Geo-Social Media Analytics

Cheng-Te Li
Research Center for Information Technology Innovation,
Academia Sinica, Taipei, Taiwan
ctli@citi.sinica.edu.tw

Hsun-Ping Hsieh
Graduate Institute of Networking and Multimedia,
National Taiwan University, Taipei, Taiwan
d98944006@csie.edu.tw

Categories and Subject Descriptors
H.2.8 [Database Management]: Database Applications – Data Mining

General Terms
Algorithms, Experiments

Keywords
Social network, social media, geo-social media, link prediction, location recommendation, influence propagation, online, offline

1. INTRODUCTION
With the maturity of wireless communication techniques, GPS-equipped mobile devices (e.g. mobile phones and tablets) become ubiquitous, and location-acquisition technologies and services are flourishing. These location applications as well as mobile devices, developed and combined with the social networking services, foster the emergence of geo-social media, a novel type of user-generated spatio-social data. The typical services of geo-social media include Facebook, Twitter, and Foursquare. In geo-social media, social connections and geo-location information of users are the essential elements, which keep track of their user interactions and their spatio-temporal activities. While social interactions are depicted by online network structures, and geographical activities are usually represented as check-in records, which consist of sequences of data points with latitude-longitude records, time stamps, and venue information. Due to the pervasive mobility of users that leads to their ubiquitous social interactions, a huge amount of user-generated geo-social data is rapidly generated and accumulated. Such big geo-social data not only collectively represents diverse kinds of real-world human activities, but also serves as a handy resource for various geo-social applications. Therefore, we think having this tutorial is timely, useful, and practical for both academia and industries to systematically know the recent advances on geo-social media analytics. What follows provides a description about the organization of topics for our tutorial.

We will introduce the background, motivation, challenges of geo-social media analytics, followed by an over-view of data sources of geo-social media. We systematically elaborate the essential elements that composing geo-social media data: social connections, geographical information, temporal factors, and interaction contents. Then we will present various definitions of geo-social networks with corresponding examples in different geo-social media services. The main content of this tutorial consists of the following five topics.

(1) Properties of Geo-Social Networks [4, 6, 16, 17, 18]**.** Network properties, such as high clustering coefficient, low average path length, and power-law degree distribution, have been well studied by conventional SNA. With location information revealed by geo-social media, some researchers have shifted to investigate the network properties under both geographical and social contexts. This topic consists of three parts. First, to understand geo-social networks, from a series of recent studies, we will introduce various relationships between (a) geographical distance and conventional network properties, (b) geographical distance and social position (e.g. measured by *social overlap* and *k-index*), and (c) human mobility and social structures. Second, we will present a series of new proposed socio-spatial measures, including *node locality*, *geographic clustering coefficient*, *average friend distance*, *distance strength*, *triangle link length*, etc. to characterize the properties of geo-social networks. Third, to measure the node importance in geo-social networks, we introduce several geo-social metrics, including *spatial degree centrality*, *spatial closeness centrality*, *spatial efficiency*, and *local spatial clustering coefficient*.

(2) Geo-Social Link Prediction [15, 16, 19, 20, 21]**.** Link prediction is the essential task for friend recommendation in online social networks. Both structural and content features that depict common friends and capture shared interests respectively, which are extracted from online social interactions, have been validated to be effective for predicting future links. In geo-social services, the *check-in* function further allows users to specify *where* they locate, *who* they are with, and *when* they do such actions. Therefore, geographical and temporal information hidden in check-in records are available to be exploited for link prediction. This part aims to introduce how to leverage check-in data collected by geo-social media to boost the performance of link prediction. From recent literatures, we categorize check-in features into three types: (a) *geographical features* that characterize the spatial and temporal frequency and distribution of check-ins at the same or neighboring places between two users, (b) *human mobility features* that models the temporal and geographic *periodicity* of user movements, and (c) *venue features* that capture the place information (such as category, activity, and popularity). Based on these features, we will further elaborate a series of spatio-social link prediction models.

(3) Location Recommendation [9, 10, 13, 22, 23]**.** Location recommendation is to recommend new locations that users have never visited before. The general considerations of recommendation include the current geo-location, time, user preference, and social interactions. We aim to review a series of recent studies on location recommendation using geo-social media data, including Foursquare, Facebook Place, Gowalla, and Yelp. The approaches to recommend locations can be divided into three categories: *geo-social link modeling*, and *geo-social content modeling*, and *hybrid modeling*. Social link modeling aims to recommend either travel companions or venues using the connections between users, between locations, and the interaction links between users and locations. Social content modeling is to extract and model contextual information, including preferences, opinions, sentiments, and attractiveness, from the user-interaction data (e.g. tips in Foursquare and tweets in Twitter), to describe the preference on each venue, and recommend locations such that the user satisfaction can be maximized. Hybrid modeling combines both geo-social links and contents to seek a balance or trade-off among geographical influence, social factors, and user satisfaction.

WWW'15 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2741985

(4) Geo-Social Influence Propagation and Maximization [2, 8, 11, 12]**.** The mechanisms of *influence propagation* characterize the behaviors of how people adopt a new action because of their friends and how certain information spread over individuals in social networks. The goal of *influence maximization* (IM) is to select a small set of individuals in the network such that the expected number of activated ones is maximized. While conventional IM studies mainly target at pure graph structures, rich location information carried by individuals gain much less attentions currently. However, many real-world applications require location-aware influence maximization. For example, local business which creates new campaigns may want to the potential customers within particular geographical area for the promotion and marketing. In disaster management, the local government agency needs to find regionally influential individuals such that the help and support can be widely distributed. In this part, we aim to present recent advances of influence propagation and maximization in the presence of geo-social media. We elaborate various definitions of geo-social influence maximization, different geo-social diffusion models, and greedy and heuristic solutions to geo-social IM.

