Dynamic Place Profiles from Geo-Folksonomies on the GeoSocial Web

Soha Mohamed and Alia Abdelmoty

Abstract The growth of the Web and the increase in using GPS-enabled devices, coupled with the exponential growth of the social media sites, have led to a surge in research interest in Geo-folksonomy analysis. In Geo-Folksonomy, a user assigns an electronic tag to a geographical place resource identified by its longitude and latitude. The assigned tags are used to manage, categorize and describe place resources. Building data models of Geo-folksonomy data sets that represents and analyses, tags, location and time information can be helpful in studying and analysing place information. The aim of my research is to use the spatio temporal data available on the Social web, to extract dynamic place profile. Building a dynamic profile involves including the temporal dimension in the Geo-folksonomy. Indeed, adding the temporal dimension can provide an understanding of geographic places as perceived by users over time.

1 Introduction

Recently, media has become popular and attracts many researchers from a variety of fields. Social networks enable users to post their user generated content at any time and from any location. With the emergence of Web 2.0, tagging systems evolved. Users can assign keywords of their choice to web resources. These keywords are called tags, and the process of assigning keywords to resources is termed as tagging. A typical examples of tagging systems are Delicious, Flickr and Tagazania where uses enters tags of their choices to annotate resources (url, images, places, .etc). Consequently, with the emergence of new social tagging systems such as Twitter and Foursquare, the structure of tags changed. Traditional tags consist of one word, while new tags may contain a sentence of more than one word. The result of tagging resources is called folksonomy. Folksonomies are created by tagging applications

Soha Mohamed, Alia Abdelmoty
Cardiff University, e-mail: (AlySA, A.I.Abdelmoty)@cardiff.ac.uk

via users' interaction on web 2.0. It consists of three main entities: users, tags and resources. Geo-Folksonomies are a special kind of folksonomies where the users assigns tags to geographical place resources. Studying tags from users visiting places can enhance the understanding of the geographic places visited, and its characteristics. Understanding places and their use can help cities understand social dynamics of these places and improve service provision and planning.

This paper proposes a framework (Fig.1) for constructing a dynamic place profile from the GeoFolksonomies. Mining the spatio-temporal data to extract the implicit semantics would give insight into the characteristics of users and place and interactions between them. In this work, Foursquare is used as a data source. Users can share their current location using venue check-in and/or leave a Tip about their experience in the visited venue. The framework involves different stages of tag extraction and folksonomy preparation in order to apply the temporal decay model for dynamic place profile construction. The derived place profile contains tags associated with the place and their weights. Furthermore, a temporal place similarity measure is used to measure the change of the similarity between places over time. Studying geofolksonmies was introduced before in [10], [22], and [23] but they didnt consider how the places change over time. Furthermore, similar work was done in [17] and [14] but they used the time factor to compare the folksonomy structure at different points of time and they didnt include the time in the equation of calculating the similarity between places and users.

The paper is structured as follows. A review of the related work on folksonomy analysis, geofolksonomy analysis, and the geo-temporal models are described in section 2. The proposed model for dynamic place profile is explained in section 3. Some results and analysis are discussed in section 4.

2 Related Work

In this section a summary of the related work is introduced. The following subsections will explain the the folksonomy analysis, geofolksonomy and the geotemporal models.

2.1 Folksonomy Analysis Methods

Formally, a folksonomy is a tuple F=$< U,T,R,A >$where U, T, R represent users, tags and resources respectively. The relationship 'A' relates U,T and R. Consequently, the folksonomy can be represented as a tripartite graph [11]. The vertices of the graph are the users, tags and resources. Alternatively, the folksonomy graph can be represented as a three dimensional adjacency matrix. However, to ease the manipulation, the tripartite graph can be decomposed to 3 bipartite graphs which are tag-user, tag-resource, and user-resource [1].

There are three measures of tag relatedness as stated in [4]: Co-occurrence, Cosine Similarity and Folk Rank.

