[DC-02-004] Social Media Platforms

Abstract

Social media is a group of interactive Web 2.0 Internet-based applications that allow users to create and exchange user-generated content via virtual communities. Social media platforms have a large user population who generate massive amounts of digital footprints, which are valuable data sources for observing and analyzing human activities/behavior. This entry focuses on social media platforms that provide spatial information in different forms for Geographic Information Systems and Technology (GIS&T) research. These social media platforms can be grouped into six categories: microblogging sites, social networking sites, content sharing sites, product and service review sites, collaborative knowledge sharing sites, and others. Four methods are available for capturing data from social media platforms, including Web Application Programming Interfaces (Web APIs), Web scraping, digital participant recruitment, and direct data purchasing. This entry first overviews the history, opportunities, and challenges related to social media platforms. Each category of social media platforms is then introduced in detail, including platform features, well-known platform examples, and data capturing processes.

Keywords: big data, data capture, Facebook, Flickr, geotag, social media, Twitter, Web API, Web Application Programming Interface, Web scraping

Author & citation

Gong, X. and Yang, X. (2020). Social Media Platforms. The Geographic Information Science & Technology Body of Knowledge (3rd Quarter 2020 Edition), John P. Wilson (ed.).
DOI: 10.22224/gistbok/2020.3.2.

Explanation

1. Definitions
2. The Emergence of Social Media Platforms
3. Social Media Categories
4. Data Capturing Methods
5. Opportunities and Challenges
6. Social Media Platforms Today

1. Definitions

Social media: Interactive Web 2.0 Internet-based applications that allow users to create and exchange user-generated content via virtual communities.

Web 2.0: The second generation of the World Wide Web which changes from static webpages to more interactive and user-generated content.

Web APIs: Defined interfaces to certain services over the web which can be accessed using the Hypertext Transfer Protocol (HTTP).

Rate limit: The number of allowed Web API requests within a given time period.

Web scraping: The process of automatically extracting and downloading information from websites.

Open Authorization (OAuth): An open-standard authorization protocol used by Internet users to grant websites access to their information on other websites without providing passwords.

Check-in: Self-reported locations on social media platforms to indicate where users have physically visited.

2. The Emergence of Social Media Platforms

Social media is a group of interactive Web 2.0 Internet-based applications that allow users to create and exchange user-generated content via virtual communities (Kaplan & Haenlein, 2010). Although social media may be rooted in the introduction of the telegraph in the 1840s, it started to gain popularity when the Internet proliferated with the World Wide Web in the mid-1990s. GeoCities appeared in 1994 as one of the Internet's earliest social media websites. Since the late 1990's, the Web 2.0 innovations have brought in a vast development of social media platforms including LinkedIn, MySpace, Facebook, YouTube, and Twitter, among others. By January 2020, there were over 3.8 billion active social media users worldwide (We Are Social & Hootsuite, 2020). For a large portion of users, visiting social media platforms has already become a daily routine. Researchers have also seen the convergence of Geographic Information Systems and Technology (GIS&T) and social media. On the one hand, GIS has been functioning more socially with increasing users who contribute to their own virtual community for exchanging geographic information on online mapping sites; on the other hand, social media have become increasingly equipped with mapping and location-based features (Sui & Goodchild, 2011).

3. Social Media Categories

Hundreds of social media platforms are (or have been) available on the Internet. This entry only focuses on platforms that contain spatial information in different forms (such as geotags, place names, user profiles, check-ins, and other georeferenced contents), which can be used in GIS&T research. These social media platforms can be categorized into microblogging sites, social networking sites, content sharing sites, products and service review sites, collaborative knowledge sharing sites, and others (Table 1). More details about each category are discussed in Section 6.