(5) Connecting Online and Offline Social Networks [1, 3, 5, 8, 14, 15, 24]**.** In geo-social media, users are allowed to specify their social connections in their friend lists, which can be regarded as a kind of *online* social network. Geo-social media also provide convenient platforms for people to create, distribute, and organize *social events*, ranging from informal gathering (e.g. dining out and seeing movie) to formal activities (e.g. business meeting and technical conferences). People participating events are supposed to show up and have *offline* face-to-face interactions with each other in the physical world. That says, offline social networks can be constructed through social events. While users in geo-social media are interweaved within both online and offline social networks, one may think are there any characteristics of offline social networks that differ online, how to connect online and offline social worlds together to revisit convention SNA, such as community detection and information diffusion, and whether or not online and offline networks co-evolve over time together. This part will uncover the connections and differences between online and offline networks, revisit SNA tasks under online-offline contexts, and pointing out potential applications inspired from bridging virtual and physical social worlds.

In the end of this tutorial, we will highlight the unsolved problems for each of the abovementioned topics and some future directions of geo-social media analytics. In addition, we will also provide a list of tools and packages for analyzing geo-social media.

2. AUDIENCE

We expect to attract the audiences from several research communities: (a) *mobile web*: understanding and exploiting user behaviors from their historical interactions in geo-social media, (b) *data mining*: mining the complex data containing social graphs of users, spatial deployment of locations, and time-evolving information flows, (c) *information retrieval*: integrating geo-social media data with location information sensored by mobile devices to provide geo-social search engine, (d) *social network analysis*: revisiting the conventional SNA tasks considering both social connections and moving behaviors of users, (e) *recommender system*: item recommendation with geographical and social factors, (f) *marketing*: devising the marketing strategies based on how users move and interact with each other in the city, and (g) *advertisement*: placing the ads on mobile devices according to the visiting preferences of users.

3. REFERENCES

[1] L. A. Adamic, T. M. Lento, A. T. Fiore. How You Met Me. ICWSM 2012

[2] P. Bouros, D. Sacharidis, N. Bikakis. Regionally Influential Users in Location-Aware Social Networks. SIGSPATIAL/GIS 2014.

[3] C. Brown, V. Nicosia, S. Scellato, A. Noulas, C. Mascolo. Where Online Friends Meet: Social Communities in Location-Based Networks. ICWSM 2012.

[4] E. Cho, S. A. Myers, J. Leskovec. Friendship and mobility: user movement in location-based social networks. KDD 2011.

[5] K. Feng, G. Cong, S. S. Bhowmick, S. Ma. In search of influential event organizers in online social networks. SIGMOD 2014.

[6] H. Gao, J. Tang, H. Liu. gSCorr: modeling geo-social correlations for new check-ins on location-based social networks. CIKM 2012.

[7] M. Gomez-Rodriguez, M. Rogati. Bridging offline and online social graph dynamics. CIKM 2012.

[8] D. Hristova, M. Musolesi, C. Mascolo. Keep Your Friends Close and Your Facebook Friends Closer: A Multiplex Network Approach to the Analysis of Offline and Online Social Ties. ICWSM 2014.

[9] H.-P. Hsieh, C.-T. Li, S.-D. Lin. Measuring and Recommending Time-Sensitive Routes from Location-Based Data. ACM TIST 2014.

[10] H.-P. Hsieh, C.-T. Li. Mining and Planning Time-aware Routes from Check-in Data. CIKM 2014.

[11] K. Y. Kamath, J. Caverlee, Z. Cheng, D. Z. Sui. Spatial influence vs. community influence: modeling the global spread of social media. CIKM 201.

[12] G. Li, S. Chen, J. Feng, K.-L. Tan, W.-S. Li. Efficient location-aware influence maximization. SIGMOD 2014.

[13] C.-T. Li, H.-P. Hsieh. MobiCrowd: simulating crowds with periodic and social mobility. AAMAS 2014.

[14] K. Li, W. Lu, S. Bhagat, L. V. S. Lakshmanan, C. Yu. On social event organization. KDD 2014.

[15] X. Liu, Q. He, Y. Tian, W.-C. Lee, J. McPherson, J. Han. Event-based social networks: linking the online and offline social worlds. KDD 2012.

[16] A. Sadilek, H. A. Kautz, J. P. Bigham: Finding your friends and following them to where you are. WSDM 2012.

[17] S. Scellato, A. Noulas, R. Lambiotte, C. Mascolo. Socio-Spatial Properties of Online Location-Based Social Networks. ICWSM 2011.

[18] S. Scellato, C. Mascolo, M. Musolesi, V. Latora. Distance Matters: Geo-social Metrics for Online Social Networks. WOSN 2010.

[19] S. Scellato, A. Noulas, C. Mascolo. Exploiting place features in link prediction on location-based social networks. KDD 2011.

[20] C. Scholz, M. Atzmueller, A. Barrat, C. Cattuto, G. Stumme. New Insights and Methods for Predicting Face-To-Face Contacts. ICWSM 2013.

[21] D. Wang, D. Pedreschi, C. Song, F. Giannotti, A.-L. Barabási. Human mobility, social ties, and link prediction. KDD 2011.

[22] M. Ye, D. Shou, W.-C. Lee, P. Yin, K. Janowicz. On the semantic annotation of places in location-based social networks. KDD 2011.

[23] M. Ye, P. Yin, W.-C. Lee, D. L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. SIGIR 2011.

[24] P. Yin, Q. He, X. Liu, W.-C. Lee. It Takes Two to Tango: Exploring Social Tie Development with Both Online and Offline Interactions. SDM 2014.