

Modeling Characteristics of Location from User Photos

Vikas Kumar*
University of Minnesota
Minneapolis, USA
kumar093@umn.edu

Saeideh Bakhshi
Facebook Inc.
San Francisco, USA
saeideh@gatech.edu

Lyndon Kennedy
Futurewei Technologies, Inc.
San Francisco, USA
lyndonk@acm.org

David A. Shamma
Centrum Wiskunde &
Informatica
Amsterdam, Netherlands
aymans@acm.org

ABSTRACT

In the past decade, location-based services have grown through geo-tagging and place-tagging. Proliferation of GPS-enabled mobile devices further enabled exponential growth in geo-tagged user content. On the other hand, location-based applications harness the abundance of geo-tagged content to further improve user experience and more relevant localized content. We show in this paper that geo-tagged content can vary significantly based on whether they are captured by a local versus a tourist to the location. Using photos shared by online users, we also show how we can learn unique characteristics about a given location. We finally discuss an effective metric to rank the most representative photos for a given location by combining visual contents and their social engagement potential.

ACM Classification Keywords

H.4.0. Information Systems Applications: General; H.3.3. Information Systems: Information Search and Retrieval

Author Keywords

location-aware; content ranking; information theory; geo-tagged photos;

INTRODUCTION

Location plays a critical role in personalizing and identifying contextual meaningful content for users. Mobile services like Google Now, Yelp, Foursquare etc leverage location to improve user experiences [17, 25]. Other applications in online games [22] or shopping [12] are shown to adapt effectively with location to provide more personal experience and assistance respectively. At the same time with wide availability of

*Work was conducted while first author was intern at Yahoo Labs while other authors were part of Yahoo Labs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

HUMANIZE'17, March 13 2017, Limassol, Cyprus.

Copyright is held by the owner/author(s). Publication rights licensed to ACM
ISBN 978-1-4503-4905-5/17/03 ...\$15.00.
<http://dx.doi.org/10.1145/3039677.3039683>



Figure 1. Distinct group of users in San Francisco. Spots of photos taken by locals are in blue and tourists spots are in red. Yellow trails might be either. Image © Eric Fischer on Flickr: <https://flic.kr/p/87P5qP>.

location-sensing mobile devices users now find it more easy to share their own content tagged with location, for example photos. This exponential growth of geo-tagged contents provide valuable opportunity to study user photography patterns and learn about location.

Geographic patterns in photography have been studied extensively in the past decade. While geo-tagged photos have been used in tag-aggregation research [3, 4, 13, 21]; social interaction data and modern computer vision techniques can offer new insights in understanding user photographic behaviors. How do people interact with photos from places they are currently visiting or want to visit? Can computer vision and social interaction patterns be used to change the way we

visually summarize locations? What implications would these technologies have on location-based applications?

In most location-based services, the underlying assumptions are that online shared geo-tagged contents are representative of the location and that users are always *local* with respect to their shared content. For example, set of photos captured by a local who is a resident of a city are considered equally representative to those by a tourist or visitor. Johnson et al. [11] in their analysis of localness of geo-tagged content suggests that only 70% of contents represent what is local to the location and that this percentage decreases as we move from urban to rural location. In this paper, we further investigate the content of shared photos to understand the differences in characteristics locals would capture compared to tourists.

Using a large-scale geotagged photos dataset available from Flickr [23], an online social photo sharing platform, we demonstrate how existing location-based services can leverage the content and users' familiarity with location to learn about the location. We utilize computer vision techniques to infer the visual contents of photos and an information theory approach to learn descriptive spatial characteristics from photos for both locals and tourists.

Finally, we propose a characteristic score for each geo-tagged photo based on its visual characteristics and social engagement potential to rank the most representative photos of a city while maintaining the diversity in the content. In summary, we highlight the limitation of underlying assumptions in traditional location-aware application and emphasize the need to include the users' familiarity with location in considering the geo-tagged content in design of more effective and accurate location-aware systems.

