A Review Paper On Exploring Text, Link And Spacial-Temporal Information In Social Media Networks

Dr. Mamta Madan, Meenu Chopra, Vani Nijhawan

ABSTRACT: The objective of this paper is to have a literature review on the various methods to mine the knowledge from the social media by taking advantage of embedded heterogeneous information. Specifically, we are trying to review different types of mining framework which provides us useful information from these networks that have heterogeneous data types including text, spacial-temporal and data association (LINK) information. Firstly, we will discuss the link mining to study the link structure with respect to Social Media (SM). Secondly, we summarize the various text mining models, thirdly we shall review spacial as well the temporal models to extract or detect the frequent related topics from SM. Fourthly; we will try to figure out few improvised models that take advantage of the link, textual, temporal and spacial information which motivates to discover progressive principles and fresh methodologies for DM (Data Mining) in social media networks (SMNs).

Index Terms: Social Media Networks (SMNs), Social Media (SM), Text Mining (TM), Link Mining (LM).

1. Introduction

With the vast, phenomenal success of SMNs like Facebook, Flickr, Twitter and LinkedIn, has drastically changed the thinking as well as communication way of the masses. In contrast to earlier datasets, today SM has not only huge amount of data but also it in heterogeneous which include Data association (The Link), text, and spacial information. Our aim in this paper is to review various methods to extract the knowledge from SM which has heterogeneous format of data and has three dimensions (The Link, Text and Spacial-Temporal).

2. Exploring associations through The Link Mining (LM) or Link Prediction (LP) in SM

Getoor et al. [1], classified the Link Mining (LM) tasks into three categories: Graph related tasks (for e.g. Generative and Classification Models or Sub graph theory), Object related tasks (for e.g. Classification, Ranking, Identification, Cluster Analysis) and Link related tasks (for e.g. Type, Strength or Cardinality). Our main focus is on the related LM tasks including Link prediction (LP) and Group Discovery (GD). The Link is the important terminology underlying the SMNs. The famous social networking site (SNS) Facebook consists of the Links as directionless relationships, where as Twitter utilize the follower-followee directed relationships, and LinkedIn projects the colleagues and classmates relationships. Not only the data association (The link) presents the pair-wise links but also point towards the social and community behavior. The following are the various methods which can be used for the Link Mining (LM) or Link Prediction (LP) in Social Media (SMNs) Networks.

- *Dr. Mamta Madan is working as Professor in Vivekananda Institute of Professional Studies, India. E-mail: mamta.vips@gmail.com*
- *Ms. Meenu Chopra is working as Assist. Professor in Vivekananda Institute of Professional Studies, India. Email: meenu.mehta@rediffmail.com*

2.1 Uniform-Node:

The basic idea underlying in these methods is to calculate an approximate distance between two objects. . Kashima et al. [3] estimated information from the node for the LP and used label configuration over the paired network nodes with many link types and analyze the relationships among the nodes. Researcher Debnath et al [4] calculated the weighted values associated with the nodes (in a social network graph), from the equations of linear regression that includes human assessment about the uniformity of nodes.

2.2 Geographical:

This method focus on the exploring either global or local patterns that describes the social network graph. Chen et al. [5] describes a data clustering algorithm for K-destination on directed graphs using random hitting time. In [6], authors suggested the use of unplanned-walk. Researchers in [7], discussed about unseen bi-grams and low rank approximation meta-approaches for LP.

