ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Analyzing Flickr metadata to extract location-based information and semantically organize its photo content

Evaggelos Spyrou [a], Phivos Mylonas [b,*]

[a] Technological Educational Institute of Central Greece, 3rd km Old National Road Lamia-Athens, P.C. 35100, Lamia, Greece
[b] National Technical University of Athens, Iroon Polytechneiou 9, P.C. 15780 Zografou, Athens, Greece

ARTICLE INFO

Article history:
Received 2 November 2013
Received in revised form
1 December 2014
Accepted 26 December 2014

Keywords:
Social media content analysis
Geo-tagging
Tags
Content clustering
Information processing

ABSTRACT

The first step towards efficient social media content analysis is to understand it and identify means of user interaction. Trying to study the problem from the user perspective, we analyze user-generated photos uploaded to famous Flickr social network, in order to extract meaningful semantic trends covering specific research aspects, like content popularity, spatial areas of interest and popular events. Initially, we select a geographical area of social interest, like a city center, defined by a strict bounding box. We then cluster photos taken within the box based on their geo-tagging metadata information (i.e., their latitude and longitude information) and divide large areas into smaller groups of fixed size, which we will refer to as "*geo-clusters*". Within these geo-clusters, we further identify semantically meaningful "places" of user interest, by analyzing any additional textual metadata available, i.e., user selected tags that characterize each place's photos. By post-processing the latter, we are then able to rank them and thus select the most appropriate tags that describe landmarks and other places of interest, as well as events occurring within these places of interest. As a next step, we place these tags on a map and help users to intuitively visualize places of interest and the actual photo content at a glance. Finally, we examine the temporal dynamics of analyzed photos over a long period of time, so as to obtain the underlying trends to be identified within this kind of social media generated content.

1. Introduction

The recent growth of social networks coupled together with an extreme amount of multimedia content data, mostly in the form of digital still images, deriving from both personal and social media, gave rise to interesting applications and technologies that support them. In this work we initiate our research by first trying to understand the mechanisms that allow users to interact and exchange photo content on social media platforms such as Flickr[1] and by analyzing the underlying trends that accompany mass online multimedia content sharing. Being the center of attention, online user-generated multimedia content met an unprecedented interest increase in terms of its organization and manipulation. Consequently, there is an urgent and growing need to facilitate effortless user access and manipulation to these rather unorganized and unsorted media archives, in order for typical users (a) to take advantage of the inherent additional meta-information that is present within them (e.g., geo-tags) and (b) to exploit it. Typical approaches for assisting such information access, like browsing, searching, filtering, or recommendation techniques, although quite advanced in the textual domain, are still in their early steps with respect to the mass online multimedia content domain.

The latter observation may be attributed in the most part to the lack of sufficient – additional to the actual content itself – textual annotations, tags or geo-tags associated with multimedia content, which firstly hinders the application of text-based retrieval techniques and secondly, obstructs efficient organization of such enriched multimedia content. In addition, the art of analyzing and identifying patterns of temporal variation with respect to online content in general, forms another difficult task, mainly due to the fact that human behavior – that is inherent behind the temporal variation – is considered to be highly unpredictable and outside of any known model; the latter ranging typically between "random" [40] and "highly correlated" [10] states.

In this paper we shall focus on a subset of the above described information handling problems, which, however, lies within current top research trends and applied services: we aim to analyze large user-generated digital photos collections (such as the ones derived from Flickr), in order to select the most appropriate meta-tags to describe a geographical area of interest and thus characterize the content itself in terms of its semantics, spatial and chronological context. In the following we present a holistic attempt of our work methodology, starting from its very first steps on photo clustering

* Corresponding author. Tel.: +30 2107724351.
E-mail addresses: vspyrou@teilam.gr (E. Spyrou), fmylonas@image.ntua.gr (P. Mylonas).

[1] http://www.flickr.com

http://dx.doi.org/10.1016/j.neucom.2014.12.104
0925-2312/

based on their respective geo-information, up to: (a) each geo-cluster's textual metadata analysis, (b) the extraction of meaningful semantic trends covering the specific geographical areas of interest and (c) the final computation of their temporal dynamics over a long period of time. In the evolution of our work description we shall also illustrate some additional intermediate steps, such as placing clusters and tags on a map to help readers intuitively understand both the reasoning in the points of interest selection and the actual visual content associated. As a result, the *main contributions* of this work may be summarized as follows:

(i) we propose a two-level, semantically meaningful clustering scheme on geo-tags, based on KVQ [55]. We utilize this scheme in order to create fixed-size clusters that would semantically correspond to "places"; we define the latter to be a rather compact and meaningful geographic area. We only select "places" that involve the collective intelligence of Flickr users, or in other words "places" that show at least some user interest,

(ii) we introduce an innovative probabilistic approach for selecting the most important tags, which considers certain interesting aspects of tags,

(iii) we provide a principal trend analysis and classify tags as landmarks and events based on the temporal distributions of their textual metadata, and

(iv) we place the most important metadata on a map and visualize their level of importance.

At this stage it is also worth pointing out some novel aspects of this work. First of all we made a choice to deal with tags and geo-tags by utilizing fixed-size clusters. In this manner, we are sure that tags that belong to landmarks or area-specific events always end up to the same cluster. Secondly, we consider the user factor in the process, since the clusters that occur do not have predefined boundaries, but quite on the contrary we only predefine the shape of the clusters, while their centers are determined after an automatic, unsupervised approach. As already mentioned, we also propose a probabilistic framework, in order to select the most representative tags, characterized by novel notions in the modeling of tags and their spatial neighbors and also in the modeling of geo-places. All in all, this research work attempts to broaden the scope of tag-recommendation approaches by providing a broader, semantic-based view on it. Last, but not least, the herein proposed methodology is fully automated, as it demands only two user-defined parameters, i.e., the radii of geo-clusters and geo-places; a fact that to the best of our knowledge constitutes it rather unique in nature.

A last justification is required for selecting Flickr social network to base all our observations, studies and applications; Flickr was favored due to the main fact that it has been very popular during the last few years, both for being the largest collection of community collected geo-tagged photos and for offering a public Application Programming Interface (API)[2] for accessing these photos along with their metadata information. Each Flickr uploaded photo may contain metadata added by its photographer, such as textual tags that describe either its visual content or location, or even a free text that describes the photo from its uploader's point of view. It may also contain metadata added by the hardware equipment used to capture it or by the photographer, such as date taken, camera settings (e.g., ISO, shutter speed or aperture values), equipment type (e.g., camera, smartphone), etc.

The rest of this paper is organized as follows. In Section 2 we begin by presenting relevant recent research works on handling social media content collections, focusing on metadata available from Flickr, as well as other online social networks, micro-blogging platforms and content collections. Then, in Section 3 we present the main aspects of our work, that may be summarized briefly in the clustering technique we apply on photos based on their geo-data, the tag-ranking algorithm we apply on each cluster and the clusters' modeling and transition to semantic "places", as well as the definition of trends and time exploitation. Our experimental results derived from the application of discussed approach on two datasets are provided through a detailed case study in Section 4. Finally, in Section 5 we briefly discuss our future plans.

2. Related work

2.1. Exploiting information

As expected, the tasks of semantically characterize, organize and efficiently exploit user-generated multimedia content towards the meaningful exploitation of its carrying information are of great importance within recent research community efforts. Starting back in 2009, Cha et al. [12] collected and analyzed large-scale traces of information dissemination derived from Flickr, aiming at answering a set of information propagation questions. More recently, Kalantidis et al. [28] proposed a visual-based photo image retrieval and localization approach, which exploited low-level image characteristics similarities in order to achieve accurate results. Another interesting approach is [34], where meaningful travel route recommendations are proposed, utilizing Flickr's user histories and past actions behaviors. Still, other approaches focus on mobile platforms and try to investigate whether knowledge extracted from massive content user contribution and interaction may offer any kind of added-value services [64].

Lately, research interest has been given also on statistical approaches to the problem, i.e., Yang et al. [59] developed a k-spectral centroid clustering algorithm in 2011, so as to identify temporal patterns in online media. Huberman et al. [24] studied the social interactions on the famous social microblogging network Twitter, and came to the conclusion that the underlying driving usage process is a sparse hidden network of friends and followers, while most of the links represent meaningless interactions. The almost real-time nature of information exchange inherent within this social medium constitutes it as the ideal candidate for related trend research, so Java et al. [26] investigated its social structures and managed to isolate different types of user intentions, whereas the same social network has been also examined later on by Jansen et al. [25] as a mechanism for word-of-mouth advertising.

In an effort to address and overcome some of these issues that hinder effective content access and interaction, researchers have focused on the notion of *collective intelligence*, [19] trying to identify potential sources of knowledge that would lead to efficient multimedia content characterization and thus, manipulation. Towards this direction, the addition of the notion of collectiveness aids the overall pattern deviation and complexity increase, considering all possible differentiations in interactions between small or larger groups of people. Given the fact that online user-generated multimedia content is increasingly popular, several research methods for organizing and providing access to its data have been emerged on this topic since the last few years, constituting the fulfillment of our motivation an extremely intriguing research task.

2.2. Exploiting traditional image analysis techniques

In the seek of efficient social media photo content analysis many research works exploit the fact that good, old traditional visual content image analysis may indeed provide a rather powerful description. As a

[2] http://www.flickr.com/api

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

result, many research efforts try to combine visual descriptions with textual metadata in order to acquire the most out of photos. Crandall et al. [15] use visual, temporal and geospatial information to automatically identify places and/or events in city and landmark level. They also add temporal metadata information to improve classification performance. With the same motivation, Quack et al. [48] divide the area of interest into non-overlapping, square tiles, then extract and use visual, textual and geospatial features. They handle tags by a modified TF-IDF ranking and link their results to Wikipedia.[3] Gammeter et al. [20] overlay a geospatial grid over earth and match pairwise retrieved photos of each tile using visual features. Then, they cluster photos into groups of images depicting the same scene. The metadata are used to label these clusters automatically, using a TF-IDF scheme. Moëllic et al. [44] aim to extract meaningful and representative clusters from large-scale image collections. They propose a method based on a shared nearest neighbors approach that treats both visual features and tags.

Li et al. [38] propose an algorithm that learns tag relevance by voting from visually similar neighbors. They do not use geospatial data, nor limit their approach on landmarks/places of interest and aim to retrieve semantically similar images. Moxley et al. [45] classify mined geo-referenced tags as places, by extending [49], landmarks by clustering image datasets considering mutual information and prior knowledge from Wikipedia and visual terms using the mutual information between visual descriptors and tags. Hays and Efros [23] advance to world-scale geographic estimation by searching into a database of 6M geo-tagged Flickr images; however, since their images were represented by global features like color histograms, GIST descriptors [47], etc., matching accuracy is not even comparable to that of local features and the output is a geo-location probability map. Kalogerakis et al. [29] build also on previous results by exploiting in addition the time each photo is taken, much like [15].

2.3. Exploiting only textual metadata

However, since the above described extraction and manipulation of visual content may be proved slow and even computationally difficult in some cases, many researchers propose to work solely on the textual part of image descriptions, i.e., the available user-provided textual metadata. Lee et al. [35] create overlapping geographical clusters for each tag and then, for a pair of two tags they calculate their geographical similarity. Then they introduce weighted similarities for both tags and geographical distributions and use the mutual information of tagging and geo-tagging. Rattenbury et al. [49] aim to extract semantics such as places and events from tags and unstructured text-labels. They observe that event tags follow certain temporal patterns, while place tags follow certain spatial patterns. They use methods inspired by burst-analysis techniques and propose scale-structure identification. Abbasi et al. [1] identify landmarks using tags and Flickr groups without exploiting any geospatial information. They use SVM classifiers trained on thematical Flickr groups, in order to find relevant landmark-related tags. Ahern et al. [4] analyze tags associated with geo-referenced Flickr images so as to generate knowledge. This knowledge is a set of the most “representative” tags for an area. They use a TF-IDF approach and present a visualization tool, namely the World Explorer, which allows users explore their results. Serdyukov et al. [50] adopt a language model which lies on the user collected Flickr metadata and aims to annotate an image based on these metadata. The goal herein is to place photos on a map, i.e., provide an automatic alternative to manual geo-tagging. Venetis et al. [56] examine techniques to create a “tag-cloud”, i.e., a set of terms/tags able to provide a brief yet rich description of a large set of terms/tags. They present and define certain user models, metrics and algorithms aiming at this goal. Ye et al. [61] develop a semantic annotation algorithm, which is based on SVM classifiers. They use check-in information from users and extract features from places. Their goal is to determine the probability of each tag for a specific place. Finally, Biancalana et al. [11] deal with personalization aspects through the implementation of a social recommender system involving an experimental empirical framework. It allows users to freely leverage and assign tags, by employing a user-based tag model that derives correspondences between tag vocabularies and folksonomies.