CONCEPTUAL BACKGROUND

Places in the same city having different associations and effects on human emotions and behavior is a concept emphasized in psychogeography [5]: a field that discusses the laws, methods, and inventive strategies for exploring an urban landscape or a city. Consumer psychology of tourism [18, 14], for example, suggest differences in tourists' and locals' behavior depending on what a city has to offer. Tourists in their short stay are more likely to explore well known or famous attractions in the city while locals are likely to be interested in food, parks or other means of entertainments. For example, in San Francisco, tourists can be often seen at the ferry building, piers, or the Golden Gate bridge compared to locals who prefer to be at local restaurants, breweries, parks, stadiums etc. These differences are clearly shown in maps of various cities in United States by Eric Fischer [7]. An example map of San Francisco is shown in Figure 1.

Items with spatial footprints (e.g. geo-tagged photos, news, and restaurants) provide valuable opportunity for location-based systems. These systems leverage user's current location to recommend more contextual meaningful content [26] leading to more pleasant experiences [17, 25]. Studies show that spatial limitations on items and users are more efficient and accurate compared to traditional recommendation techniques [16, 15]; however, these location-based systems assume

every geo-tagged content to be representative of location. Instead, we find that cities are places where locals and tourists have different interests. Our design allows existing systems to reconsider the definition of locals in the location-aware systems and identify content uniquely representative of location.

Only a few recent studies like that of Zheng et al. [27] and Quercia et al. [19] measure the differences in users perception about a location. However, a major distinction lies in the way we model these perceptions. Previous studies have considered location or GPS traces to infer user interests and consider every content to be local; whereas we learn these preferences from the very *content* shared by these users.

MODELING VISUAL CONTENTS

In this section we describe the dataset, and explain how we derive descriptive characteristics of a given location from visual content of geo-tagged photos. We then summarize differences that exist among locals' and tourists' photos based on their spatio-temporal patterns.

Dataset

We use the YFCC100M image dataset [23] consisting of 100 million publicly-available Creative Commons images from Flickr. The images have attributes such as the owner, acquisition timestamps, user-provided titles, descriptions, tags, and geo-tagging. For our analysis, we consider only the subset of geo-tagged images taken in United States. Further, using Flickr API, we retrieved a multitude of social metrics such as the number of favorites, number of views and number of shares for each photo. For analysis, we only consider photos with at least 10 views and 10 favorites. We use these thresholds to consider photos shown to have potential for social engagement [1]. The final dataset consists of approximately 4.5 million images.

Location Characteristics

To find the characteristics photos capture about a city we leverage the visual content of shared photos. We learn meaningful descriptive characteristics from the photos in the form of keywords. These keywords are derived using a computer vision deep convolutional neural network based technique. The technique learns discriminative image representations from large-scale collections of training examples pre-trained on the ImageNet dataset [6] provided by the Caffe framework [10]. The output of the last fully-connected layer (fc7) delivers a 4096-dimensional feature representation of each image. Using a linear support vector machine [24], the images are then classified along 1700 different ImageNet concepts. We refer to these automatically-detected visual concepts as "characteristics".

In order to understand what characteristics *uniquely* identifies with the location, we model the descriptions (or characteristics) and locations (cities) into a information theory metric known as *conditional entropy*. The metric, defined as $H(X|Y)$, measures the certainty of variable X (bits of information) in a system given the knowledge about variable Y . Smaller metric value implies higher certainty about variable X . We model set of characteristics C and the set of locations L as the random

variables X and Y respectively; and formally define $h(c|l)$ as shown in Eqn. 1. This measures the certainty a characteristic (c) carries given the location (l).

$$h(c|l) = p(c, l) \times \log \frac{p(l)}{p(c, l)}; \text{ where, } c \in C, l \in L. \quad (1)$$

In equation (1), $p(c)$ is the ratio of the number of photos with visual descriptor c to the total number of photos; and, $p(l)$ is the ratio of number of photos taken at the location l to the total number of photos. Finally, $p(c, l)$ is the joint probability of characteristic c and location l . A summation of values for all characteristics C and locations L gives us the conditional entropy $H(C|L)$ of the system.