2.3 Probabilistic:

These methods work on the concept of correlations that exists between nodes on social graph [8,9]. Link Prediction is the problem of predicting the existence of a link between two nodes in a relationship in a social media graph, where prediction is based on the properties of the objects and other observed links. LP has been studied on various kinds of graphs like metabolic pathways, protein-protein interaction, social networks, etc. Various researches use different measures such as user-wise similarity and location-based similarity to predict the existence of the links. These existing measures use different models that have been analyzed for the LP tasks which include relational Bayesian networks and relational Markov networks. LR in social media is closely related to LP, but has its own specific attributes. Social media networks can be considered as a graph where each node has its own properties. Linked nodes share certain similarities with respect to property information linked with entities and structure information linked with the graph. Where some others researchers have combined the above mentioned techniques, Popescul and Ungar [12] had suggested the relational statistical learning models for the Link Prediction. In Rattigan and Jensen [10], gave the demonstration of the effectiveness of LP models to give the solution of abnormal link discovery. Madadhain et al. [11] has given learned classifiers like naïve Bayes and logistic regression for predicting temporal link using network as well as object features. Group discovery (GD) a.k.a Group Detection, is to segregate the network vertices or nodes into densely connected sub graphs [2], which is a crucial milestone in creation of datasets in any of the network like social, web graphs, co-authorship, biological, etc. In [2] Newman et al. projected an algorithm to divide the network into groups by removing edges recursively. Tang et al. [13], using network structures, gave the appropriate review of group detection algorithms.

3. Exploring associations through Text Mining (TM) in SM

Online 'Big data' is everywhere and is in the format of text. The web users use the text to tag objects, to comment, to update their status or share something interesting with others. The following are the various models for text mining in social media which are used to discover the relevant topic structure from the stream of documents that to with the timestamps, which is a basic problem in TM.

3.1 Using Links Topic Modeling:

These models are called as probable models used to discover the meaningful structural traits of the document based on the notion of Bayesian analysis. For example, the traditional models which represent the model of semantic structure of the text collected. Using links many others models were proposed [14] like NetPLSA, Author-Topic model, Group-Topic Models, Block-LDA, Author-Recipient-Topic ,Topic-Link LDA and Topics-on-Participants model[15,16].

3.2 Spacial Topic Modeling:

This method focus on the exploring either global or local patterns that describes the social network graph. Chen et al. [17] describes a data clustering algorithm for K-destination on directed graphs using random hitting time. In [5], authors suggested the use of unplanned-walk. Researchers in [7], discussed about unseen bi-grams and low rank approximation meta-approaches for LP.

3.3 Temporal Topic Modeling:

In this type of models mining of the documents captures the topic structure associated with timestamps. For example some researchers have proposed busty topic patterns from text document, some have discovered state space models, and Iwata et al. [18] has given an online model with time evolution factor. All the research work on the temporal topic relies on the topics of the evolutionary pattern.

3.4 Event Detection and Tracking;

In [19], Fung et al. discovered Time-Based Documents –partition model to build a property –based event hierarchy for a text document. Some proposed the difficulties of event detection and tracking, others have given a probabilistic model to inculcate both text and time information in a single framework to discover news events. In [9], He et al. used the ideas from physics to model busty patterns and at last but not least; Becker et al. [20] exploited various techniques for the SM documents to detect events. Others researchers like Sakaki et al. [21], proposed an algorithm for real time events such as earthquakes, Yang et al. [22],explored online content to detect temporal patterns and Leskovec et al. [23], proposed an approach to generate a news cycle representation using daily rhythms in news media.

4. Exploring associations through Spacial and Temporal Information in SM

This kind of information is generally embedded in Social media for example some Social networking sites record the timestamps of online registered users, their actions and fetch the relevant information like geographical or IP address through various mobile applications. This kind of spacial-temporal information can help us to detect and analyze the online user community behavior. Some researchers proposed a model known as Web-a-Where for relating topography with online pages while others discovered a Scale-structure recognition method to retrieve the event and place connotation from Flickr tags based on the region or position and time metadata. Crandall et al. [24] predicted the areas for the online pictures on Flickr from the visual, temporal and textual features. Serdyukov et al. [25] discovered the topological locations for the pictures uploaded on Flickr by a language model on the end-user comments. They elaborated the language model by tag-based smoothing and cell-based smoothing. Instead of mining the geographical information for Flickr images, the blog is also a valuable source to fetch landmarks. Silva et al. [26] proposed a system for retrieving multimedia travel stories by using region metadata. Popescu et al. [27] displayed how to get clean tour related information from Flickr metadata. They extracted the geographical names from Wikipedia and generated the tour by mapping the picture tags to geographical names. In [28, 29], Choudhury et al. designed tour planning as directed orienteering problem. In [30], Lu et al. used dynamic programming for tour planning. In [62], Kennedy et al. used location, tags and visual features of the pictures to produce different and representative pictures for the landmarks.