2.4. Exploiting the geographical aspect

Focusing more on the challenging geographical aspect of the problem, Lee et al. [35] created overlapping geographical clusters for each tag, calculated geographical similarity for pairs of tags, and then introduced similarities for both tags and geographical distributions. Rattenbury et al. [49] extract semantics such as places and events from tags and unstructured text-labels, observing that event tags follow certain temporal patterns, while place tags follow certain spatial patterns. In the same manner, Abbasi et al. [1] identify landmarks using tags and Flickr groups, without exploiting geospatial information, aiming to find relevant landmark-related tags, whereas the work presented in [4] analyzes tags associated with geo-referenced Flickr images and uses a TF-IDF approach to generate knowledge as a set of the most “representative” tags for an area. Continuing, Serdyukov et al. [50] adopt a language model which lies on user-collected Flickr metadata and aims to annotate an image based on these metadata and place photos on a map, i.e., provide an automatic alternative to manual geo-tagging. Kennedy et al. [31] could be considered as pioneers into mining popular locations and landmarks from more than 10M Flickr images including metadata like tags, geo-tags and photographers. Their approach, although it performs rather poorly due to subsequent visual clustering steps based on global image features involved, is ideal for constructing tag maps for arbitrary areas in the world. In another interesting approach and in an attempt to combine both geographic and visual clustering worlds, Zheng et al. [63] perform also a similar combination coupled together with an inverse search by travel guide articles containing landmark names, tackling the huge computational cost of their approach by simply utilizing parallel computing in the process. Finally, in another interesting recent work, Stepanyan et al. [52] exploit geo-data in order to semantically annotate places and toponyms in weblog posts.

2.5. Exploiting the chronological aspect

Last thing to consider, one of the challenges when dealing with the chronological aspect in terms of trends' identification in social media, and one we try to partially address within our proposed work, is to automatically detect and analyze the emerging topics (i.e., the ‘trends’) [41]. Most works exploit social and human activity media dynamics to focus on prediction of real-world events and tendencies [6]. Patterns of human attention [58,60], popularity [36,54] and response dynamics [10,16] have been extensively studied in the literature. Recently, researchers investigated temporal patterns of activity of news articles, like Backstrom et al. [9] and Szabo et al. [54], blogposts, like Adar et al. [3] and Mei et al. [42], videos, like Crane et al. [16] and online discussion forums, like Aperjis et al. [5]. Gruhl et al. [22] showed how to generate automated queries for mining blogs in order to predict spikes in book sales. Joshi et al. [27] use linear regression from text and metadata features to predict earnings for movies, whereas Sharda and Delen [51] transformed the prediction problem into a typical classification problem tackled by neural networks to classify movies into meaningful categories. On the other hand, there is also related

[3] http://www.wikipedia.org

research on the interesting problem of general time series clustering, whose main apparatus are typically a distance measure [18] and a clustering algorithm [57]. The most classical distance measure utilized is, of course, the Euclidean distance, although more sophisticated measures such as Euclidean squared distance, Minkowsky distance, or minimax distance may be used, as well. In the clustering domain, there are different types of clustering algorithms for different types of applications and a common distinction frequently used is between hierarchical agglomerative [33] and the K-means clustering [43]. Due to its simplicity and scalability, the latter algorithm introduced many optimizing variants, e.g., fuzzy K-means clustering [30] applicable for the topic at hand.

In this context, our work lines up with related researches on traditional Web search queries that find temporal correlation between social media [2] or queries whose temporal variations are similar to each other [13]. After efficient identification of temporal patterns, one may then focus on optimizing media content management, e.g., in order to maximize clickthrough rates [9], predict news popularity [54] or find topic intensities streams [32]. Still, from all the content that people create and share online within social media, only a few topics manage to attract enough attention to rise to the top [7] and become so-called *temporal trends* which are meaningful to their end-users. In addition to this, there is also little research in the field of identifying and mining such trends from Flickr, where in opposition to social micro-blogging platforms like Twitter, the temporal aspect is not of first level importance to its users.

2.6. Exploiting combined methodologies

Lastly, there exist several works that aim to consider the majority or even all the aforementioned aspects. These works typically aim at tourist applications and focus mainly on (a) the recommendation of places of interest, (b) the automatic discovery of main attractions, which allows users to decide which to visit and also (c) route recommendation algorithms that not only recommend main attractions, but also try to organize the users' schedule and help them visit as many as they wish, using as criteria time efficiency and/or popularity and interestingness. To begin with, Zheng et al. [62] propose a novel navigation system, namely *GPSView*, which is an augmented GPS navigation system that aims to incorporate a scenic factor into the routing. It plans a route not by means of traveling distance and time, but by taking into account certain tourist attractions that may not be visible when passing from the shortest road. Nie et al. [46] proposed a multimedia topic modeling approach which aims to extract venue semantics from heterogeneous location-related user generated contents, which are related by leveraging on multiple data sources. They used a graph clustering method and proposed a semantic based venue summarization approach. De Choudhury et al. [14] constructed itineraries by following a novel two-step approach, as they worked on each city individually and they begun by extracting Flickr photo streams of individual users. They constructed a POI graph, in order to aggregate all the streams of information and to automatically construct itineraries based on the popularity of the POIs and subject to the user's time and destination constraints. Finally, Sun et al. [53] built a recommendation system that aims to recommend to users the best travel routings and also suggest the most popular landmarks. Their dataset was a set of Flickr geo-tagged photos and they used spatial clustering and machine learning methods so as to calculate the popularity of the roads based on the number of users and the number of PoIs. It should be obvious that our current work differentiates significantly from all the above, since we are not focusing on either typical time series clustering methodologies or identifying a unifying global model of temporal variation, but rather explore techniques that allow us to meaningfully quantify what kinds of temporal variations exist on social network user generated content.

3. Processing photo metadata

The main goal of our work is to analyze large user-generated photo collections in order to select the most appropriate meta-tags to describe a geographical area of interest and thus characterize the content itself in terms of its semantics, spatial and chronological context. The very first step in this process would be to cluster available photos, based on their respective geo-information. Thus, in the next subsections we shall initially focus on the notion of a *geo-cluster* and a *(geo-)place* and present the algorithms and techniques we propose in order to perform efficient textual metadata analysis for each geo-cluster. Continuing, we shall present the tag-ranking algorithm we apply on each resulting cluster, as well as our novel proposition to introduce a preliminary intelligent content interpretation step in our approach by re-applying KVQ technique on the latter, so as to obtain a refined set of places. Furthermore, we attempt to understand how, given a set of places, tags and geo-tags associated to them do vary over time. Finally, we first apply a post-processing step to identify semantically meaningful user interests among processed content and then correlate it to time series of corresponding mentions or social interactions.

3.1. Selecting geo-clusters

As in many recent approaches, we choose not to work on the full set of metadata at once, but instead follow a clustering scheme according to their location, i.e., the latitude and the longitude where their corresponding photo has been taken. This location is generally being manually tagged by their owner. However, in some few cases, geo-tagging is automatically added if an appropriate camera/smartphone is used. We should emphasize that in the first case, the accuracy depends on the user's knowledge/memory and is subject to errors. However, in the latter case, the accuracy is higher since it depends on the GPS metadata of the capturing device. Such devices (mainly smart-phones) have become common only during the last few years.

In the following, we shall refer to the clustering procedure as *geo-clustering* and to the resulting clusters as *geo-clusters*. These geo-clusters are allowed to overlap, if necessary, and they do not cover the entire area of interest, e.g., omitting parts where no single photo has been geo-tagged at. Our main objective is to group photos that may have been tagged with semantically similar terms in the same cluster(s). We expect intuitively that photos sharing specific tags should not have been captured at locations very far apart. For example, photos tagged with term *Acropolis* must have been taken within a radius of a few hundred meters. Of course, photos tagged with more “general/vague” tags, like, e.g., *Athens*, are expected to spatially spread over a significantly larger area, thus taken even a few kilometers apart. It is clear that we should select an appropriate clustering algorithm, i.e., an algorithm able to cluster together photos that share certain specific tags. For reasons that will be clarified in the next subsections, in order to perform geo-clustering on a given set of photos, we adopt the *kernel vector quantization* (KVQ) approach of Tipping and Schölkopf [55]. We begin by summarizing the properties of KVQ and present examples of the resulting geo-clusters after its application on a large Flickr photo dataset.

3.1.1. Kernel vector quantization

It is common knowledge that the selection of a cluster analysis approach on a specific dataset is generally subject to the problem at hand. There does not exist an algorithm that may be used in

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

every application and few are the completely automatic algorithms. Thus, in the majority of the cases, one should carefully select both an appropriate algorithm and a distance measure. Many well known algorithms, such as K-means [43], require a user-defined number of clusters and create unequally sized clusters. Accordingly, should we apply in our problem an algorithm such as K-means, in few cases we would expect that geographically adjacent images could easily end up in different geo-clusters, i.e., those that reside near the extracted cluster boundaries.

Based on this observation, and since we require that any two given images, whose distance is below a given threshold (i.e., within a "small enough" radius from a place of interest), to end up in the same geo-cluster. As we shall see, this behavior is guaranteed by the KVQ algorithm. If we consider KVQ as an encoding method, the maximal distance between images belonging to the same cluster may be regarded as the maximum level of *distortion*.

In other words, using KVQ we have a guaranteed *upper* bound on distortion, which corresponds to the radius of the cluster. Accordingly, the number of clusters is automatically adjusted in order to satisfy this property, while also covering the whole dataset.

Vector quantization [21] is a common approach for data compression that uses an appropriate set of vectors in order to model probability density functions. The problem of vector quantization may be formulated as: "Given a set of m data vectors $x_1, x_2, \ldots, x_m \in X$, represent them by a subset of X: $y_1, y_2, \ldots, y_n \in X$, where each x_i is then represented by the nearest y_i, in terms of a pre-defined metric d and so as to minimize a distortion measure". We say that X and d form a metric space, namely (X,d). We say that y_i form a "codebook" and we may formulate the aforementioned error as

$$E = \sum_{i=1}^{l} d(x_i, y(x_i)), \tag{1}$$

where d may be e.g. an L_2 metric function and

$$y(x_i) = \underset{y_i}{\operatorname{argmin}}\ d(x_i - y_j). \tag{2}$$

KVQ chooses to solve the aforementioned minimization problem using linear programming in the following way: Given a generic metric space (X,d) and a finite dataset $D \subseteq X$, $D = x_1, x_2, \ldots, x_m$, the objective is to select a subset $Q(D) = y_1, y_2, \ldots, y_m$ that is as small as possible, under the constraint that all points in D are not "too far" from some point in Q. Let us denote this maximal distance as r. We may then denote an area of X with radius r, centered at x as

$$B_r(x) = \{y \in X : d(x,y) < r\}. \tag{3}$$

This set contains all the data points of X that are not "too far" away from a point x, i.e. their distance is smaller than r. We may now formulate an indicator function $\mathbf{1}_{B_r(x)} : X \to 0, 1$ of set $B_r(x)$ as

$$\mathbf{1}_{B_r(x)} = \begin{cases} 0 & \text{if } y \in B_r(x) \\ 1 & \text{if } y \notin B_r(x). \end{cases} \tag{4}$$

Using $\mathbf{1}_{B_r(x)}$, we are now able to define a kernel function $k : X \times X \to \mathbb{R}$ as

$$k(x,y) = \mathbf{1}_{B_r(x)}(y). \tag{5}$$

This function indicates whether points $x, y \in X$ lie "too far", e.g. more than $r > 0$, when it is equal to 1 and not "too far" when it is equal to 0, in terms of the metric d. We shall refer to r as the *scale* parameter.

Given a point x we may now define the empirical kernel map as $\phi(x) = (K(x_1, x), \ldots, K(x_n, x))^T$. Now we may observe that if there exists a vector $\mathbf{w} \in \mathbb{R}^n$, such that

$$\mathbf{w}^\top \phi(x) > 0, \quad \forall x \in D, \tag{6}$$

then all points $x \in D$ lie within distance r of some point $y_j \in D$, with a corresponding weight $w_j > 0$. The goal is to find a solution that satisfies Eq. (6).

To calculate the optimal solution for $\mathbf{w}$, we would end up to a problem that requires combinatorial optimization. Instead, Tipping and Schölkopf suggest the calculation of a sparse solution $\mathbf{w}^*$, which results by solving a simpler, linear programming problem. Since this problem may result to the determination of clusters that contain the exact same set of vectors with a different center, a pruning step is applied subsequently. We should note that still even after this step, the set of cluster centers will still remain a cover for D.

Given this sparse solution $\mathbf{w}^*$, we are now able to define the resulting codebook $Q(D)$ as

$$Q(D) = \{x_j \in D : w_j^* > 0\}. \tag{7}$$

For the interested reader, more details for the solution of the resulting linear programming problem are given in [55,8].