In Co-occurrence measure, the tag-tag co-occurrence graph is defined as a weighted undirected graph whose set of vertices is the set T of tags, and two tags t1 and t2 are connected by an edge, iff there is a least one post. The weight of the edge is given by the number of posts that contain both t1 and t1.[5]

In Cosine similarity, the measure of tag relatedness is computed by using the cosine similarity of tag-tag co-occurrence distributions. Two tags are considered related when they occur in a similar context, and not when they occur together.[20]

The FolkRank method is derived from the PageRank algorithm which reflects the idea that a web page is important if there are many pages linking to it, and if those pages are important themselves.[2] The same principle is employed for Folkrank, a resource which is tagged with important tags by important users becomes important itself. The same holds for tags and users.[12]

2.2 Geo-Folksonomy

[10] introduced an approach for extracting place semantics embedded in geo-folksonomies. Social tags about places from Tagazania were used. In particular, perceptions of users about place and human activities related to places are captured to build place type and activity ontologies. The approach addresses the quality problems evident in the tags and place resources through a cleaning process; it also provides a place ontology model to capture the desired place semantics, and utilises external semantic resources and statistical co-occurrence methods to build the place ontology.

A suite of methods to extend folksonomies (SWE-FE) was presented by [22] SWE-FE utilizes the geospatial information associated with the three key components of the tagging system, tags resources and users. The authors extend the formal definition of folksonomy (user, tag, and resource) to contain the geographic location. They also extended the nave method to calculate the similarity between the users to include the distance as a factor in calculating the similarity. The geospatial folksonomy scheme described was implemented on GeoCENS. The authors argued that including the distance as a factor in the similarity can enrich the relationships among users and thus can provide better recommendation.

A model-based framework GeoFolk was introduced in [23] which combines both tags and spatial information to better content characterization. In Geofolk model, Bayesian statistics models were employed to represent Geodata, combined with tag co-occurrence patterns. The data set was taken from the publicly accessible CoPhIR2 dataset that contains metadata for over 54 million of Flickr resources. Each resource is annotated with one longitude and latitude and a set of tags. GeoFolk aims to explain the semantic relatedness of tags by means of latent topics in a probabilistic Bayesian framework. Latent Topics Models (LDA) are mechanisms for discovering the number of topics in a document in a probabilistic manner. In GeoFolk model, tag similarity/relatedness is estimated in a natural manner, by comparing tag distributions over latent topics. The author shows that GeoFolk works better than text-only analysis in tag recommendation, content classification and clustering. However, GeoFolk is not suitable for region clustering as it fails to find common topics in different geographical sites.

All the above work didn't consider the temporal dimension in the construction of folksonomy models. The following subsection introduces the temporal folksonomy that can be used to include the temporal dimension the in geo-folksonomy.

2.3 Geo-Temporal Modeling

A recent work was done in [17] to extract topics from user generated tips assigned to venues in the Foursquare dataset. The latent Dirichlet Allocation (LDA) topic model discussed earlier was used. A "topic"consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings. Using this model, each venue is expressed as a mixture of a given number of topics. An individual's activity identity is a combination of the venues in which he checks in. User similarity measure is calculated using the distinct venue topic distribution of every user in the dataset. *Jensen-Shannon divergence*(JSD) is used to compute a dissimilarity metric among users' topic distribution. The resulting metric is bound between 0 and 1 where 0 indicates that the two users topic signatures are identical and 1 representing complete dissimilarity. The limitation of their work is that a 3 hour time window was used as the temporal bound for users' activities collected, and the user similarity measure was used to compare users temporally based on this time window throughout the day. Further temporal factoring to reflect day of the week and month can enhance the ability of the model to discover similar users. Moreover, it is better to include time in computing the similarity between users and not only comparing similarities between user tags at different point of time.