5. Few Improvised Models for Mining in SM

All the above mentioned methods integrate the link, text and spatial & temporal information in social media from different perspectives and provide new principles and novel methodologies for data mining in social media using heterogeneous dimensions. There are a few detailed or improvised studies for social media mining, which researchers have found out are summarized as follows.

Link Recommendation (LR)

Online social networking sites (SNS) such as Facebook, Twitter, and LinkedIn are drawing much more curiosity than ever before. The users not only use the social media networks to stay connected with old friends, but also use the sites to search new friends with common interests and for business networking. Since the only thing i.e. the Link among people is the underlying key concept for online social networks, it is not surprising that LR [31, 32] is an essential link mining (LM) task. The following are the advantages for the Link Recommendation.

1. LR can help users to search the number of potential friends, a task that improves user experience in social media networks (SMN) and attracts more users subsequently.
2. LR helps the SMNs to expand fast in terms of the social linkage. A complete social graph not only improves user

involvement, but also provides the financial benefits linked with a number of users such as publisher network for advertisements.

Link recommendation in social network is closely related to link prediction, but has its own specific properties. Social network can be considered as a graph where each node has its own attributes. Linked entities share certain similarities with respect to attribute information associated with entities and structure information associated with the graph. In future we can study the problem of expressing the link relevance to incorporate both attributes and structure in a unified and intuitive manner.

5.2 Dormant Topological Topic Analysis (DTTA)

With the development of web 2.0 the GPS chips and smart phones, geographical records have become popular. A geographical record is defined as a two dimensional vector, latitude and longitude, depicting a unique location on the Earth. There are different popular ways like GPS receivers, some applications like Google Earth, Smart phones having GPS etc. to obtain topological data on the Web. In all these application or tools, GPS data is provided together with different documents including tags, user posts, etc., those documents with GPS records are defined as GPS-linked documents. The amount of GPS-linked documents is increasing exponentially. For example, social network media, Flickr hosts more than 100 million photos linked with tags and GPS regions. Mostly, GPS-linked documents make it possible to figure out the topological traits of various subjects. The topological traits of the hot or new topics call for effective approaches to study the GPS-linked documents on the Web. In recent years, some studies have been conducted on GPS-linked documents like managing geo-tagged photos [24] and finding large topological datasets [61]. However, none of them addressed the following two needs in analyzing GPS-linked documents.

- Discovering different topics of interests those are coherent in topological regions.
- Evaluating many topics across different topological locations.

There are three different approaches for topological topic discovery which are as follows:

- Location-driven model: In this, cluster GPS-linked documents are based on their regions and each document cluster considered as one topic. This model works well in existing apparent location clusters.
- Text-driven model: Which finds topics based on topic modeling with regularization by spatial information.
- Dormant Topological Topic Analysis (DTTA): In this we explored the location-text joint model for both topic discovery and comparison called as DTTA (Dormant Topological Topic Analysis), which combines and unified topological clustering and topic modeling into one single framework. Not only we can fetch the topological topics of good quality, but also able to calculate the topic distribution in different topological locations for topic comparison.

5.3 Dormant Frequent Topic Analysis (DFTA)

Frequent phenomena exist ubiquitously in our real-world, and many of natural and social topics have frequent re-occurring patterns for example music and film festivals, Hurricanes strike, Sales offered, annual reports, TV programs etc. occur in the similar seasons annually. Due to the prevalent existence of periodic occurring topics, frequently analysis is crucial in the real world. Generally, users can not only analyze the human behavior and the natural phenomena, but also predict the future trends which help them in decision making from the discovered frequent patterns. Presently with the development and ease of the Web, many textual data is present with timely information, e.g., photos tagged with their taken dates and time in Flickr, news articles with publishing dates and in Twitter users published tweets along with their upload times. An enormous amount of useful information is embedded in these textual data, and it is interesting to analyze and discover topics that are frequent and having their temporal patterns. Many researchers' had worked on the frequently detection for the database based on time series. Some studies worked on the strategies to speculate the time distribution of the query to detect frequent patterns of a single tag [33]. However, most of the studies are restricted to only time series database as they cannot be applied on textual data directly.