Now, given a point $x \in D$, we begin by defining a *cluster* $C(x)$ with its center at x as

$$C(x) = \{y \in D : d(x,y) < r\}, \tag{8}$$

or in other words as the set of all points $y \in D$ that lie within distance r from x. It should be clear now that this distance is the aforementioned upper bound on distortion, i.e., any two given points in the same cluster are guaranteed to lie "not farther" than r from the cluster center. By applying KVQ on D, we shall obtain the codebook $Q(D)$, which defines the resulting set of clusters of our interest. As a final remark, $Q(D)$ satisfies the following properties:

(a) $Q(D) \subseteq D$, that is, *codebook vectors* are points of the original dataset. Alternatively, we shall refer to such points as *cluster centers*.
(b) By construction, the *maximal distortion* is upper bounded by r, that is, $\max_{y \in C(x)} d(x,y) < r$ for all $x \in Q(D)$.
(c) The *cluster collection* $\mathcal{C}(D) = \{C(x) : x \in Q(D)\}$ is a *cover* for D, that is, $D = \bigcup_{x \in Q(D)} C(x)$. However, it is *not* a *partition*, as $C(x) \cap C(y) \neq \varnothing$ in general for $x, y \in D$.

The third property denotes that all points of D are contained within the cluster collection, while *clusters may overlap*. The latter observation is very useful for our approach, since it guarantees that images taken within a nearby distance (i.e., lower than r) are never separated. We should finally note that contrary to other clustering techniques in the literature, the number of clusters is automatically adjusted to the user-defined maximal distortion r, so as to cover all images and is not user pre-defined. Such a user-defined distance r is strongly desired in our approach, for reasons that will clarify in the following subsections.

3.1.2. Photo geo-clustering

We now have the theoretical background to continue with our problem, i.e. the application of KVQ on the clustering of a large set of geo-data. Let P denote a set of photos. Then, each photo $p \in P$ may be represented in terms of its geographic coordinates by (p_{lat}, p_{lon}), where p_{lat} and p_{lon} define its geographical capture location, i.e., its *latitude* and *longitude* coordinates, respectively.

Let $\mathcal{P}$ denote the set of all "possible" photos. In order to cluster P in geo-clusters, we apply KVQ in metric space $(\mathcal{P}, d_g)$, where metric d_g denotes the *great circle distance*.[4] We also set r_g as the scale parameter, which we remind that it is the only user-defined parameter of our methodology. In accordance to Eq. (8), given a photo $p \in P$, a geo-cluster, i.e. a cluster of photos may now be defined as

$$C_g(p) = \{q \in P : d_g(p,q) < r_g\}. \tag{9}$$

[4] http://en.wikipedia.org/wiki/Great-circle_distance

This is, the set of all photos $q \in P$ that lie within a circle of radius r_g, centered at p. The resulting codebook $Q_g(P)$ is the set of the photos that correspond to the centers of the geo-clusters. Given $Q_g(P)$, we may define the *geo-cluster collection* as

$$\mathcal{C}_g(P) = \{C_g(p) : p \in Q_g(P)\}. \tag{10}$$

In order to illustrate the above, in Fig. 1 we present a map of Athens depicting all geo-clusters at different zoom levels, for $r_g = 700$ m. We should note the density of photos in the city center (Fig. 1) and particularly in the area of the *Acropolis* (Fig. 1). Photos taken even 1 km apart, e.g., on different sides of a landmark, may be included in the same cluster. The important and novel point of our approach is the fact that the total number and position of produced clusters is automatically inferred solely from the data. In addition, in Fig. 2 we illustrate in detail a randomly picked geo-cluster from the Athens area (depicting also the notion of *places* – to be explained in forthcoming Section 3.3).

3.2. Tag ranking

In this subsection we shall describe the steps towards our second goal, i.e., ranking of tags within extracted geo-clusters. To achieve this, we shall use a probabilistic model on the set of terms that users use to tag their photos (similarly to Serdyukov et al. [50]) and work for each geo-cluster separately, whereas we shall also exploit some global statistical properties of tags in the process. Still, our work clearly differentiates from [50] since we aim to find the most important tags of a geo-cluster, targeting to landmarks, places of interest or even events, while they try to discover the actual location of a photo based on its tags. In the following we initiate the presentation of our approach by modeling the relations among *geo-clusters* and *tags* and we continue by modeling the relations among *clusters* and *users*, as well as among *tags* and their *nearest neighbors*. We conclude by combining our measures into a single one, which we shall use for the selection of the top-ranked tags, a.k.a. the most "important" ones, within each geo-cluster.

3.2.1. Modeling clusters and tags

Like previously, let $P = \{p_i\}$ represent the set of collected photos. The nature of the problem at hand demands that set P should be tagged, geo-tagged and collected from a large region of interest. In our case, where we have chosen online content gathered within Flickr social network, we are able to collect all geo-tagged photos from large urban areas surrounding large European cities, such as Athens, Greece and London, United Kingdom. Following the process already described in Section 3.1.2, we first apply KVQ within such each region, so as to extract a set of geo-clusters denoted as $\mathcal{C} = \{C_j\}$. We then denote by

$$P_j = \{p_i \in P : p_i \in C_j\} \tag{11}$$

the set of all photos taken within geo-cluster C_j. Let T be the set of all tags that accompany all collected photos in our region of interest. For a given set of photos P_k, we will denote the set of all tags these photos have been tagged with, by

$$\mathcal{T}(P_k) = \{t \in T : t \in P_k\}. \tag{12}$$

Then, $\mathcal{T}(P_j)$ is the set of all tags of cluster C_j.

We may now define the probability to obtain a specific geo-cluster C_i, given a tag t_j, as (cardinality of a set is denoted by $|\bullet|$):

$$P(C_i|t_j) = \frac{P(t_j|C_i)P(C_i)}{P(t_j)}, \tag{13}$$

where we calculate the probability of a tag t_j given a specific geo-cluster C_i as

$$P(t_j|C_i) = \frac{|p_j \in P_j : t_j \in \mathcal{T}(p_j)|}{|p_j|}, \tag{14}$$

or in other words as the ratio of the number of all photos of C_i being tagged with t_j to the number of all photos of C_i. Next we define the probability of geo-cluster C_i as

$$P(C_i) = \frac{|P_j|}{|P|}, \tag{15}$$

i.e., as the ratio of the number of photos of C_i to all photos, and the probability of a tag t_j as

$$P(t_j) = \frac{|p_j \in P : t_j \in \mathcal{T}(p_j)|}{|P|}, \tag{16}$$

i.e., as the ratio of the number of photos being tagged with t_j to all photos.

The probability $P(C_i|t_j)$ defined in (13) may be viewed as a means of defining the "importance" of tag t_j for geo-cluster C_i. In other words, tags spread in many geo-clusters will be ranked lower than those unique to C_i. To provide an example from the Athens urban area paradigm, photos tagged with "*Patission*" (a name of a downtown city street which spans across more than one geo-clusters) should be ranked lower to those tagged with "*Polytexneio*" (a script in "greeklish" denoting the National Technical University of Athens, a place of interest located in Patission str.).

3.2.2. Modeling clusters and users

In order to extend the basic approach presented so far, we now take into account the *popularity* of a tag. It should be obvious that tags selected by a large number of users within a specific geo-cluster, i.e., the most "popular" ones, should be ranked above those selected by a small number of users. To make this clear, we should consider a typical case in social media geo-tagging: a single photographer uploads a large number of photos depicting a non-landmark scene (e.g., a friend of his or an animal/pet) and uses the same tag(s) for all. Sometimes the number of such photos may be large enough, biasing those photos to a higher ranking. Since it would not be research-wise to ignore this case in our approach, we extend our baseline method in a way that such photos, i.e., of not significant importance, get a lower ranking. In order to formalize this effect we choose a similar approach to the one of Venetis and al. [56]. Let us first define:

- U as the set of all users,
- U_i as the set of all users whose photos are contained in geo-cluster C_i, and
- U_i^j as the set of all users who have tagged their photos in geo-cluster C_i with tag t_j.

Then, we define the *popularity* (*Pop*) of a tag t_j in geo-cluster C_i as

$$Pop_j^i = \frac{|U_j^i|}{|U^i|}, \tag{17}$$

where U^i denotes users whose photos are contained within geo-cluster C_i.

3.2.3. Modeling tags and their nearest spatial neighbors

One way to select meaningful tags for untagged photos would be to first localize them based on their low-level visual features and then select the most appropriate tags from their most distant and visually similar neighbors [28]. Another one, proposed herein, would be not to exploit any kind of low-level visual information in the process, but exploit the fact that all photos are geo-tagged, thus allowing us to identify the spatial neighbors of each photo. The latter is considered to be a novel, semantically enhanced

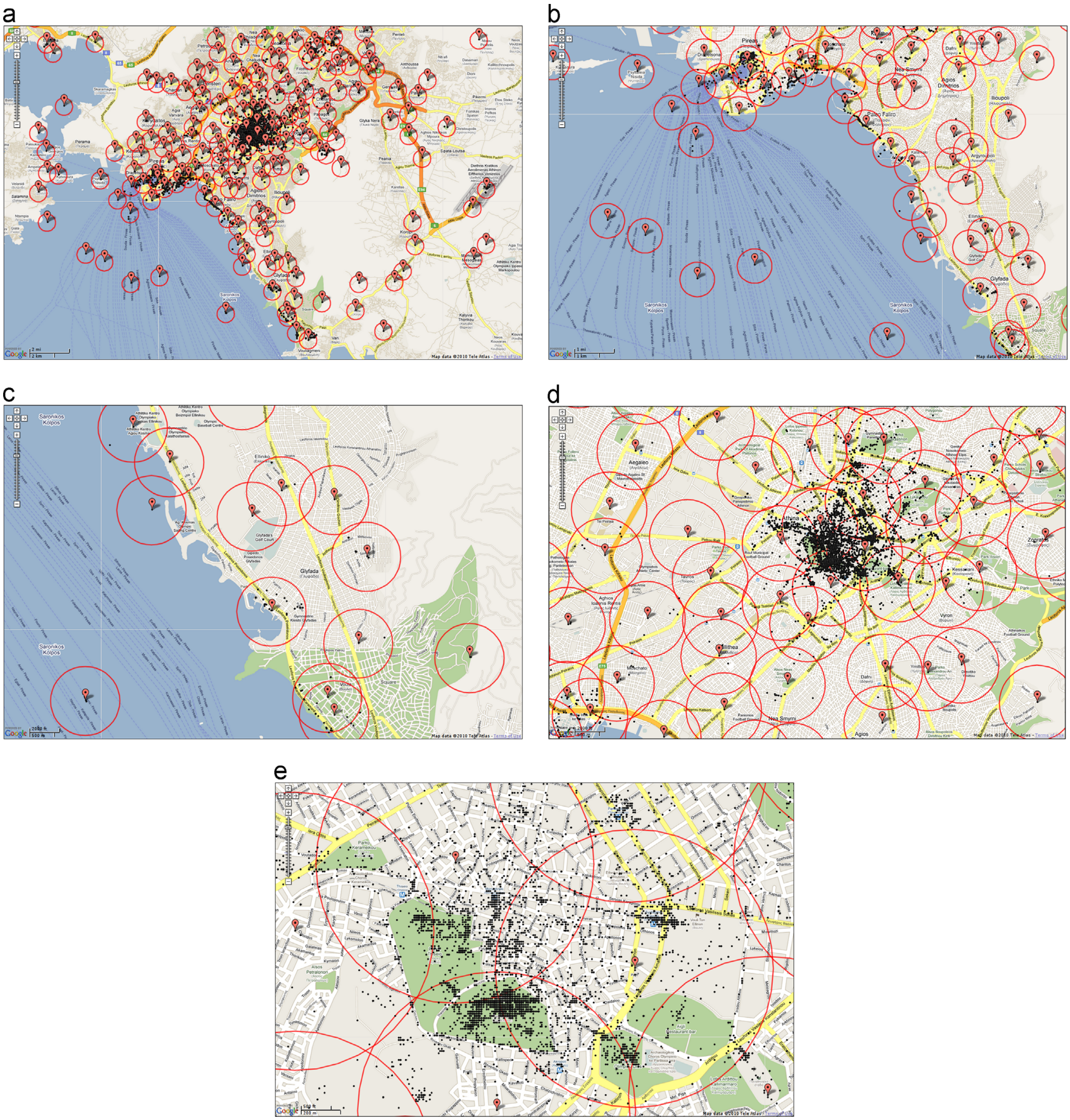

Fig. 1. A map of Athens depicting all geo-clusters. By (black) dots, (red) markers and (red) circles we mark photos, geo-cluster centers and geo-cluster boundaries, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

approach and its first step would be to define a meaningful spatial neighborhood for each photo; let us first define this *neighborhood* $N_i^{D_{\max}}$ of a photo P_i as

$$N_i^{D_{\max}} = \{p_j \in P_i : d_g(p_i, p_j) < D_{\max}\}, \tag{18}$$

where $D_{\max}$ denotes the maximum distance of a given photo to p_i in order to be considered as its neighbor.