A smaller entropy (higher certainty) implies a higher chance that the characteristic c is unique to the location l . Likewise, higher entropy (less certainty) implies that the characteristic c is less likely to be representative or unique to the location. For instance, the characteristic “outdoor”, one of the most common visual descriptor in photos fails to uniquely identify with any given location thus having higher conditional entropy value. Whereas a tag like “latte” uniquely identifies with city of Seattle having a smaller conditional entropy value.

Locals Versus Tourists

To determine if users based on their familiarity capture unique photos of a location, we summarize photos into two distinct sets. First set consists of all photos taken by locals (more familiar of the location) and the second set are photos taken by tourists (less familiar of the location).

To determine locals and tourists, the timestamps of photos taken in succession by user was leveraged to identify their temporal associations with location. We visualize the distribution of differences in timestamp of first and last photos taken at a given location by the user. Based on the peak of activities, we segregate users into two sets with activity periods either to be (1) under 30 days, or (2) more than 30 days. We label users with shorter activity at a location (city in our case) as *tourists*, and the latter group of users as *locals*. The choice of 30 days is also found to be consistent with the prior definitions [7, 8] used to identify locals.

In the process however, it is important to account for the fact that an individual is likely to be recognized as local in more than one location. These locations can be the places where a user may have spent time during her childhood, or college, or currently as a resident. Likewise, the same individual is likely to be a tourist in more than one location, where she may have shorter periods of interactions. A user do not identify as both a local and a tourist for the same location.

It is also important to note that we find percentage of locals and tourists are different for different cities. For instance, in San Francisco, the percentage of local users is 68%, while tourists constitute the remaining 32%. And, the percentage of local users decreases slowly as we move from larger urban cities to more rural cities with moderate population. In smaller cities, most users are identified as tourists due to lack of active

Table 1. Top 5 tags unique for Seattle and San Francisco sorted by increased value of conditional entropy.

San Francisco		Seattle	
<i>Locals</i>	<i>Tourists</i>	<i>Locals</i>	<i>Tourists</i>
texture	urban	ocean	latte
graffiti	architecture	sunset	urban
people	skyline	sidewalk	architecture
monochrome	ferrybuilding	urban decay	skyline
portrait	bridge	biking	outdoor

periods of photography [11]. Thus, in this work, we focus only on urban cities where we find fair representation of both locals and tourists.

Finally, for each location we segregate the photos taken by users into local and tourist sets. Using the conditional entropy metric, we then identify unique characteristics for each location for both locals’ and tourists’ photos respectively. In Table 1, the top 5 descriptors for Seattle and San Francisco are shown with increasing metric value of conditional entropy (or decreasing certainty).

The top descriptors in Seattle for locals are “ocean”, “sunset” and “sidewalk” compared to those of tourists’ “latte”, “urban” and “architecture”. A visual inspection of the photos from Seattle taken by locals are shown to include sunset and sunrise photos. We believe that this is due to time of a day when residents are involved in casual walk or jog along the lakeshores and parks. Whereas tourists’ photos include shots of Pike Place Market, Starbuck’s first cafe, the urban-architecture of Space Needle and Seattle skyline. Likewise, for San Francisco, photos taken by locals are found to consist of pictures of local pride parades, or the graffiti in Clarion Alley (a popular area known for street art) while photos taken by tourists, similar to those in Seattle, include urban settings and architecture influenced by the skyline, famous Golden Gate Bridge, and Bay Bridge. The descriptions imply a strong evidence of varying characteristics that users capture in their photos based on their how they associate with location. These observations underline the limitation of assumption of local-ness in existing location-based services.

LOCATION-BASED PHOTOS RANKING

To examine if the characteristics learned about a location can help find more relevant photos for a location, we propose a score to re-rank the photos that we believe are most representative of location. In this section outline the motivation and the scoring metric for each photo.