5.4 Dormant Group Topic Analysis (DGTA)

DGTA is a classical approach for the text mining which is to explore the unknown topics that exits in a document collection. The known examples of DGTA, such as PLSA, LDA and their mutants [34], use a multinomial word distribution to represent a meaningful coherent topic and model the generation of the text collection with a mixture of such topics. With huge and enormous amount of text content online, it is cumbersome for us to read and digest all the information from all the documents. DGTA not only provides an effective approach to help understand these large amounts of information but also helps in discovering the topics which are also useful to organize. To discover the group-based topics in text-associated graphs, we have to do the following three tasks.

- We would like to discover the group composition in the graph, so we can know the relationships among different users. To identified groups for not only to provides summarization of network structure and help understand the graphs, but also are crucial to analyze user behaviors in the setting of social media networks.
- We would like to discover the dormant or frequent topics in text-associated graphs. In this way we can know the interests of the users in the social media graph.
- We would like to learn the relationship between groups and topics, so we can know which groups are interested in a specific topic or which topics a specific group cares about.

With the development of social networks, discovering groups in graphs draws much more attention than before [35]. A group in a social media network is considered as a group of nodes with have more interactions and common topics among its members than between its members and others, and group discovery is the process to group the nodes into the clusters of close proximity in terms of interaction and common interests. To discover groups in graphs, typically an objective function is chosen to gets the intuition of a group as a set of nodes with better internal connectivity than external connectivity based on link structure [36]. However, if we only use link to discover groups, we cannot derive the coherence of common interests inside groups. A good group should be coherent in interaction patterns as well as shared topics. To find the group-based topics in text-associated graphs, we have to follow the previous text mining studies. Instead of modeling topics by considering pair-wise link relationships, we consider topic modeling in the group level. We assume that one group can correspond to multiple topics and multiple groups can share the same topic. For example, in a network one group can be interested in both sports and education topics, while multiple groups can be interested in a sports topic. The analysis of topics and groups could benefit each other.

5.5 Ranking Based on Pattern Trajectory

Social media is becoming increasingly popular with the development of Web 2.0 which includes websites such as Flickr, Facebook, and YouTube which further host enormous amounts of photos and videos. These images or videos in media sharing community, image or video files are contributed, tagged, and commented by all online users with the extra information such as topological information captured by GPS devices. Our aim is to analyze the common knowledge in photo sharing community for that we studied millions of personal photos in Flickr, which are associated with user tags and geographical information. The topological information is stored by low-cost GPS chips in cameras and mobile phones, and generally saved in the header of image files. The millions of photos in Flickr which contains user tags and topological information, which could be used to match tourist interests in terms of regions or countries or locations [27, 37], we analyzed that the trajectory patterns having two kinds of users. First, some users are interested in the most important trajectory patterns (follow those trajectories of popular locations). Second, some users are interested in discovering a new location in different ways. Instead of focusing on how to mine frequent trajectory a pattern, our focus is to explore that how ranking takes place in the discovered trajectory patterns and differentiate the ranking results. Ranking based on trajectory pattern helps the first kind of users who are interested in top important trajectories, while diversification helps the second kind of users that are willing to explore the diverse routes. Like some studies on trip planning using Flickr. For example Choudhury et al. [28,29] designed trip planning as directed orienteering problem,[30], Lu et al. explored dynamic programming for trip planning and Kurashima et al. [38] recommended tourist routes by combining topic models and Markov model. Trajectory mining has been investigated in many datasets including human travel history [39], animal movement, urban traffic, hurricane tracking etc. Some developed a spatial-temporal pattern mining paradigm that discovers trajectory patterns. We have explored that the trajectory patterns from geo-tagged social media are represented by a sequence of locations according to temporal order. Usually the pattern mining result is a set of mined patterns with their frequencies. However, this kind of representation has several disadvantages. Firstly, there are too many trajectory patterns in the result which is difficult to mine from the massive result set. Secondly, the top frequent trajectory patterns are usually short, not informative and less interested by all the users. Thirdly, redundancy information exists in the results because of the similar patterns. Many frequent trajectory patterns share common sub-patterns, so it is not interesting to output all of them. To avoid the above problems, in future research can be done to design an algorithm to rank trajectory patterns by taking into considerations like the relationships among online-users, regions and path tracking, and introduce an exemplar-based algorithm to diversify trajectory pattern ranking results.