Now we are able to define the *influence* of the neighbors as

$$NB_i = \frac{|p \in N_i^{D_{\max}} : t_j \in \mathcal{T}(P_n)|}{|\mathcal{T}(P_n)|}. \tag{19}$$

3.2.4. Combining measures

The last step towards the efficient selection of the most "important" tags within each geo-cluster would be the combination of all three aforementioned modeling relations. Thus, we shall combine the three discussed measures (Eqs. (14), (17) and (19)) and produce a single measure of *importance* R_i^j for a given tag t_j in geo-cluster C_i as

$$R_j^i = P(C_i | t_j) \times Pop_j^i \times NB_i. \tag{20}$$

Within our current work all co-efficients of the above combination step are limited to an equally weighted scheme or in other words

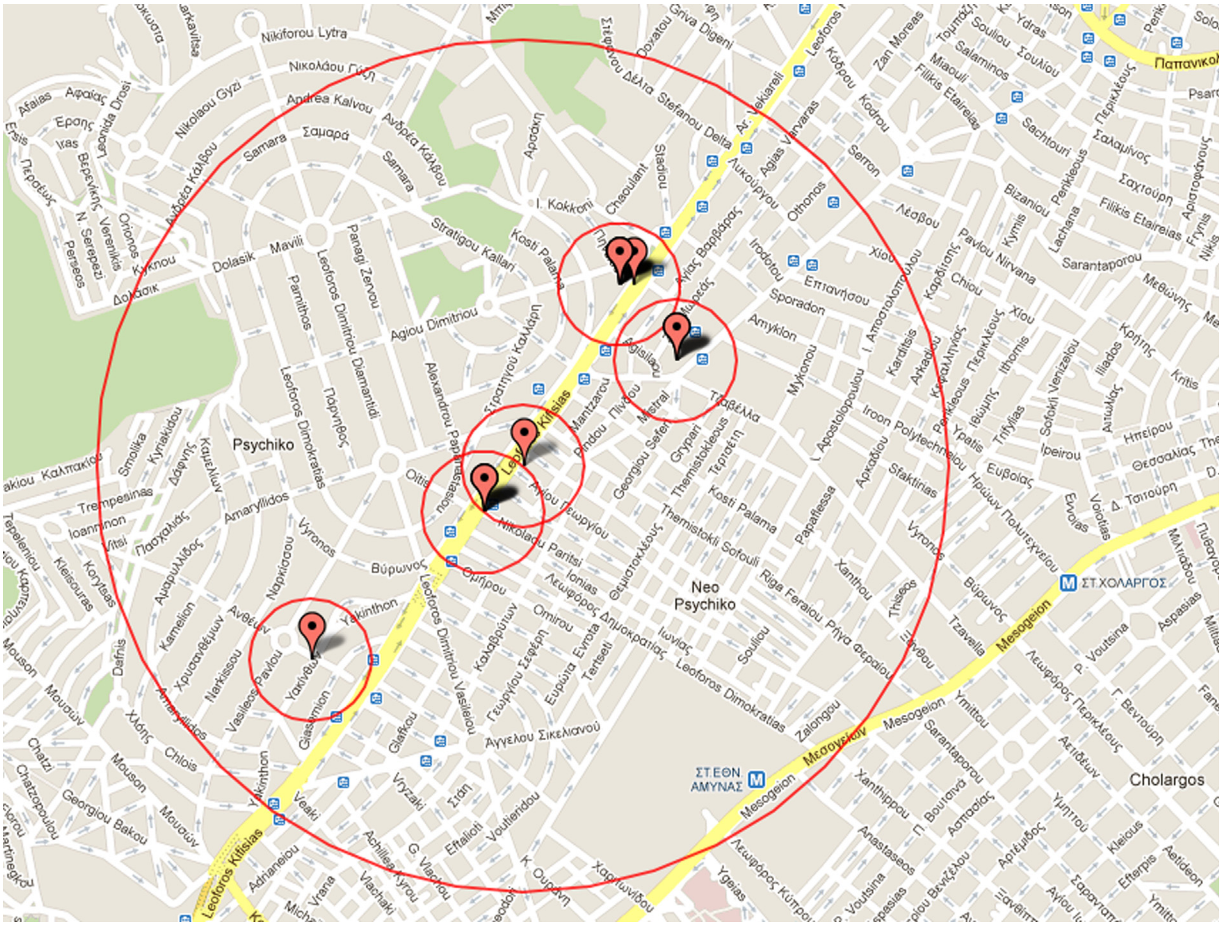

Fig. 2. A geo-cluster and all places extracted within it, using KVQ. Red markers mark photos. The radius of a geo-cluster utilized is $r_g = 700$ m, while the one of a place is $r_p = 100$ m. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

all discussed measures equally affect the construction of importance R_i^j.

3.3. From geo-clusters to places

Continuing and in order to advance our approach from the plain geo-cluster level to the semantically enhanced level of "places", we re-apply initial KVQ clustering step on each geo-cluster. In this case, we select an appropriate percentage of radius r_g of a geo-cluster, namely p_r, as the distortion of clustering. This way, each geo-cluster is further clustered into a second level, which we denote as its corresponding set of *places* (see Fig. 2).

We may now assume that a given place L_j contains a set of photos $P = \{p_i\}$. Let T_p be the set of all tags in this place. For a given set of photos P_k, we will denote the set of all tags these photos have been tagged with, by

$$\mathcal{T}(P_k) = \{t \in T_p : t \in P_k\}. \tag{21}$$

Then, $\mathcal{T}(P_j)$ is the set of all tags of place L_j. For each place and using the dates its photos have been taken, we are then able to create a cumulative distribution of its "popularity" through time, for a given date d, as

$$F_p(d) = |\{p_i \in P : d_i \le d\}|, \tag{22}$$

where again by $|\bullet|$ we denote set cardinality.

3.4. Exploiting data interrelations

At this point we should note that so far our analysis relies on the properties of the content of each place in a separate manner, without taking into consideration its neighbors. Continuing, we shall work on these specific sets of tags and try to understand how tags and geo-tags vary over time, for different types of places. First, we group similar tags, using the well-known and widely adopted Levenshtein distance [37], denoted by d_L in the following. In principle, the Levenshtein distance between two tags is defined as the minimum number of edits needed to transform one tag into the other, with the allowable edit operations being insertion, deletion, or substitution of a single character. We compute this distance for any two given tags t_i, t_j, that are considered similar; the latter are *merged* together, if and only if

$$d_L(t_i, t_j) < T_{lev} \tag{23}$$

where T_{lev} is an appropriately chosen threshold. We treat each such tag group:

$$\mathcal{T}_d = \{t_i, t_k \in \mathcal{T}(P_j) : d_L(t_i, t_k) \le T_{lev}\} \tag{24}$$

as a single tag, which will be denoted as *the "representative" one*. As an additional verification step, we also calculate the cosine similarity of the two given tags t_i, t_j, in order to make the above algorithmic methodology more robust and since cosine similarity measure is often paired with other approaches, to also limit the dimensionality of the problem at hand. Cosine similarity is a common, vector-based similarity measure, whereby the input string is transformed into vector space so that the Euclidean cosine rule can be used to determine the actual similarity. Scoring obtained from the latter procedure is again compared against by an appropriately chosen threshold of 80%. More specifically, the cosine similarity $CS(t_i, t_j)$ between tags t_i, t_j is calculated by

$$CS(t_i, t_j) = \frac{t_i \cdot t_j}{\| t_i \| \cdot \| t_j \|}. \tag{25}$$

We depict an example of the output of the aforementioned process in Table 8. In this table we present groups of the 6 most frequent tags for a place near Acropolis, whereas we have also estimated the most "representative" tag for each group.

The (combined) "representative" tag resulting from the above methodology is considered to be the most frequent tag of the group and "inherits" the dates of all tags belonging to its group. We then calculate the cumulative distribution of each "representative" tag, for a given date d as

$$F_g(d) = |\{t_i \in \mathcal{T}_g : d_i \le d\}|. \tag{26}$$

The above algorithmic analysis is depicted in the following pseudo-code:

Algorithm 1. Calculation of ''representative'' tag.

```
FOR any two given tags from dataset /* Specify
  tags to consider */
CALCULATE their Levenshtein distance d
SELECT appropriate threshold T
IF d < T
GROUP any two tags into new representative tag r,
  if d < T
END IF
FOR the two given tags from dataset
CALCULATE cosine similarity CS
MAINTAIN r if and only if CS score < 80% /* Verify
  rept. tag r */
END FOR
FOR EACH representative tag r
FOR a given date
CALCULATE cumulative distribution F
RETURN F
END FOR
```

3.5. Defining trends

Up to the previous subsection, we have analyzed available user selected tags that characterize each place's photos. By post-processing them, we are allowed to select the most appropriate tags that characterize landmarks and other places of interest that are contained within, as well as events occurring within these places of interest. In order to identify the semantically meaningful "places" of user interest, one should be able to specifically describe the latter. Thus, in the following subsections we shall specifically define (a) *landmarks*, (b) *events* and (c) *places of no-interest* that facilitate the notion of *trends* within our work.

3.5.1. Landmarks

It is well acknowledged in the literature that the most significant places of interest of a given city are denoted by the term *landmarks* and may include buildings, statues, squares, archaeological sites and so on. In other words, landmarks do denote the most popular places of a city for its visitors. Since "popularity" is definitely a vague term, which in turn may not be able to be measured precisely, in this work, we tend to define and estimate it in a threefold sense explained in the following:

First One should expect that a large amount of photos is taken in the close spatial area surrounding a potential landmark.

Second These photos are generally taken by a large number of people (Flickr users, in our case), since landmarks are places of general interest.

Third Since a popular landmark is generally of interest all year, we expect that photos taken thereby should be distributed uniformly through time, or, in general, under a "predictable" distribution.

The latter statement means that, e.g., one should expect an increasing number of photos taken between June and August in Athens (i.e., exactly when tourist season reaches its peak) and a decreasing one between August and April, and so on.

3.5.2. Events

Merriam-Webster dictionary[5] defines *events* as "competitive encounters between individuals or groups carried on for amusement, exercise, or in pursuit of a prize". By the term *events*, in this work, we comply to this definition by considering events like concerts, festivals, musicals, theatrical performances and so on. We also consider athletic events such as a marathon race, a football game or even Olympic Games as a whole. Considering its nature, duration of an event typically varies, still, no periodic behavior is to be anticipated. More specifically, some such as a football game or a concert may last a few hours, while a festival and some athletic events may span across many days. Since events often attract interest, one should expect an increased number of photos during an event's lifetime, concentrated either to a small spatial area, e.g., a football stadium, or to a significantly larger one, e.g., the Marathon route, which extends itself over more than 40 km.

3.5.3. Places of "no-interest"

Since Flickr defines itself as an online media hosting website, it is addressed to all kinds of users and particularly to artists, or tourists. Consequently, one should expect to find many photos of non-landmarks, or non-events in the considered dataset. Typical examples of photos that may be denoted as of "no-interest" for our work may depict a house, a family meeting, an object, an animal/pet, et al.. We make the assumption that these photos have been taken at non-popular places (i.e., at least when considering the notion of popularity from the general public point of view!); such places generally contain less than 10 photos, typically taken by a single user and tagged with the same tags. At this point we should note that this kind of content/photos may also appear in so-called popular places, but due to their limited number, they may be acceptably considered as "noise" and have a limited effect on the overall analysis process.

3.6. Analyzing time

Finally, within this last subsection, we try to identify a time series of mentions or interactions (i.e., through tagging actions) with a particular piece of Flickr photo content for the aforementioned "places". In the current approach this includes a time series of tags of a popular photo, but it could also be further extended to incorporate additional features, like the number of views of a popular photo on Flickr, the number of times a photo was viewed, or the number of times that a popular tag was used to describe a Flickr photo. Of course, the latter demands that all of the above lie

[5] http://www.merriam-webster.com/

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

either within a local (i.e., the specific geo-cluster considered) or overall geo-location of interest (i.e., the overall urban area geo-window considered). The next thing would be to identify patterns within the temporal variation of those time series that are shared by many pieces of such photo content.

More specifically, let us select P content items (photos) and for each photo $p \in P$ let us have a set of outlines in the form $(u_j, z_j)_p$, which means that user u_j mentioned photo p at time z_j. From these P outlines, we then construct a discrete time series $s_p(z)$ by counting the number of tags of photo p at time interval z. Simply, we create a time series of the number of tags of photo p at time z, where z is measured in a meaningful time unit, i.e., hours. Intuitively, s_p depicts efficiently the popularity or attention given to photo p over time. In other words, time series s_p simply represents how the popularity or attention to photo p changed over time and is the one that really interests us. At a later stage, it would then be possible to consider the most popular time series s_p and cluster together photos based on this temporal parameter. The center of each such cluster would be the representative common temporal pattern for the overall group of popular photos, thus reducing the overall information overload.

4. Experimental results

The so far herein described methodology has been applied to a small, yet indicative part of the VIRaL framework,[6] which as of December 2012 includes a total of 2,221,176 digital images collected from 39 cities around the world. In order to collect the entire dataset we have utilized the quite popular Flickr API to design and develop a custom Flickr content batch downloader tool, available to download at http://www.image.ntua.gr/~espyrou/flickrDloadr. The partial VIRaL dataset utilized herein consisted of a total of 160,684 geo-tagged photos collected from two large urban areas surrounding the historical centers of the cities of Athens (18,356 photos) and London (142,328 photos), between years 2004–2009 (both subsets are available online at http://viral.image.ntua.gr/?city&name=Athens and http://viral.image.ntua.gr/?city&name=London).