Characteristics Score

Photos play a major role in illustration of human geography [20]. Photographs help us choose our travel destinations; they influence our decision making, planning, cognition and behavior at a destination [9]. Thus, we hypothesize that users would be more likely to explore about a location and thus browse photos before they actually post photos when they actually visit the location.

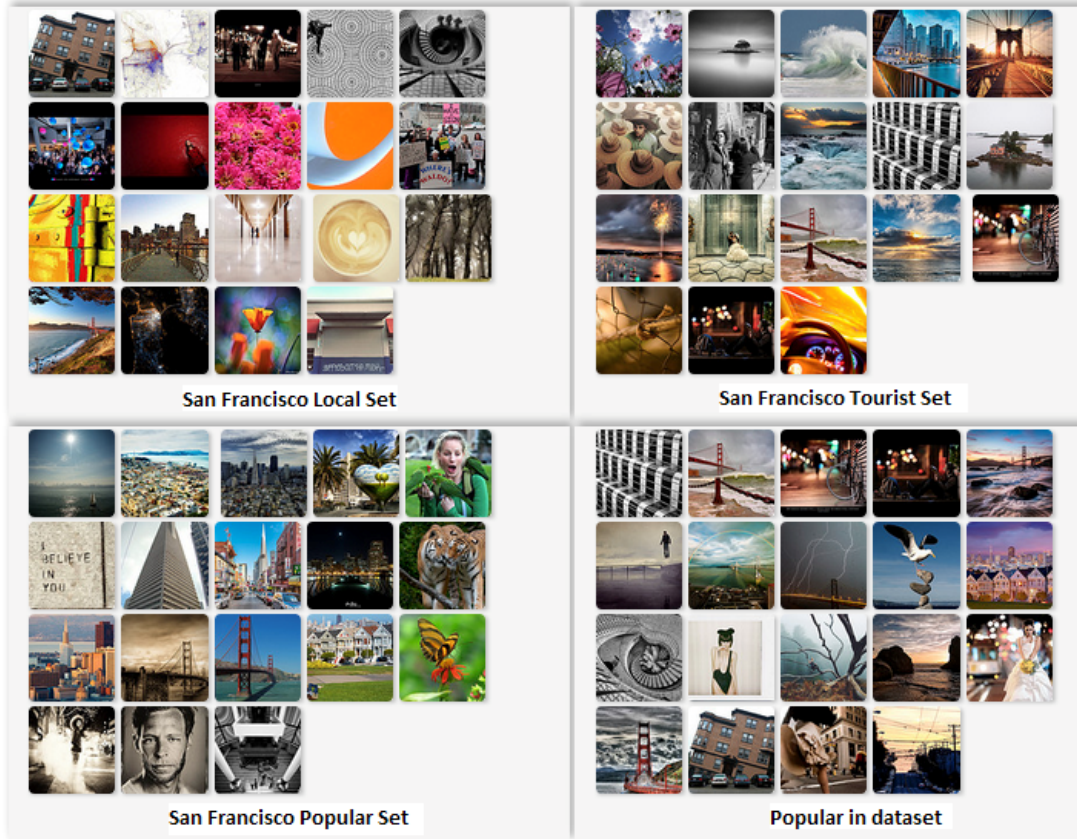


Figure 2. Characteristic based ranking of geo-tagged photos from San Francisco. First two sets (1st row) show the characteristics popular among locals and tourists respectively. The other two sets (2nd row) show the photos that are most popular in San Francisco and most popular in the dataset respectively. Note: Most popular photos in dataset are not necessary geo-tagged with San Francisco but have similar characteristics to those of found in San Francisco.

We test the hypothesis by comparing the timestamps of users' first *like/favorites* of existing photos to those photos taken by user for same location. The difference is calculated between these timestamps for all users. We find the difference to be smaller than zero (i.e. timestamps of favorites were smaller than timestamps of photo taken) and is statistically significant ($p\text{-value} < 0.001$) with an effect size of -0.89 . This highlights the importance of photographs to cater to users while they explore or learn about new location. Hence, we now propose a metric that is aimed to identify the most representative and diverse set of photos to aid users in their exploration about a location.

