Discovery and Visualization of Scholarly Information Diffusion in Twitter Networks

Anatoliy Gruzd, gruzd@dal.ca
Dalhousie University, Halifax, NS, Canada

Position paper submitted to
Workshop on Mapping Ideas: Discovering and Information Landscapes
San Diego State University, San Diego, CA
August 1-2, 2012

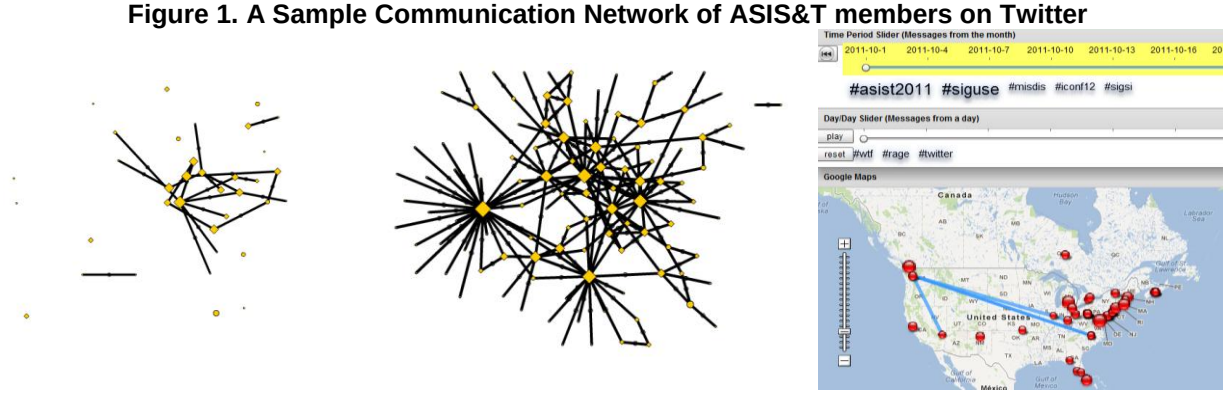
Since early 2000s, the number of people who joined various social media sites has increased exponentially. For example, according to a recent report, there are over 900 million active Facebook users (Facebook.com) and over 140 million active Twitter users (Twitter.com). Like the general public, scholars are also increasingly adopting social media for their professional work (e.g., Collins & Hide, 2010; Gruzd et al., 2011; Ross, et al, 2011; Rowlands, et al, 2011). There are many reasons why this is happening. First, social media are now widely available, easy to use and free to all users. Second, there is no traditional peer-review process to delay or even prevent publication. Social media sites allow researchers to quickly share their ideas or preliminary findings and even solicit feedback in real time. Third, most content posted on social media is instantly available to fellow community members for consumption and dissemination via virtual word of mouth. As a result, content published using a social media-type technology tends to attract a larger audience more quickly and at a much lower cost as compared to more traditional subscription-based publishing methods (whether digital or print). Aside from being easily discoverable and accessible, social media are very interactive in nature, enabling the content creators to interact more directly with the content consumers. In scholarly communities, these kinds of direct interactivities are normally found at conference presentations or panel discussions in which a presenter can easily get feedback from the audience, and the audience can ask follow-up questions to clarify some points of the presentation.

The current study seeks to better understand scholarly communication and information sharing practices on social media and whether social media can be used to form and strengthen scholarly networks on a global scale. To achieve the study objectives, we apply text mining, information visualization and social network analysis to *public* social media data in order to capture and use documentary evidence (as produced and shared by scholars) to discover, visualize and study communication and information networks that form among researchers. The types of network data that we capture is very comprehensive and include data points from across three different dimensions: distance, time and disparate forms of media. An example of the proposed approach and the type of interactive visualizations that are being developed and used as part of this research can be found at AcademiaMap.com, being developed by Gruzd and his team. One of the components of AcademiaMap is an online Geographic Information Visualization (GIV) that tracks Twitter conversations and visualizes communication connections between scholarly users of Twitter.com (a microblogging platform for sharing short messages) from across the globe.

Below is an example of how AcademiaMap analyzes the public messages of everyone who follows the Twitter account - @asist_org, which belongs to the American Society for Information Science & Technology (ASIS&T). In this example, the first step in such analysis is to transform text-based data, in this case –Twitter messages, into various network representations. For instance, Figure 1a depicts Twitter data in the form of a "*who talks to whom*" network during the ASIS&T annual conference in October, 2011. Such network visualization can be used to discover key members or groups within this community, identify isolates, and generally better understand how social media are being used for communication and information dissemination. The next step is to add the temporal component in order to examine how this

network changes over time. Figure 1a shows how the network looked before the ASIS&T 2011 conference started and Figure 1b – the network after the conference ended; when more people joined the conversation. By comparing the network in different time periods, it becomes possible to identify more or less active parts of the network and detect where the spread of information tends to begin or end. The next step is to plot this network on a geographical map by placing each node that represents people into their corresponding locations. Such geo-based network is especially useful in studying how connections with remote collaborators are formed and maintained. See Figure 1c. This particular example shows communication connections for one of the frequent posters in this network from Seattle, WA, indicating how information flows between that user and other association members. Finally, the geo-based visualization can be expanded even further by extracting important concepts that are being discussed and associating these concepts with the timeline as well as individual nodes/people (see the top part of Figure 1c). In the sample case, the topics in the visualization (also known as hashtags) represent popular concepts of importance to this scholarly community during the month of October. Besides discussion around the ASIS&T annual conference (hashtags #**asist2011**), members of this group were also interested in concepts such as #**siguse** – the ASIS&T Special Interest Group on Information Needs Seeking and Use.

Figure 1. A Sample Communication Network of ASIS&T members on Twitter



(a) before the conference (b) after the conference (c) geo-based visualization

An interactive network visualization such as the one described above is an effective method to filter out the noise from a large quantity of conversational data that are being shared among group members of scholarly community. By mapping the relationship between people and concepts over time and space, it will be easier to identify more popular and potentially more important conversations and concepts to be brought to the forefront. Furthermore, the proposed method can also be used to identify influential voices among group members, as well as core groups in the network and pin point a potential knowledge gap or potential bottleneck in information flow within online networks.

References

Collins, E. and Hide, B. (2010). Use and relevance of web 2.0 resources for researchers. Paper presented at the International Conference on Electronic Publishing, Helsinki, Finland, 16-18 June, 2010.

Gruzd, A., Wellman, B., and Takhteyev, Y. (2011). Imagining twitter as an imagined community. American Behavioural Scientist, 55 (10), 1294-1318.

Ross, C., Terras, M., Warwick, C., and Welsh, A. (2011). Enabled backchannel: Conference Twitter use by digital humanists. Journal of Documentation, 67(2), 214-237.

Rowlands, I., Nicholas, D., Russell, B., Canty, N., and Watkinson, A. (2011). Social media use in the research workflow. Learned Publishing, 24(3), 183–195.