For each dataset photo we have also downloaded all its available textual and location metadata. An indicative set of typical photos collected from Athens, accompanied with the full set of their textual metadata is illustrated in Table 10. As one may observe, the utilized set consists not only from typical "landmark" photos, i.e., those of tourist interest, but also from "non-landmark" photos, either of personal interest (e.g., deriving from family occasions) or of public interest (e.g., deriving from sport or music events). The locations of these photos span across the geographic areas depicted in Figs. 3 and 4, for Athens and London, respectively. In the following we present our experimental results in a twofold approach, with respect to the extraction of geo-clusters and their meaningful interpretation towards trends' identification.

4.1. Baseline clustering approach

As a first step, we applied KVQ on the geo-data of the Athens and London datasets. The clustering process applied with radius r_g = 700 m produced 193 geo-clusters for Athens (see Fig. 1) and 356 geo-clusters for London. The choice of this value for the scale parameter r_g originates from our previous work [8,28]; results were satisfactory both for the geo-clustering problem and also for a visual clustering problem, where we clustered images based on their visual features.

[6] http://viral.image.ntua.gr

Then, we worked on each geo-cluster separately and collected all tags available from the user-defined metadata. In order to remove "noisy" words, such as the camera model used to take a photo (i.e., tags automatically added in EXIF metadata by certain camera models), we used a manually created stop list for each city. On this set of tags we first calculated all tag frequencies. We then applied the proposed probabilistic model of Section 3.2 first considering only the analysis of Section 3.2.1 and then the effects of both users and nearest neighbors (Sections 3.2.2 and 3.2.3). In the following, we shall refer to this series of experiments as "all tags", "baseline" and "baseline + users + NN", respectively.

After calculating all combined measures presented in Section 3.2.4, we obtained a ranked set of tags for each geo-cluster. The next step was the selection of an appropriate set of tags that describes in the best possible way each geo-cluster. We used a manually defined threshold $T_m = 10$ in order to decide which tags to keep.

Algorithm 2. `KVQ application.`

```
1: Apply KVQ algorithm to cluster dataset
2: FOR each cluster /* Specify radius to be taken
   into consideration*/
3: SET radius r_G returning set of geo-clusters
4: FOR each geo-cluster in geo-clusters
/* Re-apply KVQ for each geo-cluster*/
5: SET radius p_r returning set of places
6: Apply KVQ algorithm
7: RETURN set of places
8: END FOR
9: END FOR
```

4.2. The use case scenario – placing user-generated metadata on a map

In order to evaluate how representative the above datasets are for the content of their corresponding geo-clusters, we present the following simple use case scenario: a user of our VIRaL system visits Athens and/or London as a tourist. This user wishes to discover popular places of interest within a small city region, so as to better plan his available time. Thus, he uses his mobile phone to connect to our system[7] and zooms its map at the appropriate zoom level while centering it at an appropriate position. This could be for example his hotel or his current position. Our system then presents a set of tags. The user may then click on each of them, and is presented with the corresponding set of photos, along with their position on the map. He could then decide which places he wishes to visit. In Fig. 5 we illustrate a map that depicts the most representative tags for an area near *Acropolis*, Athens, Greece and in Fig. 6 an area near *Big Ben*, London, United Kingdom. The font size of each tag is proportional to the measure of importance we calculated as described in Section 3.2.4.

We are now able to evaluate the aforementioned scenario. Since our current goal is to assist users to discover popular places based on the set of tags that our system extracts for each geo-cluster, we choose to evaluate focusing on user satisfaction for the set of tags they have been presented with. We should emphasize that in general, evaluation of tasks aiming at users' satisfaction is known to be a difficult and expensive task, which may involve empirical issues in the process [39]. Having said that, for the sake of evaluating our system, we have conducted a user-centered evaluation by involving 58 real-life users from

[7] http://viral.image.ntua.gr/?mobile

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

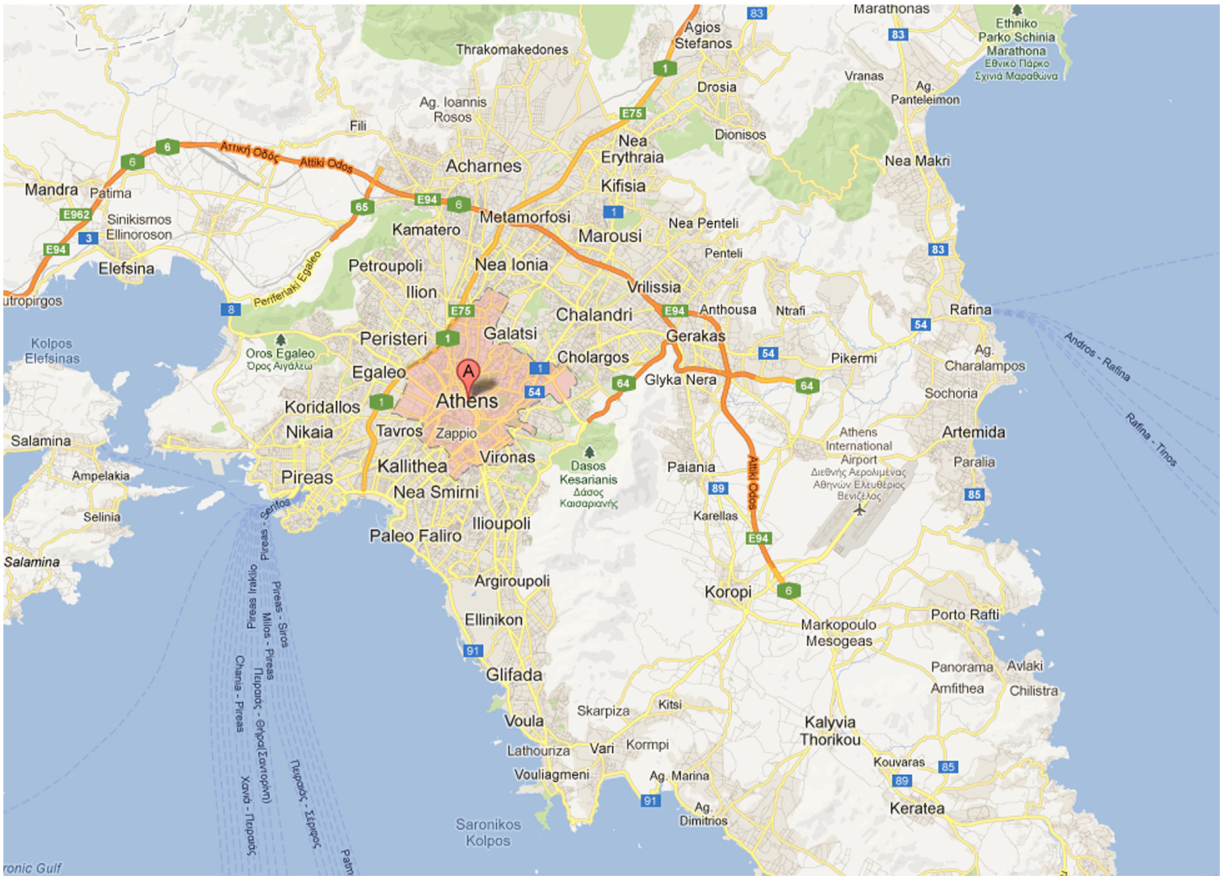

Fig. 3. A map depicting the geographical area surrounding Athens. All geo-tagged photos have been extracted by querying Flickr API with the boundaries of this area.

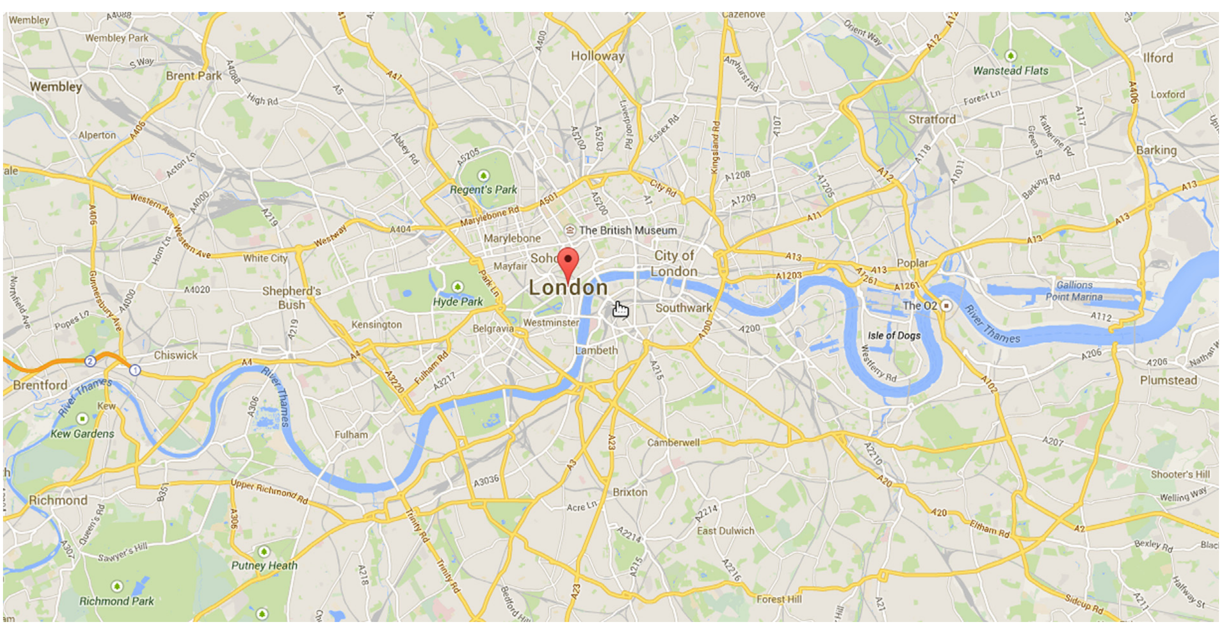

Fig. 4. A map depicting the geographical area surrounding London. All geo-tagged photos have been extracted by querying Flickr API with the boundaries of this area.

four (4) academic institutions.[8] These users were to a great extend familiar to Athens city center (all 58 of them were current or previous local residents) and up to an extend familiar to London city center (45 of them have visited London at least once in their lifetime).

Our first experiment was to present each student photos deriving from 52 Athens geo-clusters, i.e., those nearest to the historical Athens city center, as well as 153 London ones. We then presented them three (3) sets of tags per each:

[8] More specifically, (a) Technological Educational Institute of Central Greece, Lamia, Greece, 23 students, (b) University of Central Greece, Lamia, Greece, 16 students, (c) Ionian University, Corfu, Greece, 7 students, (d) National Technical University of Athens, Athens, Greece, 12 students.

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

Fig. 5. A map of an area near Acropolis, Athens, Greece. For this zoom-level system suggested tags are "*Acropolis*", "*Parthenon*", "*Caryatid*", "*ancient*", "*theatre*".The font-size of tags is proportional to their importance.

Fig. 6. A map of an area near Big Ben, London, UK. For this zoom-level system suggested tags are "*Big Ben*", "*Westminster*", "*tube*", "*bridge*". The font-size of tags is proportional to their importance.

(i) the first consisting from unfiltered tags ranked by their frequency,
(ii) the second by our probabilistic model of Section 3.2.1 and
(iii) the latter by incorporating filtering and re-ranking achieved by modeling of users and nearest neighbors, as described in Sections 3.2.2 and 3.2.3.

We asked the students to select those that describe appropriately the given geo-cluster. As summarized in Table 1, it turned out that in all cases and for both cities students were more satisfied from our system's produced tags.

Our second experiment was to ask them to create a list of 10 tags that best describe each geo-cluster, according to their intuition/experience. We then estimated *precision*, p and *recall* r measures for the sets of tags that our system provides, for both cases (Sections 3.2.1 and 3.2.2 and 3.2.3, respectively). Results provided in Table 2 depict a small, but still measurable improvement in terms of precision over the utilization of the enhanced approach proposed herein (and as expected, recall is slightly worse than the baseline one). It should be noted that the rate at which p increases is much better than the r decreases, resulting in an overall optimized behavior. Thus, we choose to calculate and demonstrate the results of appropriate *F-measures* since these metrics are able to weight in the importance of precision versus recall.