Another recent work was claimed to be the first spatio temporal topic (STT) modelling for Location recommendation[14]. In their work, they addressed the problem of recommending right locations to users at the right time. Basically, they represented a check-in as a user, a location with a pair of coordinates, and a relative timestamp,which are all considered as observed random variables. Similar to LDA, a set of latent topics is defined. Each user is associated with a probability distribution over topics, which captures the user interests, and each topic has a probability distribution over locations, which captures the semantic relationship between locations. They performed an experimental evaluation on Twitter, Gowalla, and Brightkite data sets from New York. STT assumes that the topic distribution doesn't only depend on user topic distribution but also on the time's topic distribution. Moreover, the probability of recommending location depends on the given time's location distribution. The dynamic topic model they used captures the evolution of topics over time. The limitation of this model is that topics are modelled at each time slot, so the overall distribution of the topics in places can not be monitored.

3 The Proposed model: Dynamic place profiles

The proposed model is composed of different components as shown in Figure 1. Each component is explained in the following subsections.

3.1 Data Collection

In this work, different location based social network applications like Foursquare, Twitter, Flickr were studied and compared to choose the appropriate application to collect data from. Foursquare was chosen as it contains venues in which users can check-in and leave their tips. This data allows studying the three main aspects of human mobility which are geographic movement , temporal dynamics and interaction. The data collection system was implemented using Java and the Foursquare API.

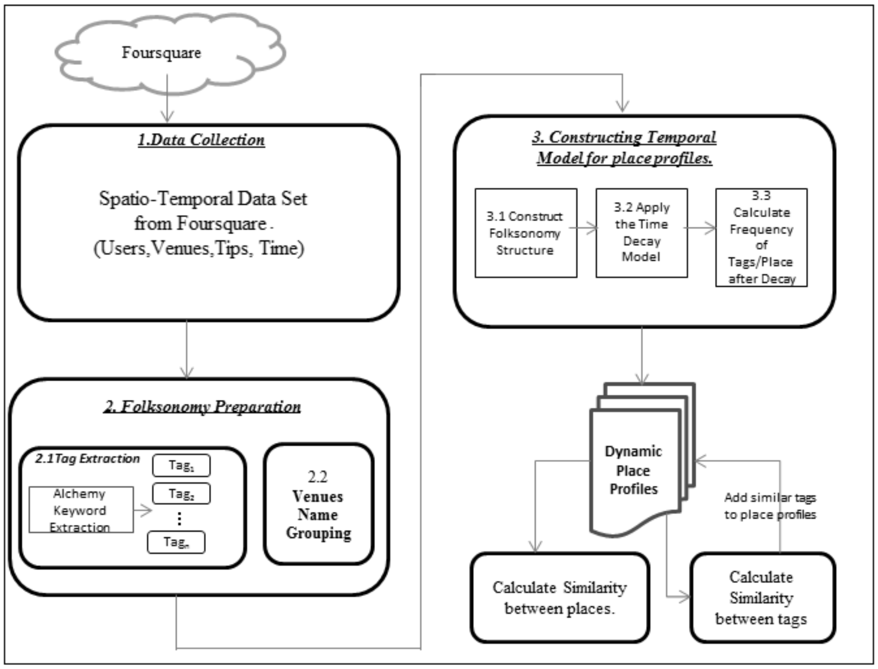
Foursquare
1.Data Collection
Spatio-Temporal Data Set
from Foursquare .
(Users,Venues,Tips, Time)
2. Folksonomy Preparation
2.1Tag Extraction
Alchemy Keyword Extraction
Tag$_1$
Tag$_2$
⋮
Tag$_n$
2.2
Venues Name Grouping
3. Constructing Temporal Model for place profiles.
3.1 Construct Folksonomy Structure
3.2 Apply the Time Decay Model
3.3 Calculate Frequency of Tags/Place after Decay
Dynamic Place Profiles
Add similar tags to place profiles
Calculate Similarity between places.
Calculate Similarity between tags

Fig. 1 Proposed Model Framework

The two functions venuesSearch() and venueTips() were used to get all the venues in a specific longitude and latitude and the tips within each venue. Cardiff, Bristol were chosen for the data collection. The three data sets will be used in the evaluation of the proposed model. A summary of some statistics of the data collected in Cardiff and Bristol are in Table 1.