Metric: To identify and rank quality photos, an ideal system is one that optimizes both content and perceived relevance of items [2]. Using the characteristics derived from visual content; and social engagement potential (likes/favorites) as proxy for their relevance, we define a characteristic-score or $charScore_p$, for each geo-tagged photo (p). We use the summarized descriptive characteristics c we learn for a location l to find the most characteristic photos of location defined in a score as:

$$charScore_p = \frac{\log(\#faves_p)}{\sum_{c \in chars(p)} \frac{h(c|l)}{size(c)}} \quad (2)$$

The metric aims to identify photos that has higher social engagement potential and capture unique characteristics of the location. Higher the social engagement (that is, more number of favorites¹) imply higher relevance with users; and smaller the sum of conditional entropy imply higher chances the candidate represents location.

The characteristic-score is calculated for each individual photos in the locals and tourists set respectively. Ranking the photos in decreasing order of the score leaves us with the most representative photos for location as per locals and tourists respectively.

¹We also investigated other social metric signals to rank photos such as number of views, number of shares etc. in characteristic score. However, in preliminary analysis by a group of 8–10 users, photos are found to be very similar to ranking based on number of favorites. We choose to keep the social metric that is more available and intuitive across other domains.

DISCUSSION

To gather insights on quality of scoring of candidate photos and to understand how well photos actually capture the characteristics of location, we generate four different set of photos. Including the photos from two distinct photo sets, one from locals and one from tourists, we also include two other sets as baselines which include most popular photos at the given location, and the most popular photos in the dataset. We choose six major cities of United States for evaluation that includes Boston, Chicago, Los Angeles, New York, San Francisco, and Seattle. The choice of the cities are based on fairly balanced percentage distribution of locals and tourists.

We use the top 20 photos in each set for evaluation. However, in our preliminary analysis, photos in top of the list are found often to be taken by an expert user with very specific interests (like wedding, or landscape, or portraits) and huge fan-following resulting in unusual higher social engagement than others. This results in lack of *diversity* in the photo sets representing the location [28]. Although the quality and aesthetics of these photos are un-questionable, the high correlation in their main theme results in a repetitive candidate set. To address this issue, we adopt a simple approach of limiting the number of photos to *two* per photographer [in the set]. This helps us achieve a satisfactory level of diversity across the candidate set of photos. The sample photo sets for San Francisco is shown in Figure 2.

In our preliminary analysis with lab members, we find that tourists photos indeed show the most well known representative characteristics of location whereas photos based on local set tend to capture more of diverse and unknown characteristics of location, for example, the parade photos in San Francisco or the popular evening destinations of Seattle. We plan to continue the evaluation to gather feedback from more users and from both locals and tourists of the given location.

CONCLUSION

In this work, we propose and show how shared online photos help us identify representative characteristics of a location. We learn these characteristics to be unique to locations and that locals and tourists tend to capture different perspectives of the locations, an important result challenging the underlying assumptions in location-based services. We show and propose an effective method to not only segregate these contents between locals and tourists but also to learn and model different characteristics they illustrate about a location. Our evaluation results show that the scoring metric is able to capture the unique differences that exist between locals and tourists; and identify more representative and relevant photos for a location. We plan to continue this work and evaluate our results with more larger set of users in an online experiment to understand more about perceived quality, aesthetics and diversity of the photos. We could also be able to include assessment from users who are actually local or tourist to a location.