6. Results

In this paper we explored the interconnections among the heterogeneous dimensions in social media including link, text and spatial-temporal data and mine useful knowledge from the integration of different data types. The table below shows our finding about the above improvised models for social media mining. Unlike traditional datasets, social media data has heterogeneous data types including link, text and spatial-temporal information. New methodologies are urgently needed for data analysis and many potential applications in social media. Ours focuses on exploring the various models to mine the knowledge from social media by taking advantage of embedded heterogeneous information.

Model	Features	Finding
1. LR	This model combines both link and attributes information in social graphs.	It lacks multiple intuitive factors that influence link creation.
2. DTTA	This model combines text and spatial information to discover topological topics	It works in finding regions of interests but also providing effective comparisons of the topics across different locations i.e. provide important clues to group different topological regions
3. DFTA	This model combines text and temporal information to discover frequent topics.	It lacks the periodicity of the terms as well as term co-occurrences and leverages the information from both text and time well.
4. DGTA	This model combines text and link information to discover group-based topics	It incorporates group discovery into topic analysis in text-associated graphs to guarantee the topical coherence in the groups so that users in the same group are closely linked to each other and share common dormant topics.
5. Ranking Based on Pattern Trajectory	It works for geo-tagged social media.	Through leveraging the relationships among users, locations and trajectories, we rank the trajectory patterns to reveal the collective wisdom in the seemingly isolated photos.

7. Conclusion

In this paper we have explored the heterogeneous data types in SM inclusive of Link, Spacial, Text and temporal Information which is used for mining useful knowledge. First, we try to explore the methods of Link Mining, Text mining, spatial temporal mining and then further discussed about few models like the Link recommendation model which further improves the structure of links in the SM. Secondly, we discussed a Dormant Topological Topic Analysis (DTTA) model that aggregates text and spacial information to explore the topological topics. Thirdly, we talked about the Dormant Frequent Topic Analysis (DFTA) model that inculcates text and temporal information to find frequent topics. Fourthly, we analyzed the Dormant Group Topic Analysis (DGTA) model that includes group discovery into topic analysis in text-linked graphs to ensure the topical coherence in the groups so that persons in the same group are closely associated to each other and share common dormant topics. At Last, we explored a Ranking Based on Pattern Trajectory in which finding the relationships between end-users, geographical locations and trajectories for the social media networks and then rank patterns to disclose the real picture in the geo-tagged isolated photos. All our studies in this paper, we have concluded that instead of mining these heterogeneous data in an isolated manner, we try to attempt to integrate link, text and spatial-temporal data in social media to mine useful knowledge from social media from different perspectives. We also address the urgent needs of analyzing the novel social media datasets in an integrated manner and advance both the principles and methodologies for data mining in social media.

8. References

[1] Lise Getoor and Christopher P. Diehl. Link mining: a survey. SIGKDD Explorations, 7(2):3-12, 2005

[2] Jure Leskovec, Kevin J. Lang, and Michael W. Mahoney. Empirical comparison Of algorithms for network community detection. In Proc. of 2010 Int. World Wide Web Conf. (WWW'10), pages 631-640, 2010.

[3] Hisashi Kashima, Tsuyoshi Kato, Yoshihiro Yamanishi, Masashi Sugiyama, and Koji Tsuda. Link propagation: A fast semi-supervised learning algorithm for link prediction. In Proc. of 2009 SIAM Int. Conf. on Data Mining (SDM'09), pages 1099-1110, 2009.