Table 1 Data Sets

	Cardiff Dataset	Bristol Dataset
Venues Count	1,627	2,082
Venus with Tips	446	409
Tips Count	1,084	1411
Tags Count	2,976	3,260
Users Count	876	1,094

3.2 Tags Extraction

Tips are different from tags because they contain more than one word. So, there is a need to extract tags from tips as a pre-processing step. In this work, Alchemy API[1] is used to extract important keywords from the tips. The Alchemy API employ sophisticated statistical algorithms and natural language processing technology to analyse the data, extracting keywords that can be used to index content. The following are some examples of extracting keywords from tips.

Tip: Try the double chocolate brownie and enjoy
Keywords: double chocolate brownie, enjoy

Tip: Was in Millennium Stadium for Judgement Day on Easter Eve
Keywords: Millennium Stadium, Easter Eve, Judgement day.

3.3 Database Design

The database engine used in this research is MySQL V.5. The database is stored on the school server (emphesus.cs.cf.ac.uk). The Geofolksonomy database is designed to support storing and searching of the collected Geo-Folksonomy datasets as well as the output of the Folksonomy analysis methods implemented. The data model of the database is shown in Figure 2.The three distinct components of the

[1] http://www.alchemyapi.com/api/

geo-folksonomy are modelled using the Place table representing folksonomy venue resources, the Tag table representing folksonomy tags, and the User table representing folksonomy users. The user_tag_place table relates the three tables user, place, and tag. The timestamp is placed in the user_tag_place table in preparation to apply the time model to the data. The database also contains several tables for storing the output of the folksonomy analysis such as tags similarity and place similarity.

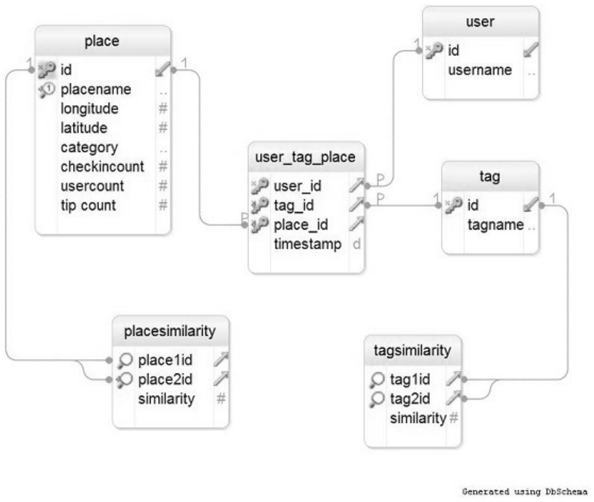

Fig. 2 Database Schema

3.4 Dynamic Place Profile Construction

An important model used in temporal modelling is the decay model (also called The ageing theory)[6].The Decay is an exponential function defined as the way in which a quantity naturally decreases over time at a rate proportional to its current value [9]. Tags and posts are considered to have a lifetime. When they are first defined, they are fresh and interesting to users then they decay over time and their interest value decreases. [24]. The life cycle ageing theory was successfully used in a work done on Twitter in order to detect real time emerging topics by mining terms that frequently occur in specified time interval. [3]. [24] used the decaying equation in order to track the time changing behaviour of a tag through the lifetime of all tagged posts. They used the temporally ordered sequential tag stream in order to mine burst or anomaly intervals of tags. [21] proposed a temporal semantic model to compute semantic relatedness of words. They applied the exponential weighting function on word time series in order to put more emphasis on recent correlations between words. Using the decay model for modelling the dynamic place profiles will be useful as it is important to maintain the freshness of the tags and to monitor the change of the tag behaviour over time.

In the proposed model, the Decay Model is used as it suits the nature of our problem. A quantity is subject to exponential decay if it decreases at a rate proportional to its current value. In constructing a dynamic place profile, tags should be ranked according to freshness and frequency [9]. Old tags with no repetitions should decay over time, while recent tags should have strong weight. In addition, if a recent tag has a small frequency, its weight should be high. Similarly, old tags with high frequency should have a high weight as well. Symbolically, this process can be expressed by the following equation:

$$P(t) = P_0 e^{-rt} \tag{1}$$

where:
P(t) = the amount of some quantity at time t.
P_0 = initial amount at time t = 0.
r = the decay rate.
t = time (number of periods).