REFERENCES

1. Saeideh Bakhshi, David A Shamma, Lyndon Kennedy, and Eric Gilbert. 2015. Why We Filter Our Photos and How It Impacts Engagement. In *Ninth International AAAI Conference on Web and Social Media*.
2. Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12, 4 (2002), 331–370.
3. Iván Cantador, Ioannis Konstas, and Joemon M Jose. 2011. Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web* 9, 1 (2011), 1–15.
4. Iván Cantador, Martin Szomszor, Harith Alani, Miriam Fernández, and Pablo Castells. 2008. Enriching ontological user profiles with tagging history for multi-domain recommendations. (2008).
5. Guy Debord. 1955. Introduction to a critique of urban geography. *Critical Geographies A Collection of Readings* (1955).
6. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.
7. Eric Fischer. 2014. Locals and tourists. *A+ U-ARCHITECTURE AND URBANISM* 530 (2014), 30–33.
8. Fabien Girardin, Francesco Calabrese, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. 2008. Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE* 7, 4 (2008), 36–43.
9. Olivia H Jenkins. 1999. Understanding and measuring tourist destination images. *The International Journal of Tourism Research* 1, 1 (1999), 1.
10. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093* (2014).
11. Isaac L Johnson, Subhasree Sengupta, Johannes Schöning, and Brent Hecht. 2016. The Geography and Importance of Localness in Geotagged Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 515–526.
12. Gerrit Kahl, Lübmira Spassova, Johannes Schöning, Sven Gehring, and Antonio Krüger. 2011. IRL SmartCart-a user-adaptive context-aware interface for shopping assistance. In *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 359–362.
13. Marius Kaminskas, Francesco Ricci, and Markus Schedl. 2013. Location-aware music recommendation using

- auto-tagging and hybrid matching. In *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 17–24.
14. Metin Kozak, Enrique Bigné, Ana González, and Luisa Andreu. 2004. Cross-cultural behaviour research in tourism: a case study on destination image. *Consumer psychology of tourism, hospitality and leisure* 3 (2004), 303–317.
 15. Vikas Kumar, Daniel Jarratt, Rahul Anand, Joseph A. Konstan, and Brent Hecht. 2015. “Where Far Can Be Close”: Finding Distant Neighbors In Recommender Systems. In *Proceedings of LocalRec Workshop in ACM conference on Recommender Systems*. 13–20.
<http://ceur-ws.org/Vol-1405/paper-03.pdf>
 16. Justin J Levandoski, Mohamed Sarwat, Ahmed Eldawy, and Mohamed F Mokbel. 2012. Lars: A location-aware recommender system. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 450–461.
 17. Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. 2013. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1043–1051.
 18. RR Perdue, HJP Immermans, and M Uysal. 2004. *Consumer psychology of tourism, hospitality and leisure*. Vol. 3. CABI.
 19. Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 116–125.
 20. Gillian Rose. 2008. Using photographs as illustrations in human geography. *Journal of Geography in Higher Education* 32, 1 (2008), 151–160.
 21. Pavel Serdyukov, Vanessa Murdock, and Roelof Van Zwol. 2009. Placing flickr photos on a map. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 484–491.
 22. Ning Tan, Gaëtan Pruvost, Matthieu Courgeon, Céline Clavel, Yacine Bellik, and Jean-Claude Martin. 2011. A location-aware virtual character in a smart room: effects on performance, presence and adaptivity. In *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 399–402.
 23. Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: The New Data in Multimedia Research. *Commun. ACM* 59, 2 (Jan. 2016), 64–73. DOI : <http://dx.doi.org/10.1145/2812802>
 24. Vladimir Vapnik. 1998. *Statistical Learning Theory*. Wiley, New York.
 25. Mao Ye, Peifeng Yin, and Wang-Chien Lee. 2010. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 458–461.
 26. Hongzhi Yin, Yizhou Sun, Bin Cui, Zhiting Hu, and Ling Chen. 2013. Lcars: a location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 221–229.
 27. Vincent W Zheng, Yu Zheng, Xing Xie, and Qiang Yang. 2010. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th international conference on World wide web*. ACM, 1029–1038.
 28. Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists Through Topic Diversification. In *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*. ACM, New York, NY, USA, 22–32. DOI : <http://dx.doi.org/10.1145/1060745.1060754>