[4] Souvik Debnath, Niloy Ganguly, and Pabitra Mitra. Feature weighting in content based recommendation system using social network analysis. In Proc. of 2008 Int. World Wide Web Conf. (WWW'08), pages 1041-1042, 2008.

[5] David M. Blei. Introduction to probabilistic topic models. In Communications of the ACM, 2011.

[6] Francois Fouss, Alain Pirotte, Jean-Michel Renders, and Marco Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. IEEE Trans. Knowl. Data Eng., 19(3):355-369, 2007.

[7] David Liben-Nowell and Jon M. Kleinberg. The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology (JASIST), 58(7):1019-1031, 2007.

[8] Jerome Kunegis and Andreas Lommatzsch. Learning spectral graph Transformations for link prediction. In Proc. of 2009 Int. Conf. on Machine Learning (ICML'09), page 71, 2009

[9] Dan He and D. Stott Parker. Topic dynamics: an alternative model of bursts in streams of topics. In Proc. of 2010 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'10), pages 443-452, 2010.

[10] Matthew J. Rattigan and David Jensen. The case for anomalous link discovery.SIGKDD Explorations, 7(2):41-47,2005

[11] Joshua O'Madadhain, Jon Hutchins, and Padhraic Smyth. Prediction and ranking algorithms for event-based network data. SIGKDD Explorations, 7(2):23-30, 2005

[12] Alexandrin Popescul, Rin Popescul, and Lyle H. Ungar. Statistical relational Learning for link prediction, 2003

[13] Lei Tang and Huan Liu. Community Detection and Mining in Social Media. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers, 2010

[14] Ramnath Balasubramanyan and William W. Cohen. Block-lda: Jointly modeling entity-annotated text and entity-entity links. In Proc. of 2011 SIAM Int. Conf. on Data Mining (SDM'11), pages 450-461, 2011

[15] Guoqing Zheng, Jinwen Guo, Lichun Yang, Shengliang Xu, Shenghua Bao, Zhong Su, Dingyi Han, and Yong Yu. A topical link model for community discovery in textual interaction graph. In Proc. of 2010 ACM Conf. on Information and Knowledge Management (CIKM'10), pages 1613-1616, 2010.

[16] Guoqing Zheng, Jinwen Guo, Lichun Yang, Shengliang Xu, Shenghua Bao, Zhong Su, Dingyi Han, and Yong Yu. Mining topics on participations for community discovery. In Proc. of 2011 Int. ACM SIGIR Conf. on Research & Development in Information Retrieval (SIGIR'11), pages 445-454, 2011.

[17] Mo Chen, Jianzhuang Liu, and Xiaoou Tang. Clustering via random walk hitting Time on directed graphs. In Proc. of 2008 AAAI Conf. on Artificial Intelligence (AAAI'08), pages 616-621, 2008.

[18] Tomoharu Iwata, Takeshi Yamada, Yasushi Sakurai, and Naonori Ueda. Online Multiscale dynamic topic models. In Proc. of 2010 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'10), pages 663-672, 2010.

[19] Gabriel Pui Cheong Fung, Jefrey Xu Yu, Huan Liu, and Philip S. Yu. Time-Dependent event hierarchy construction. In Proc. of 2007 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'07), pages 300-309, 2007.

[20] Hila Becker, Mor Naaman, and Luis Gravano. Learning similarity metrics for Event identification in social media. In Proc. of 2010 Int. Conf. on Web Search and Data Mining (WSDM'10), pages 291-300, 2010.

[21] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In Proc. of 2010 Int. World Wide Web Conf. (WWW'10), pages 851-860, 2010.

[22] Jaewon Yang and Jure Leskovec. Patterns of temporal variation in online media. In Proc. of 2011 Int. Conf. on Web Search and Data Mining (WSDM'11), pages 177-186, 2011.

[23] Jure Leskovec, Lars Backstrom, and Jon M. Kleinberg. Meme-tracking and the Dynamics of the news cycle. In Proc. of 2009 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'09), pages 497-506, 2009.

