located Collaborative Visual Analytics around a Tabletop Display. In Proceedings of Visual Analytics Science and Technology (VAST), pages 179–186, 2010. IEEE.
4. McSherry, F., Privacy Integrated Queries, in Communications of the ACM, Association for Computing Machinery, Inc., 1 September 2010
5. Rode, J., Johansson, C., DiGioia, P., Silva Filho, R., Nies, K.,Nguyen, D., Ren, J., Dourish, P., and Redmiles, D. 2006.Seeing Further: Extending Visualization as a Basis for Usable Security. Proc. Symp. on Usable Privacy and Security SOUPS 2006 (Pittsburgh,PA), 145-155.
6. Turner, T. C., Smith, M. A., Fisher, D., and Welser, H. T. (2005). Picturing Usenet: Mapping computer-mediated collective action. Journal of Computer-Mediated Communication, 10(4), article 7. http://jcmc.indiana.edu/vol10/issue4/turner.html