In this work, we propose two ways of creating place profile. The first is the Dynamic Direct Place Profile, in which we add all the Tags associated with the place and its frequency, and the second way is the Dynamic In direct Place Profile in which the tag-tag similarity is computed using cosine similarity measure. The strength of association measure is then assigned to place based on tag similarity. The more similar a tag is to another, the stronger the relationship between that tag and the place.

3.4.1 Direct Dynamic Place Profile

Each place in the database is associated with a set of tags, and each tag has a frequency. As the time passes, the frequency of the tag should decay slightly to maintain the freshness of tags. Algorithm 1 is a summarized version of calculating the Tag Frequency after applying the decay equation. The algorithm has three inputs, TimeCluster in which the user can specify to cluster the data by hour, day, week, month or year. The second input is the Date, that is the point of Time the user want to calculate the Tag Frequencies. The third input is r which is the decay rate as mentioned previously. After applying this algorithm, the output is a list of Tag Frequencies of each place with the dynamic effect.

```
Algorithm 1: DYAMIC TAGFREQUENCY($TimeClusters, Date, r$)

for each $place \in PlaceTable$
  do
    $Cluster\ Tag\_User\_Place\ Table\ by\ TimeClusters$
    $Calculate\ TagFrequency\ within\ each\ TimeClusters$
    for each $T \in TimeStamps$
      do
        $Calculate\ No.\ of\ periods\ between\ T\ and\ Date$
        $ApplytheDecayEquation$
        $Calculate\ TagFrequencyafter\ Decay$
    for each $Tag$
      do for all $TimeSlots$
      do $SumUp\ TagFrequency$
return $TagFrequency$
```

3.4.2 Indirect Dynamic Place Profile

In the Indirect place profile, the tag-tag Similarity Measure [10] is calculated using the cosine similarity measure [15].The tag-tag Similarity is used to find the tags that are similar to the tags in the Dynamic Direct Place Profile. The similarity between two tags is defined by the following Equation:

$$sim(T_1, T_2) = \frac{|P_1 \cap P_2|}{\sqrt{|P_1|.|P_2|}} \tag{2}$$

where T_i represents a tag and P_i represents the set of instances of place resources associated with the tag T_i in the Geo-folksonomy after applying the dynamic tag frequency algorithm.

3.5 Place Similarity

The cosine similarity method is used to calculate the similarity between places in the database. The place similarity depends on the co-occurrence of tags in different place. It also depends on the weight of each tag. The geo-folksonomy dataset is used to assess the similarity of the place instances.

4 Analysis and results

Figures 3 and 4 shows a comparison between tags similar to tag ’Bay’ during years 2011 and 2012 in Cardiff dataset. The nodes represents the tags and the weights on the edges represent the similarity measure between two tags. The similarity measure is a number between 0 and 1. As shown in figures, more tags were added during 2012. The similarity measure between tags that were not mentioned in 2011 decreased and decayed in 2012, and the similarity of tags that are mentioned in both 2011 and 2012 increased. This shows the benefit of using the decay model over other temporal models as it doesn’t forget the past tags unless they are never mentioned.

Figures 5 and 6 shows a map based comparison of the places similar to ’Millennium centre’ in Cardiff during 2011 and 2012 respectively. The similarity between two places are calculated using the cosine similarity measure between the tags mentioned in the two places. The more similar the places are the bigger the radius of the circle.