Table 1. Categories of Social Media Platforms for GIS&T Research

Category	Definition	Example (language, if not English)
Microblogging Sites	Microblogging sites restrict users to only post short messages for faster dissemination. The postings can include texts, images, videos, and links. Users can subscribe to, pass along, and reply to others' contents. Users can also create and/or share hashtags to discuss related topics.	Twitter, Sina Weibo (Chinese)

Category	Definition	Example (language, if not English)
Social Networking Sites	Social networking sites are online platforms that focus on building and visualizing social networks or social relations among people who share interests and/or activities. Each user has an individual profile. Users can post text, video, pictures, and links; they can also include location information into their profiles and shared contents.	Facebook, Foursquare, LinkedIn, MySpace, VK (Russian)
Content Sharing Sites	Content sharing sites allow users to share pictures or videos with the public or a restricted community. Users can upload, edit, organize, and comment on pictures and videos directly on the website. Users can also embed graphics, links, time, and location information in their shared contents.	Flickr, Instagram, Pinterest, YouTube, Tumblr, TikTok
Product and Service Review Sites	Product and service review sites provide general information about products and services for users to evaluate. Users can both read and write product/service reviews.	Yelp, TripAdvisor
Collaborative Knowledge Sharing Sites	Collaborative knowledge sharing sites are collectively written and maintained by a community of users. Any user can create and edit the site contents to share knowledge, and often a smaller group of users serve as moderators to oversee the content creation process.	Wikimapia, OpenStreetMap
Others	Other social media platforms containing spatial information that can be used in GIS&T research.	Tencent Product (Chinese), Airbnb, Craigslist, Anjuke (Chinese), WhatsApp, Skype, Telegram, Snapchat, etc.

4. Data Capturing Methods

Four methods are available for capturing data from social media platforms:

- **Web APIs**: A common practice among social media platforms is to offer Web APIs for data access, query, and/or update. Researchers can collect data by performing programmatically Web API requests in frequencies not exceeding the corresponding rate limits.
- **Web scraping**: Data on some social media platforms can be extracted through Web scraping. However, Web scraping could be against the platforms' terms of service and cause ethical and legal problems.
- **Digital participant recruitment**: Many social media platforms allow users to share their data across platforms through the OAuth. Therefore, researchers can digitally recruit participants at a massive scale on a platform to collect user data instantly.
- **Direct data purchasing**: Anonymized data from certain social media platforms are available for purchase. Researchers can directly contact these platforms for further purchasing details.

5. Opportunities and Challenges

Social media platforms have a large user population who generate massive amounts of digital footprints, which are valuable data sources for observing and analyzing human activities/behavior. The crowd-sourced social media data are available at a much lower cost, faster speed, and larger amount in comparison to traditional data collection methods

such as surveys and census (Rizwan et al., 2020). Additionally, many social media platforms support geo-referenced information submission and sharing. The high-volume, geo-referenced, and open-source social media data, as a type of emerging spatial big data, provide an unprecedented opportunity for uncovering the spatial-temporal patterns of human dynamics at a large scale. More details about the analysis and applications of social media data can be found in the BoK topic on Social Media Analytics.

However, challenges also exist due to the 5Vs (volume, velocity, variety, veracity, and value) and privacy issues of social media data.

- Social media data are generated and transmitted across the Internet in high velocity and volume. Traditional paradigms for small data processing (in-memory, non-real-time) do not meet these requirements. This calls for a different infrastructure including scalable distributed storage systems, efficient data transmission models, real-time stream-based processing techniques, etc. (C. Yang et al., 2017)
- Social media data contain a variety of formats, structures, and domains, which are usually incompatible with each other. New technologies are required to automatically clean, store, and integrate data of mixed structures, as well as to generate metadata.
- Social media users are unevenly distributed in space, time, and demographics. User-generated social media contents often consist of inaccurate, incomplete, or redundant data. The availability of social media data also depends on users' data-sharing settings and the black-box nature of its access APIs. In addition, as rumors and location spoofings have been noted on social media platforms, it poses unintended challenges to treat social media data as credible and accurate data source directly (Ye et al., 2020). Therefore, without awareness and proper handling of these veracity issues, social media analytics could have an increased likelihood of "false discoveries." To address some veracity challenges of social media data, automated methods have been reported to assess the credibility of shared content, to spot rumors and fake content, and to detect location spoofing (Papadopoulos et al., 2016). A more in-depth discussion of uncertainty in big geospatial data can be found in the BoK topic on Spatial Data Uncertainty.
- To actualize the value of social media data, we need to define meaningful research questions before analysis and convert results into useful knowledge or applications. The processes are not always straightforward, which might need domain knowledge and interdisciplinary collaborations.
- Social media users usually are not aware of whether, when, or how their digitally-shared information is used for research purposes (Yan et al., 2020). It poses privacy concerns over social media data and calls for new strategies to address privacy issues.