We first estimated the *F-measure* (also known as F_1 score):

$$F_1 = 2 \cdot \frac{p \cdot r}{p+r}, \tag{27}$$

which considers both precision and recall in a moderate manner and can be interpreted as their harmonic mean. We should note that the general formula for the F-measures is given by

$$F_\beta = (1+\beta^2) \cdot \frac{p \cdot r}{\beta^2 \cdot p + r}, \tag{28}$$

for any given real positive β. Consequently, we may also estimate the $F_{0.5}$ measure, as

$$F_{0.5} = 1.25 \cdot \frac{p \cdot r}{0.25 \cdot p + r}. \tag{29}$$

Results are provided in Table 3 and indicate a slight improvement in terms of both F_1 and $F_{0.5}$, as it has been expected, given the results of p, r. We should note that $F_{0.5}$ favors p over r, as it has been previously discussed, and as a result it is the one that represents more accurately the presented approach.

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

Table 1
User evaluation results; percentages indicate users' choice.

	All tags	Baseline	Baseline + users + NN (BuNN)
Athens	5.2%	15.5%	**79.3**%
London	4.9%	17.3%	**78.6**%

Table 2
User evaluation results; Average Precision–Recall values for user generated lists of 10 tags and all geo-clusters.

		Baseline	BuNN	Difference
Athens	Precision	72.41%	75.14%	+2.73%
	Recall	78.22%	76.33%	−1.89%
London	Precision	70.63%	72.93%	+2.30%
	Recall	80.33%	79.34%	−0.99%

Table 3
User evaluation results; *F*-measures values for user generated lists of 10 tags and all geo-clusters.

		Baseline	BuNN	Difference
Athens	$\boldsymbol{F_{0.5}}$ **measure**	73.50%	75.38%	+**1.87**%
	F_1 measure	75.20%	75.73%	+0.53%
London	$\boldsymbol{F_{0.5}}$ **measure**	72.38%	74.13%	+**1.75**%
	F_1 measure	75.17%	76.00%	+0.83%

Table 4
User evaluation results; comparative comparison against custom "Overlap" and "Popularity" metrics ([56]) on the Athens Flickr dataset.

	TF	WTF	POP	COV	Baseline	BuNN
Overlap	0.25	0.24	0.29	0.22	0.23	**0.20**
Popularity	0.76	0.68	–	0.71	0.78	**0.82**

A third comparable experiment was then conducted by taking into account the baseline algorithmic approaches discussed within the similar research work of Venetis et al. [56] for the particular Athens Flickr dataset we utilized. More specifically, a sub-group of 32 students was asked in the process to create a list of 5 tags that best describe each geo-cluster, according to their intuition/experience. We then followed a 3-fold evaluation approach:

(a) we adopted two custom metrics introduced in the aforementioned paper and calculated the actual "*Overlap*" and "*Popularity*" values for all 6 algorithms, namely: both Tf-idf based algorithms (TF, WTF), Popularity algorithm (POP), Maximum coverage algorithm (COV), our Baseline algorithm (Baseline) and our enhanced by users and nearest neighbors approach algorithm (BuNN); all algorithms were implemented and applied on our Flickr dataset,
(b) we calculated Average Precision–Recall values for the user-generated lists of 5 tags and all geo-clusters, and
(c) we estimated $F_{0.5}$ measure values for the user generated lists of 5 tags and all geo-clusters.

Results regarding "*Overlap*" and "*Popularity*" values are depicted in Table 4 and depict that selected tags tend to be more distinctive and more popular in the BuNN case. In addition Table 5 summarizes Average Precision–Recall values for the user generated lists of 5 tags and all geo-clusters for the particular Athens dataset and all 6 algorithms compared, whereas Table 6 depicts the corresponding $F_{0.5}$ measure value with respect to all of them. Results provided in Table 5 depict a clear and significant optimization in terms of precision values over the utilization of the enhanced approach proposed herein compared to the rest of the utilized algorithms. These optimizations range from 2.97% in our Baseline case up to 12.11% in the COV case. Finally, results of Table 6 present $F_{0.5}$ measure improvements up to a 10.23% (COV case), thus demonstrating the numerical contribution of our proposed approach on the algorithmic front.

Table 5
User evaluation results; Average Precision–Recall values for user generated lists of 5 tags and all geo-clusters of the Athens Flickr dataset.

	TF	WTF	POP	COV	Baseline	BuNN
Precision	68.23%	69.79%	65.17%	63.25%	72.39%	75.36%
Recall	75.27%	74.64%	74.33%	75.10%	78.30%	76.25%

Table 6
User evaluation results; $F_{0.5}$ measure values for user generated lists of 5 tags and all geo-clusters of the Athens Flickr dataset.

	TF	WTF	POP	COV	Baseline	BuNN
$\boldsymbol{F_{0.5}}$ **measure**	69.53%	70.71%	66.82%	65.31%	73.50%	75.54%

Furthermore, in order to enhance the herein user-tag evaluation scenario, we presented our students with sets of Athens and London landmark tags. For each set, they were asked to rank the corresponding landmark tags by composing a list according to their preferences, with the most preferred tag placed in the first position and the less appreciated in the last one. While we received their feedback, we calculated two lists of tags, namely:

- a list where landmarks were ordered only according to their appearance frequency on Flickr web-site, and
- a list based on custom user profiles defined a priori for each user/student during a pre-processing phase of the evaluation.

Then, we evaluated how much the feedback proposed by students was similar to the two aforementioned lists in order to find out whether the introduction of user profiles in ranking landmark tags delivers real benefits. To this aim we computed the distance $d(u;y)$ between the list provided by each student and the one returned by Flickr, and then the distance $d(u;s)$ between the list provided by each student and the one ranked by the system based on the user model. To compute the distance between two ordered lists of preferences we utilized the Kendall τ coefficient. If L is the number of preferences that agree and M is the number that disagree, Kendall's τ is defined as [17]

$$\tau = \frac{L-M}{L+M} \tag{30}$$

It should be noted at this point that typically this coefficient varies between 1 (when all preferences agree) and −1 (when they all disagree) and therefore it is ideal for measuring which list is closer to that one expected by the students. In Table 7 we summarize quantifiable results between all lists provided by students for both cities and from the proposed approach. It turns out that our students tend to agree in general with our system's produced tags, as the rather large value of τ in the last column of Table 7 clearly indicates.

4.3. Athens trends

The next step of our proposed approach was to identify the so-called trends. Towards this goal we utilized only the smaller Athens dataset (18,356 images), since we were more familiar and thus more confident with corresponding Greek events happened

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

Table 7
User evaluation results; average τ-values for all pairs of lists, based on Kendall's evaluation metric.

	All tags	Baseline	BuNN
Athens	0.467	0.619	**0.714**
London	0.421	0.664	**0.728**

during the years under consideration (2004–2009). In order to further elaborate on this task, let us present some qualitative characteristics of this dataset. Fig. 7 depicts five distinctive groups of the most popular (i.e., the top-23) tags used by Flickr users to semantically characterize their uploaded content, namely three groups ((a), (b) and (c)) semantically close to landmarks and two tag groups ((d) and (e)) close to events. Each group contains following tags:

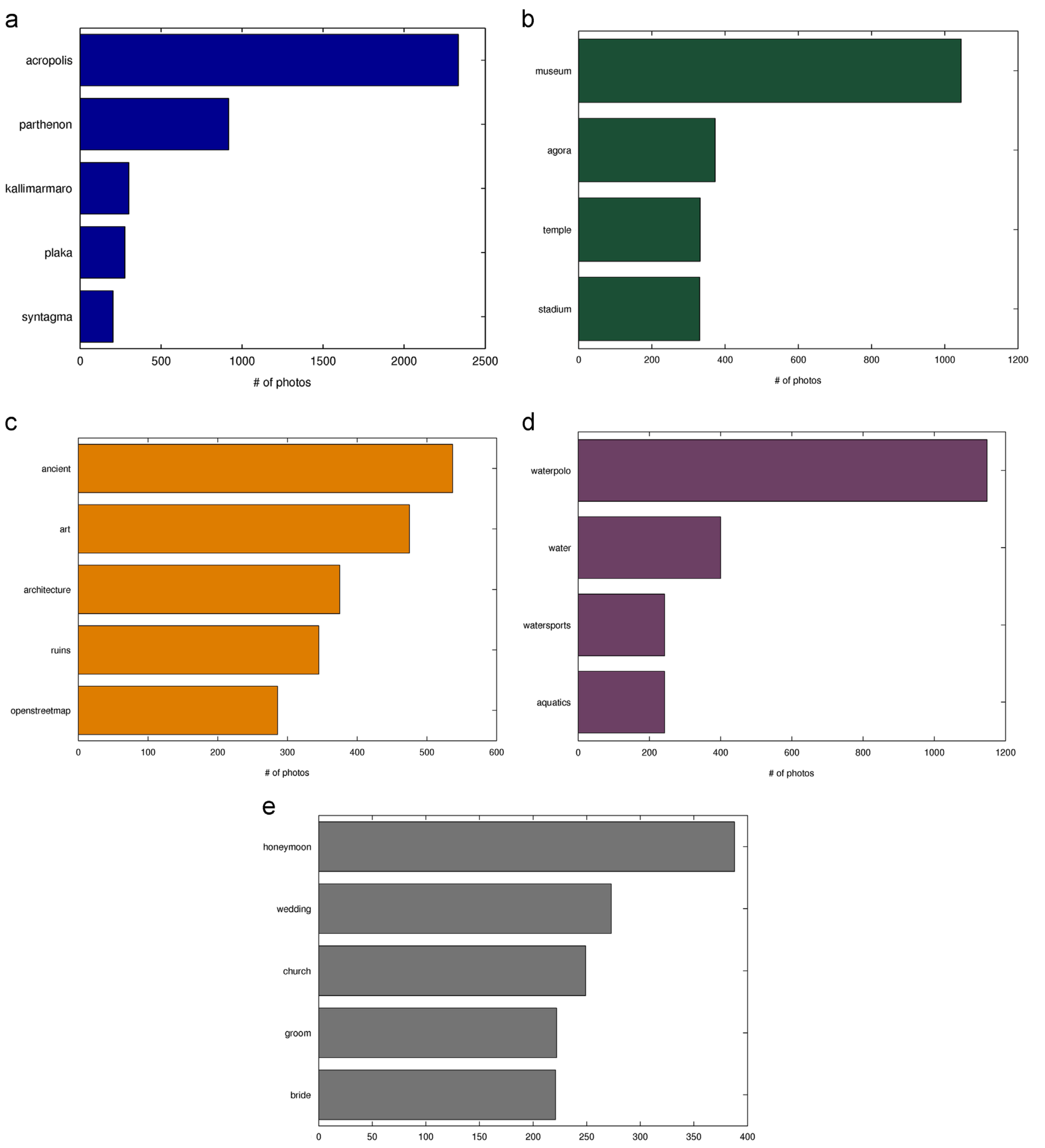

Fig. 7. Top-23 tags derived from the entire Athens urban area distributed to five groups – groups (a), (b) and (c) refer to landmark-related tags, whereas groups (d) and (e) refer to event-related tags.

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

Table 8
Six (6) groups of most frequent tags, for a place near the Acropolis. Representative tags for each group are depicted in bold, whereas Levenshtein metric has been used to implement similarity calculations (see Section 3.4).

Athens	**Acropolis**	**Greece**	**street**	**Parthenon**	**monastiraki**
Atenas	Acropole	Greek	street art	parthenon	
Atene	Acropoli	Grece	street art		
Ateny	Akropoli	Grecia	openstreetmap		
Athen	Akropolis	Grecja			
Athene	acropole	grecia			
Atheny	acropoli	Greeca			
Athina	acropolis	grece			
athen	akropoli				
athena	akropolis				
athenes					
athens					
Aten					

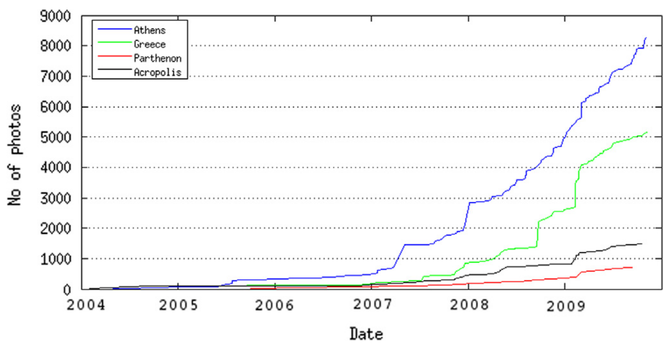

Fig. 8. The cumulative distributions of the most frequent representative tags for a place in the area of Acropolis from 2004 to 2009.

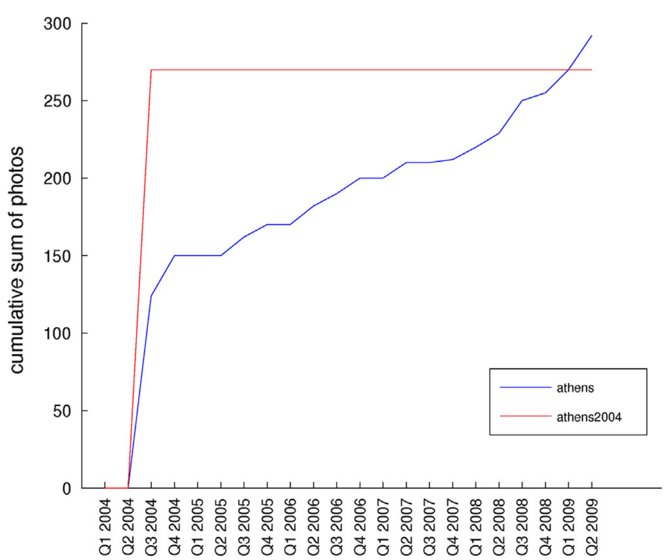

Fig. 9. The cumulative temporal distributions of tags "Athens" and "Athens 2004" for a place in the Olympic Aquatic Center of Athens from 2004 to 2009.