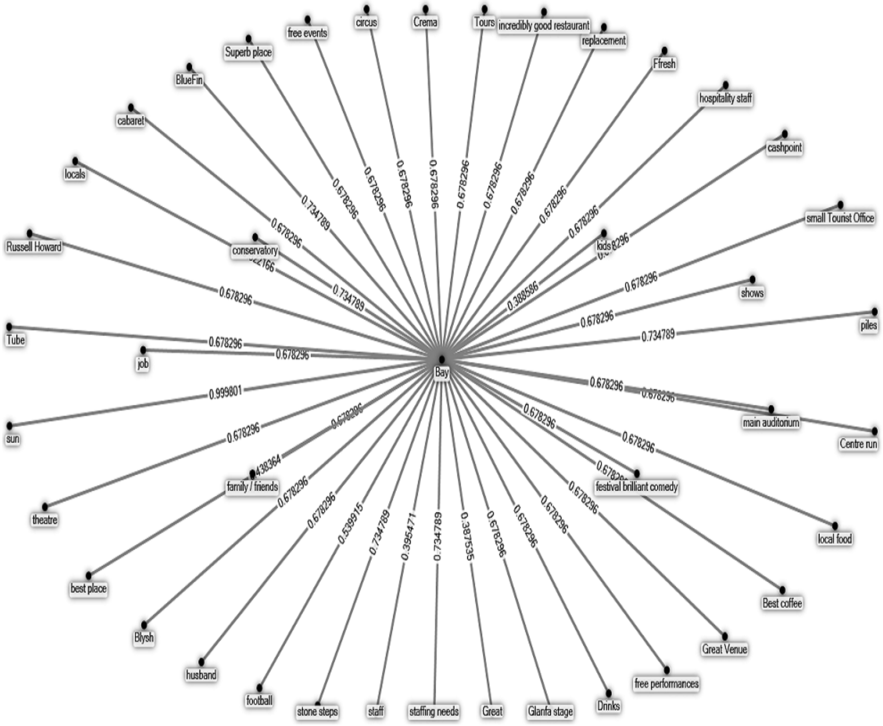

Fig. 3 Tags similar to tag ’Bay’ in Cardiff Dataset during 2011

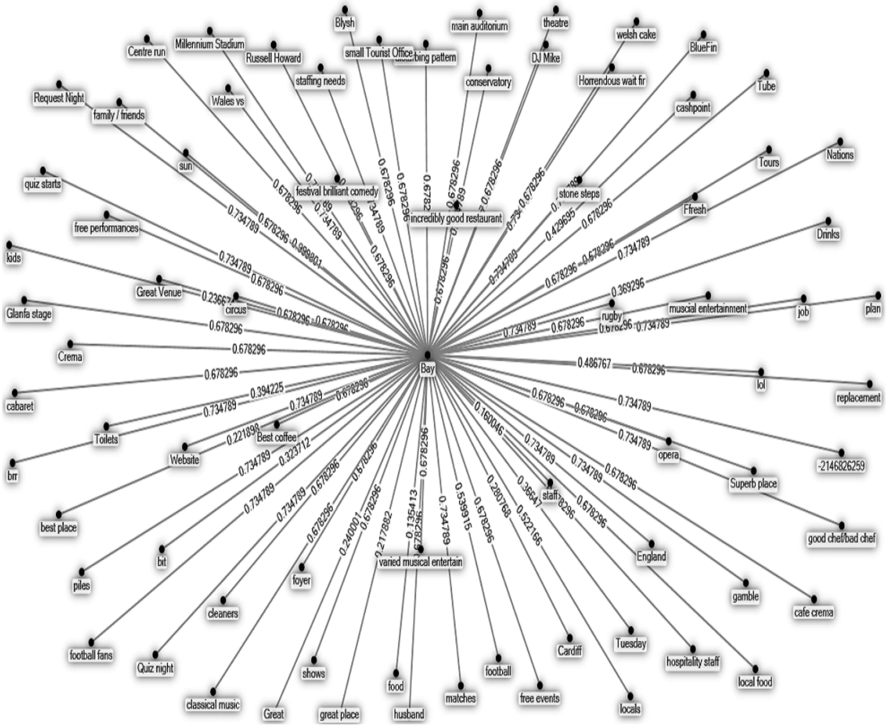

Fig. 4 Tags similar to tag 'Bay' in Cardiff Dataset during 2012

5 Conclusion and Future work

In this paper, a novel framework for dynamic place profile have been introduced. The aim of this research work is to study the spatiotemporal aspects of Social Web data and analyse their value in understanding user place characteristics visited by users. The framework developed have many potential applications and uses. In particular, it can be used to provide users with more personal search experience. It can be used by recommendation services and personalisation applications to provide users with relevant information. It can improve the recommendation services and lead to more targeted adverts and commercials. It can improve location-based service, by providing personalised access to local services.The next step is make an evaluation application in order to evaluate the place profile and to compare it against models developed for temporal geo-modeling mentioned in the related work.