6. Social Media Platforms Today

6.1 Microblogging Sites

Microblogging is a variate of blogging where the pieces of content are extremely short. The fast-paced nature of the microblog makes it perfect for sharing breaking news and real-time updates about unfolding events. Examples of microblogging sites include Twitter and Sina Weibo.

Twitter is among the most well-known microblogging sites which enjoy extreme popularity in academic research in recent years. Twitter provides shorter messaging updates (up to 280 characters per tweet) for faster dissemination as well as directed-following, hashtags, and retweeting for better user interactions (Kwak et al., 2010). Twitter users may also opt to share location metadata (exact 'point' location or a 'bounding box') with every single tweet. These unique characteristics of Twitter attract many individual and institutional users (Gong & Lane, 2020). We can make requests to different Twitter APIs to access data in frequencies not exceeding the corresponding rate limits. Commonly used APIs include the "search APIs" and the "streaming APIs". The search APIs allows searching for tweets in the past seven days (longer time periods with advanced license) based on given filter criteria (keywords, locations, languages, etc.); and the streaming APIs keep an HTTP connection open and retrieve future tweets matching given filter criteria in real time. Other details about the Twitter APIs can be found in the BoK topic on GIS APIs.

6.2 Social Networking Sites

A social networking site is an online platform that focuses on building and visualizing social networks or social relations among people who share interests and/or activities. A social network service typically provides a representation for each user (often a profile), their social links, and a variety of additional services. Examples of social networking sites include Facebook, Foursquare, Linkedin, MySpace, and VK. The two social networking sites that have been used most frequently in GIS&T research are Facebook and Foursquare.

Facebook was first launched in 2004 and has become a world-leading social networking platform ever since. There were 1.66 billion daily active users on Facebook by the end of 2019. Each registered user on Facebook has a personal profile (referred to as "Timeline") that shows a chronological view of the user's stories including status updates, photos, interactions with apps, and events. Users can opt to include spatial information (e.g. hometown, work history, and education history) in their introductions and disclose their locations via check-ins in their posts on the Timeline. Organizations can also host public pages on Facebook to share their information with the public, some are found to be location-based (Bird et al., 2012). The primary way to capture Facebook data is by using its Graph API to programmatically query data. Researchers can make requests to the Graph API to retrieve individual user's introduction and posts on the Timeline, in which user-supplied addresses and check-in information can be used in GIS&T research. Researchers can also use the Graph API to collect data from Facebook public pages, which might include location-based information that is publicly visible. Additionally, connections among users through Facebook friendships can be retrieved using the Graph API for network analysis (see the BoK topic on Social Networks for more details).

The increasing availability of location-enabled mobile devices (see the BoK topic on Mobile Devices) makes location-based social networks (LBSNs) more accessible to mobile users. Foursquare was created in 2009 and has quickly risen as the most popular LBSN with more than 55 million active users and 3 billion global visits monthly as of September 2019. Foursquare enables its users to share location information with their friends through check-ins, rating and reviewing venues they visited, and reading other users' reviews (Scellato et al., 2011). Foursquare data can be acquired by making requests to its Places API programmatically. However, this method requires user authorization to collect personal information and has rate limits set in place. Many Foursquare users choose to automatically push their check-in messages to Twitter. Therefore, researchers can first acquire the well-

formatted Foursquare check-in tweets using the Twitter APIs, then extract points of interest (POI) names from the tweets by applying handcrafted rules with regular expressions (X. Yang et al., 2016).

