Chapter 11
Capturing the Gist(s) of Image Sets Associated with Chinese Cities through Related Tags Networks on Flickr®

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT
To introduce how related tags networks may be extracted from Flickr® and used for “gist” and other analysis, this chapter describes the related tag networks associated with some of the cities of the People’s Republic of China (used as seeding terms). The software used for the data extractions (from the Flickr® API) and the creation of various graph visualizations is the free and open-source Network Overview, Discovery and Exploration for Excel (NodeXL Basic), available on Microsoft’s CodePlex platform.

INTRODUCTION
Web 2.0 brought with it the broad-scale open-access sharing of digital contents on social media platforms. People shared digital imagery, videos, “pins,” slideshows, six-word memoirs, novels, digital learning objects, and other informational contents; they shared pithy thoughts in microblogging messages. People also shared messages with others, friended and unfriended them, interacted over the digital artifacts, tagged their digital contents as well as those of others, liked and favorited digital artifacts, and engaged socially. The phenomena of “consumer-generated media” (CGM) is thought to create new dynamics (Yanai, Kawakubo, & Barnard, 2012, p. 63). Their actions on the various platforms resulted not only in publicly accessible repositories of contents, but they also enabled access to floods of metadata and metainformation, raw data about data, and more processed information about information. Some researchers summarize this phenomenon:
Social media sites share four characteristics: (1) Users create or contribute content in a variety of media types; (2) Users annotate content with tags; (3) Users evaluate content, either actively by voting or passively by using content; and (4) Users create social networks by designating other users with similar interests as contacts or friends. In the process of using these sites, users are adding rich metadata in the form of social networks, annotations and ratings. Availability of large quantities of this metadata will lead to the development of new algorithms to solve a variety of information processing problems, from new recommendation to improved information discovery algorithms. (Lerman, Plangprasopchok, & Wong, 2007, p. 65)

A number of social media platforms have created application programming interfaces (APIs) to allow broad public access to both some of those informational contents and labeling metadata. These APIs may be accessed by automated programs used to scrape information, and they may be accessed by applications that are less complex to use than command-line based programs. Researchers, particularly in the fields of data mining, have worked to turn some of that metadata into information. One type of metadata of interest are folk-created tags, freeform, unstructured keywords used to describe or label images and videos. Collective tagging—by individuals working individually and in groups—lead to a common informal and non-hierarchical classification system, dubbed a “folksonomy.” This portmanteau term of “folk” and “taxonomy” was created by Thomas Vander Wal in 2004 (Vander Wal, “Folksonomy,” Feb. 2, 2007). Others refer to informal tagging as “lightweight annotation” or “lightweight semantic scaffolding.” If formal tags come in “sets” that are non-redundant, then informal tags come in “bags” that have duplicates and redundancies. Indeed, there are some patterned differences between informal tagging and formal indexing (Rorissa, 2010). Some structured data may be extracted from freeform tags for analysis (Rattenbury, Good, & Naaman, 2007). Informal tagging is of nominal expense to the service providers as compared to formal tagging, but the captured tags have limits and enable only some kinds of information capture.

Folksonomies are understood to be amateur created, and as such, the applied terms of not formally defined or hierarchically structured; they are non-exclusive. Such tags may be explored in the context of the particular described digital artifact, but they may also be collected into networks related to certain seeding tag terms, which may enable insights into collective perceptions of particular phenomena. One writer went so far as to describe the tagging feature on Flickr® as “feral” (as in in-the-wild and non-domesticated):

The most interesting—and the most feral—aspect of Flickr is the tagging. Instead of providing a set list of possible keywords, Flickr allows users to type in any tag they like. Each photo can have as many tags as desired. If Jane clicks on one of the tags on a photo her friend Nina took, Jane is shown all Nina’s photos that have that particular tag. From that page, she can continue by clicking the link titled “see all public photos tagged with [the tag]”. This gives some very interesting results. Since there are no pre-defined rules for how to tag your photos, nobody has complete control of the ways in which photos are presented, yet vast pools of photographs of specific places or events are gathered and made accessible. Different tags produce very different kinds of description, narrative or argument. (Walker, 2005, p. 4).

Related tags networks are concept-based abstractions that exist several steps out from the original taggers (individuals who provided the keyword labels) and the digital contents (the target images, slideshows, articles, videos, or tagged objects). Related tags networks are depicted visually in node-link