[24] David J. Crandall, Lars Backstrom, Daniel P. Huttenlocher, and Jon M. Kleinberg. Mapping the world's photos. In Proc. of 2009 Int. World Wide Web Conf. (WWW'09), pages 761-770, 2009.

[25] Pavel Serdyukov, Vanessa Murdock, and Roelof van Zwol. Placing ickr photos On a map. In Proc. of 2009 Int. ACM SIGIR Conf. on Research & Development in Information Retrieval (SIGIR'09), pages 484-491, 2009.

[26] Gamhewage C. de Silva and Kiyoharu Aizawa. Retrieving multimedia travel Stories using location data and spatial queries. In Proc. of 2009 ACM Multimedia Int. Conf. (ACM-Multimedia'09), pages 785-788, 2009.

[27] Adrian Popescu and Gregory Grefenstette. Deducing trip related information from flickr. In Proc. of 2009 Int. World Wide Web Conf. (WWW'09), pages 1183-1184,2009.

[28] Munmun De Choudhury, Moran Feldman, Sihem Amer-Yahia, Nadav Golbandi, Ronny Lempel, and Cong Yu. Constructing travel itineraries from tagged geo-temporal breadcrumbs. In Proc. of 2010 Int. World Wide Web Conf. (WWW'10), pages 1083-1084, 2010.

[29] Munmun De Choudhury, Moran Feldman, Sihem Amer-Yahia, Nadav Golbandi, Ronny Lempel, and Cong Yu. Automatic construction of travel itineraries using social breadcrumbs. In Proc. of 2010 ACM Conference on Hypertext and Hypermedia (HH'10), pages 35-44, 2010.

[30] Xin Lu, Changhu Wang, Jiang-Ming Yang, Yanwei Pang, and Lei Zhang. Photo2trip: Generating travel routes from geo-tagged photos for trip planning. In Proc. of 2010 ACM Multimedia Int. Conf. (ACM-Multimedia'10), pages 1579-1582, 2010.

[31] Zhijun Yin, Manish Gupta, Tim Weninger, and Jiawei Han. A unified framework for link recommendation using random walks. In Proc. of 2010 Int. Conf. on Advances in Social Network Analysis and Mining (ASONAM'10), pages 152-159,2010

[32] Zhijun Yin, Manish Gupta, Tim Weninger, and Jiawei Han. Linker: a unified framework for link recommendation with user attributes and graph structure. In Proc. Of 2011 Int. World Wide Web Conf. (WWW'10), pages 1211-1212, 2010.

[33] Masaya Murata, Hiroyuki Toda, Yumiko Matsuura, Ryoji Kataoka, and Takayoshi Mochizuki. Detecting periodic changes in search intentions in a search engine. In Proc. of 2010 ACM Conf. on Information and Knowledge Management (CIKM'10), pages 1525-1528, 2010.

[34] David M. Blei. Introduction to probabilistic topic models. In Communications of the ACM, 2011.

[35] S. Parthasarathy, Y. Ruan, and V. Satuluri. Community discovery in social networks: Applications, methods and merging trends. In Social Network Data Analytics, pages 79-113. Springer US, 2011.

[36] Jure Leskovec, Kevin J. Lang, and Michael W. Mahoney. Empirical comparison of algorithms for network community detection. In Proc. of 2010 Int. World Wide Web Conf. (WWW'10), pages 631-640, 2010.

[37] Liangliang Cao, Jiebo Luo, A. Gallagher, Xin Jin, Jiawei Han, and T. S. Huang. A worldwide tourism recommendation system based on geotagged web photos. In Proc. of 2010 Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'10), pages 2274-2277, 2010.

[38] Takeshi Kurashima, Tomoharu Iwata, Go Irie, and Ko Fujimura. Travel route recommendation using geotags in photo sharing sites. In Proc. of 2010 ACM Conf. on Information and Knowledge Management (CIKM'10), pages 579-588,2010.

[39] Vincent Wenchen Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative location and activity recommendations with gps history data. In Proc. of 2010 Int. World Wide Web Conf. (WWW'10), pages 1029-1038, 2010.