- group A: (acropolis, parthenon, plaka, syntagma, kallimarmaro)
- group B: (museum, agora, temple, stadium)
- group C: (ancient, architecture, art, ruins, openstreetmap)
- group D: (waterpolo, water, watersports, aquatics)
- group E: (honeymoon, wedding, church, groom, bride)

Interpretation of the above statistical information provided us with interesting results, at least from the researcher's point of view. More

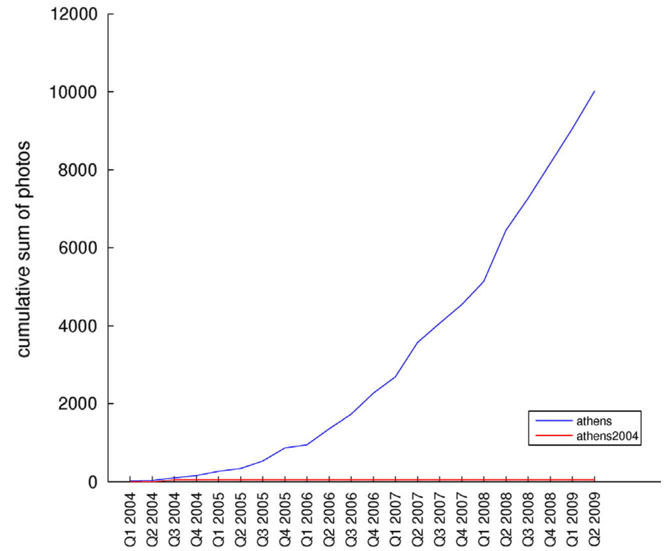

Fig. 10. The cumulative temporal distributions of tags "Athens" (blue solid line) and "Athens 2004" (red step-wise line) tags for the entire Athens photos collection between 2004 and 2009. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

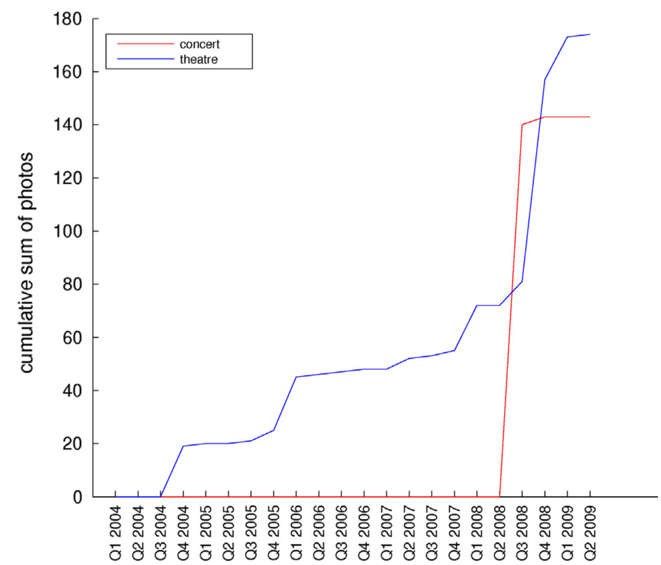

Fig. 11. The cumulative temporal distributions of tags "concert" (red single step-wise line) and "theatre" (multiple step-wise line) tags for the entire Athens photos collection between 2004 and 2009. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

specifically, in Fig. 7 we observe that tag *acropolis* is by far the most popular tag amongst the utilized famous Athens landmarks tag group, whereas the second tag *parthenon* is associated to less than half photos in total. Fig. 7 depicts another landmark-related tag group, namely tags associated to fundamental topological monuments of Athens city center. Again its most popular tag is associated to twice as much photos as the rest of the group. Fig. 7 contains popular generic tags that could have been identified in any considered dataset (i.e., in comparison to Athens-related tags/landmarks of previous groups). As expected, their distribution is rather smooth and no significant outliers are to be identified. In Fig. 7 a list of populated water sports related tags is depicted that characterize the Athens seafront area. To our interest, a specific water-polo event is related to triple as much photos as the rest of them, thus being a clear event outlier. Finally, Fig. 7 depicts popular tags associated to generic types of events in general, and a wedding occasion, in particular. In this case no significant variations are depicted with

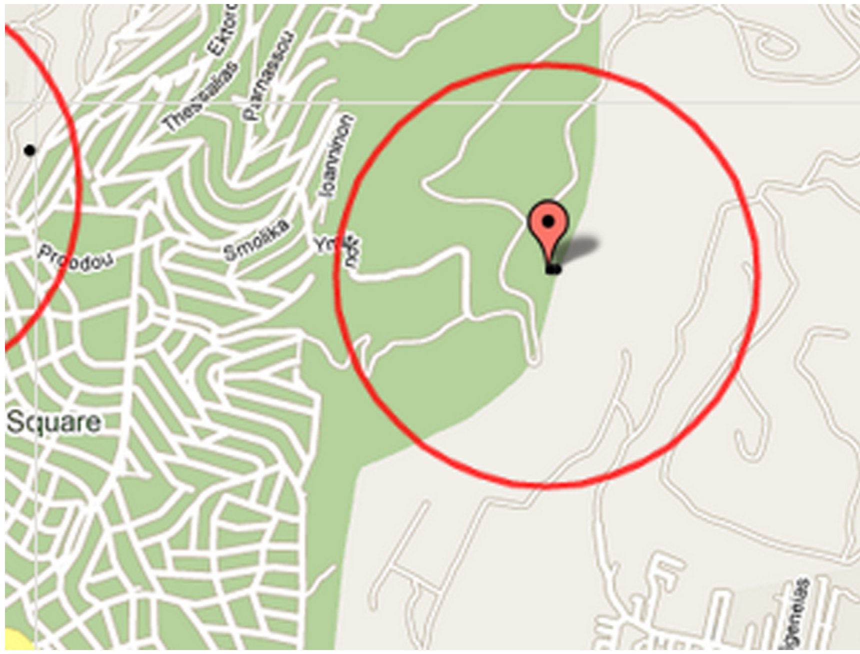

Fig. 12. A geo-cluster of no particular interest with 13 photos, all taken within the same location *Pasxa* denotes Easter in Greek language.

Table 9
Representative tags, for a place of no particular interest.

Cluster ID	# of photos	Tags
15	13	Kitties, barbecue, easter, home, pasxa

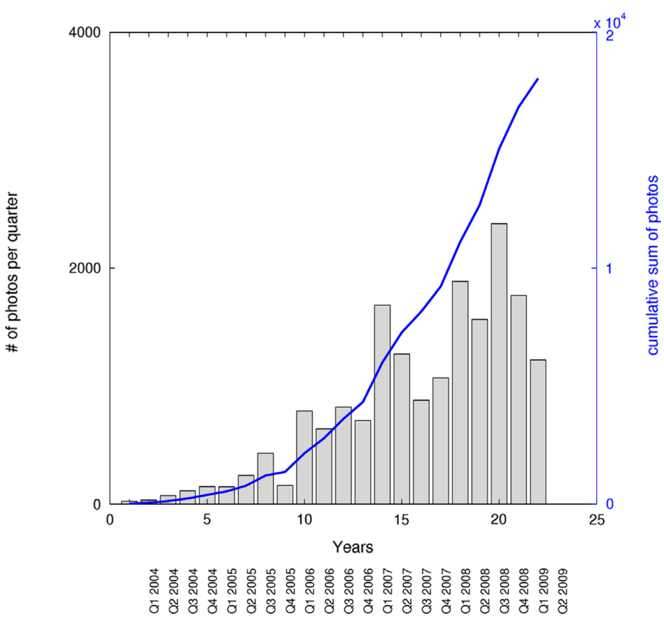

Fig. 13. Number of photos per quarter for the Athens urban area. The blue line depicts the cumulative sum of photos over the entire period. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

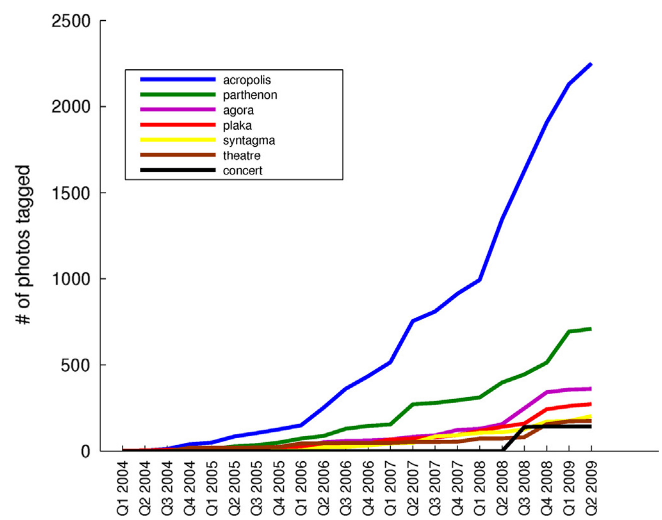

Fig. 14. Number of photos tagged for 7 selected landmark and events tags per quarter for the Athens urban area.

respect to the amount of photos containing this kind of tags. Trying to reason on all the above observations, one might identify the so-called "human factor", i.e., the fact that human content characterization might be quite peculiar, meaningful and in principle unpredictable at the same time. The latter justifies the need for research works like ours, so as to be able to identify user-based semantically meaningful information and understand its importance within the social networks environment.

Now, as mentioned in Section 4.2 and depicted in the respective pseudo-code, the first step of our methodology included the application of KVQ algorithm on to the set of available Athens tags, which resulted to 193 geo-clusters, as we used a radius of $r_g = 700$ m. In order to create a set of places we then re-applied KVQ, this time on each geo-cluster. This time we used a radius of $p_r = 100$ m. The result of this process was the division of each geo-cluster to a number of places. From each one of the 2123 resulting places, we collected all *ranked* tags.

4.3.1. Landmarks in Athens

Further focusing on the three distinct trend types introduced in Section 3.5, there may be no doubt that in Athens, *Acropolis* is by all means the most popular landmark. Thus, in order to illustrate some indicative results of our method when tackling trends within

Table 10
A couple of indicative typical Athens photos accompanied with their textual metadata.

Image	Metadata
	Title: P7190055-1 MediumUrl: `http://farm4.static.flickr.com/3222/2818012761_8a96ece8f5.jpg` PhotoId: 2818012761 Secret: 8a96ece8f5 Date Posted: Mon Sep 01 22:18:15 EEST 2008 Date Taken: Sat Jul 19 23:13:37 EEST 2003 Photo Owner: 10529391@N00,wrd39 Licence: 0 Lat: 37.9719 Lon: 23.7264 Acc: 16 Tags: Athens;Greece;Acropolis;Erechtheion; Place: /Greece/Attiki/Athens/Makrygianni
	Title: P7190044-1 MediumUrl: `http://farm4.static.flickr.com/3169/2818850479_a67b33f339.jpg` PhotoId: 2818850479 Secret: a67b33f339 Date Posted: Tue Sep 02 02:25:12 EEST 2008 Date Taken: Sat Jul 19 22:57:00 EEST 2003 Photo Owner: 10529391@N00,wrd39 Licence: 0 Lat: 37.9712 Lon: 23.7272 Acc: 16 Tags: Greece;Athens;Theater of Dionysus;Acropolis; Archaeology; Place: /Greece/Attiki/Athens/Makrygianni

landmarks, it is a wise choice to select a geo-cluster in the area of Acropolis. This geo-cluster contains 11,298 photos, divided in 19 places. We should emphasize that in this geo-cluster we encounter almost 65% of all geo-tagged photos within Athens. These photos contain a very large number of tags (i.e., more than 100,000).

The analysis of this geo-cluster using the proposed approach results to a set of 2232 representative tags. For illustrative purposes, in Table 8 we present the most popular tags, grouped as described in Section 3.4. One may clearly understand that the more tag variations available (i.e., the larger the associated tag list column is), the more difficult it is to identify the centroid "representative" tag for each cluster. Temporal distributions of certain tags, such as *Athens*, *Greece*, *Acropolis*, *Parthenon* and so on, are in a sense "predictable". These distributions are depicted in Fig. 8. We should note herein that each user has taken 25 (geo-tagged) photos on average.