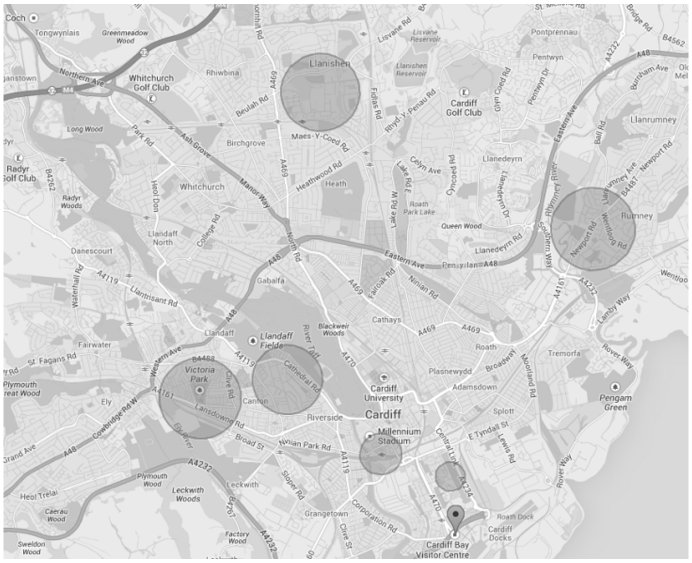

Fig. 5 The top places similar to tag 'Millennium Centre' in Cardiff dataset during 2011

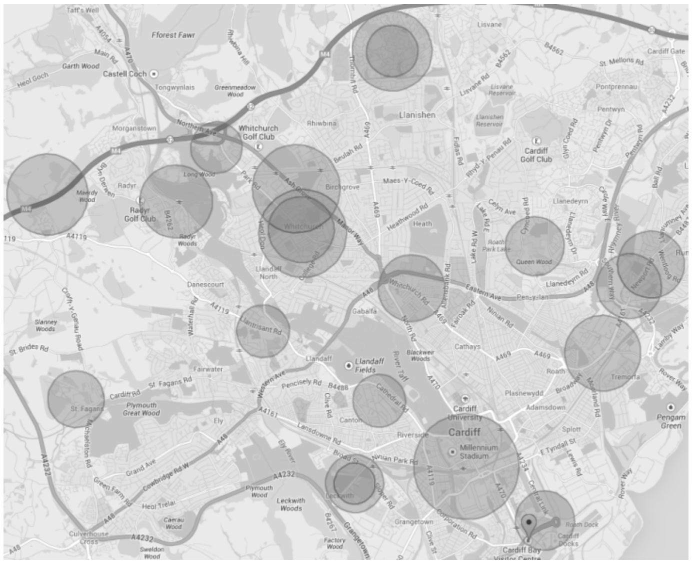

Fig. 6 The top places similar to tag 'Millennium Centre' in Cardiff dataset during 2012

References

1. Beldjoudi, S., Seridi, H., Faron-Zucker, C. Ambiguity in Tagging and the Community Effect in Researching Relevant Resources in Folksonomies. In Proc. of ESWC Workshop User Profile Data on the Social Semantic Web.(2011)