6.3 Content Sharing Sites

Content sharing sites, such as Flickr, Instagram, YouTube, and TikTok, allow users to share pictures or videos with the public or a restricted community. Users can upload, edit, organize, and comment on pictures and videos directly on the site. Users can also embed graphics, links, time, and location information in the contents they shared. The relative higher geotagging rate and easier location extraction process compared to other platforms makes Flicker the most commonly used content sharing site in GIS&T research.

Flickr is intended to help people organize and share their images and videos. It is widely used by amateur and professional photographers. Flickr officially claimed it has 90 million active monthly users in 2019. Flickr provides a geotagging function for users to attach geographic coordinates to their shared photos. The crowd-sourced geotagged photos are indeed a suitable proxy for the empirical information about where people go (Wood et al., 2013). To retrieve Flickr photos and associated metadata (including geotags, photo titles, and user-provided tags), researchers can make a Flickr API request by specifying a rectangular boundary box containing the study area (Ghermandi et al., 2020).

6.4 Product and Service Review Sites

Product and service review sites, such as Yelp and TripAdvisor, provide a social media platform for users to navigate product/service information, read product/service reviews created by other users, and contribute their own reviews. The products and services usually have associated location information; and users can also opt to include their locations in their reviews. Yelp is one of the most commonly used product and service review sites in GIS&T research.

Yelp publishes crowd-sourced reviews about businesses, such as restaurants, bars, and beauty salons. It included over 184 million reviews worldwide visited by over 178 million unique users monthly as of September 2019. Yelp offers a suite of Web APIs (Yelp Fusion API) for searching and interacting with data about local businesses. Researchers can use the search APIs to look for detailed business information and review counts by providing location, keyword, category, etc. Researchers can also retrieve limited reviews (up to three) about a given business by calling the business review API with the ID of the business. More reviews can be retrieved from the Yelp Open Dataset, which is a subset of Yelp's data for use in personal, educational, and academic purposes. The dataset includes business information, user profiles, reviews, check-ins, and pictures (Rahimi et al., 2018). Yelp also updates and enriches the dataset from time to time (as of April 2020, the dataset covered over 8 million reviews for more than 200,000 businesses).

6.5 Collaborative Knowledge Sharing Sites

Collaborating knowledge sharing sites, such as Wikis and OpenStreetMap, provide collaborative social media platforms for people to exchange information and generate knowledge. The geospatial information generated by non-professional volunteers from these sites has been coined as volunteered geographic information or VGI (Goodchild, 2007), which could be used in scientific projects (GS-24 (Citizen Science)).

The OpenStreetMap (OSM) project was founded in 2004 aiming at creating a world map freely available to anyone. The OSM is one of the most commonly cited collaborative knowledge sharing sites. It attracted 1.4 million different user contributors to archive 3.5 million map changes per day as of February 2020. There are several different methods of OSM data retrieval (Mooney & Minghini, 2017). For a small amount of OSM data, users can select a small region of the map and use the “export” feature to download the data; users can also make OpenStreetMap API requests to retrieve map data within a given bounding rectangle. For continental-, national-, and regional-sized data, data providers, such as Geofabrik, provide frequent updates and free data downloading services in different file formats. The OSM also provides access to the Planet.osm file, which is the entire OSM database contained in one very large Extensible Markup Language (XML) file or a compressed file.

6.6 Others

Besides the abovementioned five categories, there are other social media platforms containing spatial information that can be used in GIS&T research. Some examples are as follows:

- Tencent is one of the largest social media corporations in China, whose social media platforms include WeChat (a mobile chatting service), Tencent QQ (an instant messenger software), and other location-based services. Tencent had more than 1 billion users as of 2019. It provides real-time Tencent User Density (TUD) information by mapping locations of active smartphone users who are using its products. Due to its large user base, the TUD data could provide a representative depiction of human dynamics (Li et al., 2020).
- On-line housing rental websites (OHRWs) are social media platforms that are primarily used by landlords and renters. Popular OHRWs include Airbnb, Craigslist, and Anjuke. OHRW users can post and review house rental information, which often includes rich spatial data. The OHRW data can be extracted to map and monitor the fine-scale patterns of housing rental price dynamics (Hu et al., 2019).
- Social messaging apps (such as WhatsApp, Skype, Telegram, and Snapchat) might also implicitly contain geographic information in the huge amount of multimedia communications. Geographic information could be extracted from these social media platforms using text information extraction and text mining approaches (Stock, 2018).

