Chapter 14

Sampling Public Sentiment Using Related Tags (and User-Created Content) Networks from Social Media Platforms

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT

The broad popularity of social content-sharing sites like Flickr and YouTube have enabled the public to access a variety of photographs and videos on a wide range of topics. In addition to these resources, some new capabilities in multiple software programs enable the extraction of related tags networks from these collections. Related tags networks are relational contents built on the descriptive metadata created by the creators of the digital contents. This chapter offers some insights on how to understand public sentiment (inferentially and analytically) from related tags and content networks from social media platforms. This indirect approach contributes to Open-Source Intelligence (OSINT) with nuanced information (and some pretty tight limits about assertions and generalizability). The software tools explored for related tags data extractions include Network Overview, Discovery, and Exploration for Excel (NodeXL) (an open-source graph visualization tool which is an add-in to Microsoft Excel), NCapture in NVivo 10 (a commercial qualitative data analysis tool), and Maltego Tungsten (a commercial penetration-testing Internet-network-extraction tool formerly known as Maltego Radium).

INTRODUCTION

As of late 2013, 73% of online adults used social networking sites. Forty-six percent of adult Internet users posted original photos and videos online that they created, as of August 2012. (Brenner, 2013) Social media platforms are trading cloud space and cloud-delivered services in exchange for the partial or full use of the content of the photos and videos and labels provided by users. Social computing through content sharing has come to the fore in a big way. The popularization of multimedia content sharing through social media platforms has resulted in the broad avail-

DOI: 10.4018/978-1-4666-6493-7.ch014
ability of billions of user-generated texts, photos, videos, and multimedia. Various content-sharing social media sites offer a heterogeneity of digital contents (multimodal contents) from a wide range of users to a wide audience. Those who created the contents will often “tag” or label this content with keywords and / or short phrases to describe the resource, which enables smoother text-based categorization and searchability for archival and sharing. Individual tagging involves the content creators labeling their own contents, with little input by others in their ego neighborhood or larger user community (such as in photo-sharing and video-sharing social media communities); collaborative tagging involves more broad-scale labeling of contents (such as in social bookmarking sites). (Some tag clouds are access-points for users to query databases and to access particular contents by clicking relevant terms.) Some authors explain:

Tags function both as content organizers and discoverers. As content organizers, tags enable tag creators to annotate and categorize a resource so that it can be retrieved subsequently with ease. Tag consumers will use those same tags to locate that resource. As content discoverers, tags could be used as a means to tap into the collective intelligence of tag creators to make serendipitous discoveries of additional relevant resources. Furthermore, through tags, a tag consumer is able to find like-minded tag creators with resources that meet his or her information needs, potentially leading to the creation of social networks (Razikin, Goh, Chua, & Lee, 2008, p. 50).

The label metadata is value-added information; they are a form of electronically-mediated social communication. These tags result in informal “folksonomies” (non-expert-created taxonomies) that are used to identify contents, enhance user searches, and increase the findability of contents. This informal indexing involves what Hotho calls “lightweight knowledge representation” (Hotho, 2010, p. 57). They are “in vivo” codes—those used by people in the actual lived environment (even if not necessarily anything that professional archivists would use). In the attention economy of social media, having content that goes “viral” or gains exponential attention in a short time is considered grounds for bragging rights. There is empirical research that suggests that fame of contents in the photo- and video-sharing site Flickr is not generally one of exponential and sudden growth. One study found that users are able to discover new photos within hours after being uploaded and that 50% of the photo views are generated within the first two days (van Zwol, 2007, p. 184). The popularity of particular user-generated contents are a feature of social networking behavior and “photo pooling.” Photos that gain a lot of attention transcend local geographical interest. The author explains that “the geographic distribution is more focussed around a geographic location for the infrequently viewed photos, than for the photos that attract a large number of views” (van Zwol, 2007, p. 184).

In terms of tag popularity (a layer of abstraction up from the contents), some researchers suggest that the more popular social tags are the more meaningful ones (Suchanek, Vojnović, & Gunawardena, 2008). Others found that the most popular tags tend to be “visually representational of contents” for photo sharing sites (Sun & Bhowmick, 2009). The growth of popularity of images in Flickr grows in “a steady linear growth of popularity over several years” (Cha, Benevenuto, Ah, & Gummadi, 2012, p. 1066); they do not tend to acquire exponential popularity. These researchers identified two factors of a social network that affect how information spreads: what they term “the burstiness of user login times” and content aging. Bursty logins refer to intervals of short and sudden episodes or groups. However, the recent-ness (freshness) of photos—those uploaded within 30 days—does play a role in the attention they get. The 30-day old photos in Flickr receive 35% of all the favorite responses in the week during the research (at just under 70,000 “favorites”); the remaining 65% of favorite markings go to older photos over