2. Brin, S., Page, L. The anatomy of a large-scale hypertextual Web search engine. Computer networks and ISDN systems, 30(1), 107-117.(1998)
3. Cataldi, M., Di Caro, L., Schifanella, C. Emerging topic detection on twitter based on temporal and social terms evaluation. In Proceedings of the Tenth International Workshop on Multimedia Data Mining (p. 4). ACM.(2010)
4. Cattuto, C., Benz, D., Hotho, A., Stumme, G. Semantic analysis of tag similarity measures in collaborative tagging systems. arXiv preprint arXiv:0805.2045.(2008)
5. Cattuto, C., Barrat, A., Baldassarri, A., Schehr, G., Loreto, V. Collective dynamics of social annotation. Proceedings of the National Academy of Sciences, 106(26), 10511-10515. (2009).
6. Chen, C. C., Chen, Y. T., Sun, Y., Chen, M. C. Life cycle modeling of news events using aging theory. In Machine Learning: ECML 2003 (pp. 47-59). Springer Berlin Heidelberg.(2003)
7. Cho, E., Myers, S. A., Leskovec, J. Friendship and mobility: user movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1082-1090). ACM.(2011)
8. Colorni, A., Dorigo, M., Maniezzo, V. An Investigation of some Properties of an"Ant Algorithm". In PPSN (Vol. 92, pp. 509-520). (1992)
9. Fonda, L., Ghirardi, G. C., Rimini, A. Decay theory of unstable quantum systems. Reports on Progress in Physics, 41(4), 587.(1978).
10. Elgindy, E. Extracting place semantics from geo-folksonomies (Doctoral dissertation, Cardiff University).(2013)
11. Halpin, H., Robu, V., Shepherd, H. The complex dynamics of collaborative tagging. In Proceedings of the 16th international conference on World Wide Web (pp. 211-220). ACM.(2007).
12. Hotho, A., Jschke, R., Schmitz, C., Stumme, G. Folkrank: A ranking algorithm for folksonomies. In K. D. Althoff (Ed.), LWA (Vol. 1, pp. 111-114).(2006)
13. Hotho, A., Jschke, R., Schmitz, C., Stumme, G. Trend detection in folksonomies (pp. 56-70). Springer Berlin Heidelberg. (2006)
14. Hu, B., Jamali, M., Ester, M. Spatio-Temporal Topic Modeling in Mobile Social Media for Location Recommendation. In Data Mining (ICDM), 2013 IEEE 13th International Conference on (pp. 1073-1078). IEEE.(2013)
15. Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., Stumme, G. Evaluating similarity measures for emergent semantics of social tagging. In Proceedings of the 18th international conference on World wide web (pp. 641-650). ACM.(2009)
16. Mathioudakis, M., Koudas, N. Twittermonitor: trend detection over the twitter stream. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 1155-1158). ACM.(2010).
17. McKenzie, G., Adams, B., Janowicz, K. A thematic approach to user similarity built on geosocial check-ins. In Geographic Information Science at the Heart of Europe (pp. 39-53). Springer International Publishing.(2013)
18. Michlmayr, E., Cayzer, S., Shabajee, P. Adaptive User Profiles for Enterprise Information Access. In Proc. of the 16th Intl. World Wide Web Conference.(2007)
19. Pham, X. H., Jung, J. J., Hwang, D. Beating Social Pulse: Understanding Information Propagation via Online Social Tagging Systems. J. UCS, 18(8), 1022-1031.(2012)
20. Quattrone, G., Ferrara, E., De Meo, P., Capra, L. Measuring similarity in large-scale folksonomies. arXiv preprint arXiv:1207.6037.(2012)
21. Radinsky, K., Agichtein, E., Gabrilovich, E., Markovitch, S. A word at a time: computing word relatedness using temporal semantic analysis. In Proceedings of the 20th international conference on World wide web (pp. 337-346). ACM.(2011)
22. Rezel, R., Liang, S. SWE-FE: extending folksonomies to the sensor web. In Collaborative Technologies and Systems (CTS), 2010 International Symposium on (pp. 349-356). IEEE.(2010)
23. Sizov, S. Geofolk: latent spatial semantics in web 2.0 social media. In Proceedings of the third ACM international conference on Web search and data mining (pp. 281-290). ACM.(2010)
24. Yao, J., Cui, B., Huang, Y., Jin, X. Temporal and Social Context Based Burst Detection from Folksonomies. In AAAI. (2010)