References

Bird, D., Ling, M., & Haynes, K. (2012). Flooding Facebook - the use of social media during the Queensland and Victorian floods. Australian Journal of Emergency Management, 27(1), 27–33.

Ghermandi, A., Camacho-Valdez, V., & Trejo-Espinosa, H. (2020). Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. Tourism Management, 77(July 2019), 1–9.

Gong, X., & Lane, K. M. D. (2020). Institutional Twitter usage among U.S. geography departments. Professional Geographer, 72(2), 219–237.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography.

GeoJournal, 69(4), 211-221.

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. Land Use Policy, 82(129), 657–673. DOI:10.1016/j.landusepol.2018.12.030

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59–68.

Kemp, S. (2020). Digital 2020 Global Overview Report. We Are Social & Hootsuite, in the Datareportal.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In M. Rappa, P. Jones, J. Freire, & S. Chakrabarti (Eds.), Proceedings of the 19th International Conference on World Wide Web (pp. 591–600). ACM.

Li, S., Lyu, D., Huang, G., Zhang, X., Gao, F., Chen, Y., & Liu, X. (2020). Spatially varying impacts of built environment factors on rail transit ridership at station level: A case study in Guangzhou, China. Journal of Transport Geography, 82(July 2019), 1–14.

Mooney, P., & Minghini, M. (2017). A review of OpenStreetMap data. In G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, & V. Antoniou (Eds.), Mapping and the Citizen Sensor (pp. 37–59). Ubiquity Press.

Papadopoulos, S., Bontcheva, K., Jaho, E., Lupu, M., & Castillo, C. (2016). Overview of the Special Issue on Trust and Veracity of Information in Social Media. ACM Transactions on Information Systems, 34(3), 1–5.

Rahimi, S., Mottahedi, S., & Liu, X. (2018). The geography of taste: Using Yelp to study urban culture. ISPRS International Journal of Geo-Information, 7(9), 1–25.

Rizwan, M., Wan, W., & Gwiazdzinski, L. (2020). Visualization, spatiotemporal patterns, and directional analysis of urban activities using geolocation data extracted from LBSN. ISPRS International Journal of Geo-Information, 9(2).

Scellato, S., Noulas, A., Lambiotte, R., & Mascolo, C. (2011). Socio-spatial properties of online location-based social networks. In L. A. Adamic, R. Baeza-Yates, & S. Counts (Eds.), Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (Vol. 11, pp. 329–336). The AAAI Press.

Stock, K. (2018). Mining location from social media: A systematic review. Computers, Environment and Urban Systems, 71, 209–240.

Sui, D. Z., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. International Journal of Geographical Information Science, 25(11), 1737–1748.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify

nature-based tourism and recreation. Scientific Reports, 3(2976), 1–7.

Yan, Y., Feng, C. C., Huang, W., Fan, H., Wang, Y. C., & Zipf, A. (2020). Volunteered geographic information research in the first decade: A narrative review of selected journal articles in GIScience. International Journal of Geographical Information Science.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth, 10(1), 13-53.

Yang, X., Ye, X., & Sui, D. Z. (2016). We Know Where You Are: In Space and Place - Enriching the Geographical Context through Social Media. International Journal of Applied Geospatial Research, 7(2), 61–75.

Ye, X., Zhao, B., Nguyen, T.H., Wang, S. (2020). Social Media and Social Awareness. In: Guo, H., Goodchild, M.F., Annoni, A. (eds) Manual of Digital Earth. Springer, Singapore.