4.3.2. Events in Athens

Typically an event refers to a specific happening that occurs once at a specific time and place. Hence, given the set of Athens photos, an event satisfied following rules:

- the visual content of its photos should be semantically consistent; since we deal with tags, the latter should be semantically similar
- the group of its corresponding photos should have been taken within a specific time period
- the group of its corresponding photos should have been taken around the same geo-location

Once again and for illustrative purposes we selected a geo-cluster in the Olympic Center of Athens. Following the proposed approach, this geo-cluster was divided in 17 places. We selected tag "Athens 2004" and a group of tags represented by tag "Athens". Their temporal distributions are depicted in Figs. 9 and 10. More specifically, in the first picture, one may observe that tag "Athens" is evenly distributed through time, i.e., such as in Section 4.3.1, while tag "Athens 2004" forms a so-called "spike" in time. The second picture depicts the cumulative sum of photos containing both tags for the entire Athens photos collection; tag "Athens" is steadily increasing, whereas tag "Athens 2004" remains constant. Of course, such results have been expected by the assumptions we have already analyzed, given that the 2004 Summer Olympic Games lasted only for a 2 weeks time-span.

Last, but not least, in Fig. 11 we provide information for two additional events identified in the dataset, namely a single day *concert* event taken place in early 2008 and several theatrical events spread out between 2004 and 2009. We observe that due to the repetitive nature of the theatrical events, (blue) theater line depicts a step-wise performance, i.e., there are several theatrical events scattered around the utilized dataset, whereas the (red) line corresponding to the one-time concert event spikes and remains unchanged over time.

4.3.3. Places of "no-interest" in Athens

In Fig. 1, one could easily notice that in the large urban area surrounding the center of Athens, there exist many geo-clusters which contain only few photos, typically geo-tagged within small distance, i.e., smaller to the radius of a place. Such geo-clusters contain a sole place. We randomly select one of them and depict it

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

Table 11
Symbols utilized throughout the manuscript and their description.

Symbol	Description	Occurrence
$\|\bullet\|$	Cardinality of a set	Entire text
(X,d)	A generic metric space, i.e., a set on which we can measure distances d	Section 3.1.1
D	A finite dataset	Section 3.1.1
$D \subseteq X$	D is a subset of X	Section 3.1.1
$Q(D)$	The resulting codebook after KVQ clustering on D	Section 3.1.1
x	A point in X	Section 3.1.1
y	A point in X	Section 3.1.1
$B_r(x)$	An area of metric space X centered around point x	Section 3.1.1
$d(x,y)$	Distance metric between points x and y	Section 3.1.1
r	Radius of area $B_r(x)$	Section 3.1.1
$\mathbf{1}_{B_r(x)}$	The indicator function of set $B_r(x)$	Section 3.1.1
k	A kernel function	Section 3.1.1
$\mathbf{w}$	A vector in Euclidean vector space $\mathbb{R}^n$	Section 3.1.1
x_j	A point in D	Section 3.1.1
w_j	Corresponding weight of point x_j	Section 3.1.1
$\mathbf{w}^*$	A sparse vector Euclidean vector space $\mathbb{R}^n$	Section 3.1.1
$C(x)$	A cluster centered at point x	Section 3.1.1
$\mathcal{C}(D)$	A cluster collection, being a cover for D	Section 3.1.1
P	A set of photos	Section 3.1.2
p	A photo belonging to the set P	Section 3.1.2
p_{lat}	Latitude coordinate of photo p	Section 3.1.2
p_{lon}	Longitude coordinate of photo p	Section 3.1.2
d_g	The great circle distance metric	Section 3.1.2
$\mathcal{P}$	The set of all "possible" photos	Section 3.1.2
$(\mathcal{P}, d_g)$	A metric space	Section 3.1.2
r_g	A scale parameter	Section 3.1.2
q	A photo belonging to set P	Section 3.1.2
$C_g(p)$	A photo geo-cluster	Section 3.1.2
$Q_g(P)$	The resulting codebook	Section 3.1.1
$\mathcal{C}_g(P)$	The geo-cluster collection	Section 3.1.1
p_i	A photo in P	Section 3.2.1
C_j	A geo-cluster	Section 3.2.1
$\mathcal{C}$	A set of geo-clusters	Section 3.2.1
P_j	The set of all photos taken within geo-cluster C_j	Section 3.2.1
T	A set of tags	Section 3.2.1
P_k	A given set of photos	Section 3.2.1
t	A tag	Section 3.2.1
$\mathcal{T}(P_k)$	The set of all tags P_k photos have been tagged with	Section 3.2.1
$\mathcal{T}(P_j)$	The set of all tags of cluster C_j	Section 3.2.1
C_i	A geo-cluster	Section 3.2.1
t_j	A tag	Section 3.2.1
p_j	A photo in P_j	Section 3.2.1
U	The set of all users	Section 3.2.2
U_i	The set of all users whose photos are contained in geo-cluster C_i	Section 3.2.2
U_i^j	The set of all users who have tagged their photos in geo-cluster C_i with tag t_j	Section 3.2.2
U^i	Users whose photos are contained within geo-cluster C_i	Section 3.2.2
Pop	The popularity of a tag t_j	Section 3.2.2
$D_{\max}$	The maximum distance of a given photo to p_i	Section 3.2.3
$N_i^{D_{\max}}$	Neighborhood of a photo P_i	Section 3.2.3
NB_i	Influence of neighbors	Section 3.2.3
R_i^j	A single measure of importance for a given tag t_j in geo-cluster C_i	Section 3.2.4
p_r	A percentage of radius r_g	Section 3.3
L_j	A given place	Section 3.3
T_p	The set of all tags in L_j	Section 3.3
P_k	A given set of photos	Section 3.3
$\mathcal{T}(P_k)$	The set of all tags these photos have been tagged with	Section 3.3
$\mathcal{T}(P_j)$	The set of all tags of place L_j	Section 3.3
d	A given date	Section 3.3
p_i	A photo	Section 3.3
d_i	A date	Section 3.3
$F_p(d)$	A cumulative distribution of a place's "popularity" through time for a given date d	Section 3.3
t_i	A tag	Section 3.4
t_j	Another tag	Section 3.4
d_L	Levenshtein distance for any two given tags t_i and t_j	Section 3.4
T_{lev}	Levenshtein threshold	Section 3.4
$\mathcal{T}_d$	A group of tags – "representative" tag	Section 3.4
$CS(t_i, t_j)$	Cosine similarity between tags t_i, t_j	Section 3.4
$F_g(d)$	A cumulative distribution of each "representative" tag	Section 3.4
u_j	a user	Section 3.6
z_j	A moment in time	Section 3.6
z	A time interval measured in hours	Section 3.6
$(u_j, z_j)_p$	a set of outlines	Section 3.6
$s_p(z)$	A discrete time series	Section 3.6
s_p	The popularity of photo p over time	Section 3.6
T_m	A threshold on the amount of tags considered	Section 4.2
β	Any given real positive	Section 4.2
$d(u;y)$	Distance between the list provided by each student and the one returned by Flickr	Section 4.2
$d(u;s)$	Distance between the list provided by each student and the one ranked by the system based on the user model	Section 4.2
τ	Kendall coefficient	Section 4.2
L	The number of preferences that agree	Section 4.2
M	The number of preferences that disagree	Section 4.2

in Fig. 12. We depict its corresponding representative tags in Table 9. The observation of these tags indicates that photos within this geo-cluster consist a small photo album of a family gathering during Easter time. Furthermore, it is also identifiable that this set of photos has been taken over a timespan of only 3 days. As a result, one could easily classify this place as of "no-interest" for touristic purposes.

4.3.4. Time series in Athens

In an attempt to illustrate and evaluate the chronological aspect of social media content, we present in Fig. 13 a meaningful distribution of acquired photos with respect to the quarter each photo was taken.

In principle, a calendar year may be divided into four quarters (abbreviated as Q1, Q2, Q3 and Q4), where:

- First quarter (Q1) ranges from January 1st to March 31st
- Second quarter (Q2) ranges from April 1st to June 30th
- Third quarter (Q3) ranges from July 1st to September 30th
- Fourth quarter (Q4) ranges from October 1st to December 31st.

As expected and depicted also by the (blue) cumulative sum line, the amount of online available social media content increases over time following a rather smooth approach. However, this constant content increase includes also periods of low activity, e.g., the period between Q4 2007–Q1 2008 or the third and second quarter of years 2008 and 2009, respectively. On the other hand, it is dominated by high activity peaks, like the second quarter of years 2007 and 2008, as well as the fourth quarter of year 2008, that significantly boost the overall amount of available content.

As a second step, based on our expert knowledge on the Athens urban area, we selected five representative landmark tags, namely: *acropolis*, *parthenon*, *agora*, *plaka* and *syntagma* and two representative events tags, namely: *theatre* and *concert*, and analyzed their distribution over the entire large period of time under consideration. The amount of photos characterized by these semantically important tags over the 22 quarters between years 2004 and 2009 is presented in Fig. 14. For each tag, we have built a time series describing its volume using a quarter time unit. Peak volume of each landmark-related time series is chronologically closer to present time, illustrating the constant growth of online available social media information for this particular type of content (i.e., landmark photos). On the contrary, the repeated theatre event-related time series depicts a step-wise performance, i.e., several theatrical events occurred between 2004 and 2009 and were included in the examined dataset, whereas the single occurrence event *concert* peaked and remained unchanged over time.

5. Discussion and future work

The main conclusion derived from our research involvement in this work is that while mining from user-generated photos within social community network collections is becoming popular and new applications are emerging, several possibilities dealing with the underlying knowledge, intelligence and semantics remain rather untackled. In this paper we have presented our approach in order to manage inherent social media dynamics deriving from such multimedia content and provide results on the meaningful analysis of the latter with respect to specific trends, like content popularity, spatial areas of interest and popular events in time. We have shown that by being able to exploit different kinds of metadata information, such as textual tags, geo-tags, time series, as well plain visual data chunks, we are able to identify meaningful content trends over large periods of time. These trends would facilitate user interaction with generated and already stored content, allowing them to capture landmarks and events of their interest at a glance.

Among our future plans is to further build on the results of this work by studying tag behavior within a given "place" of interest, as well as by semantically combining "places" based on a meaningful semantic popularity criterion to be defined. Another future application would be to semantically explore the identified temporal patterns by which such online user-generated content grows and fades over time, and by which different pieces of content compete for attention. We also plan to further extend our evaluation process by adding another user experiment; we will hide some of the tags, ask users to fill them and compare their answers to those recommended by our system. We believe that herein presented results would have direct application to the optimal placement of user-generated online multimedia content and that our approach would offer a useful starting point for understanding the dynamics in the online social media and how these dynamics of attention may evolve over time for specific topics of interest, like landmarks and events.

Appendix

See Tables 10 and 11.

References

[1] R. Abbasi, S. Chernov, W. Nejdl, R. Paiu, S. Staab, Exploiting Flickr tags and groups for finding landmark photos, Advances in Information Retrieval, Lecture Notes in Computer Science, Volume 5478, Springer, Berlin Heidelberg (2009) 654–661.
[2] E. Adar, D. Weld, B. Bershad, S. Gribble, Why we search: visualizing and predicting user behavior, in: WWW'7, 2007.
[3] E. Adar, L. Zhang, L.A. Adamic, R.M. Lukose, Implicit structure and the dynamics of blogspace, in: Workshop on the Weblogging Ecosystem, 2004.
[4] S. Ahern, M. Naaman, R. Nair, J.H.I. Yang, World explorer: visualizing aggregate data from unstructured text in geo-referenced collections, in: 7th ACM/IEEE-CS Conference on Digital libraries, 2007.
[5] C. Aperjis, B.A. Huberman, F. Wu, Harvesting collective intelligence: temporal behavior in yahoo answers, ArXiv e-prints, January 2010.
[6] S. Asur, B.A. Huberman, Predicting the future with social media, in: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 01, 2010, pp. 492–499.
[7] S. Asur, B.A. Huberman, G. Szabo, C. Wang, Trends in social media : persistence and decay, in: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media 2011.
[8] Y. Avrithis, Y. Kalantidis, G. Tolias, E. Spyrou, Retrieving landmark and non-landmark images from community photo collections, in: ACM Multimedia, Firenze, Italy, October 2010.
[9] L. Backstrom, J. Kleinberg, R. Kumar, Optimizing web traffic via the media scheduling problem, in: KDD 09, 2009.
[10] A.L. Barabási, The origin of bursts and heavy tails in human dynamics, Nature (2005) 207–435.
[11] C. Biancalana, F. Gasparetti, A. Micarelli, G. Sansonetti, Social tagging for personalized location-based services, in: 2nd International Workshop on Social Recommender Systems, 2011.
[12] M. Cha, A. Mislove, K.P. Gummadi, A measurement-driven analysis of information propagation in the Flickr social network, in: Proceedings of the 18th International Conference on WWW, New York, NY, USA, 2009,pp. 721–730.
[13] S. Chien, N. Immorlica, Semantic similarity between search engine queries using temporal correlation, in: WWW 05, 2005.
[14] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, C. Yu, Automatic construction of travel itineraries using social breadcrumbs, in: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia (HT '10), New York, NY, USA, 2010, pp. 761-770. http://dx.doi.org/http://doi.acm.org/10.1145/1526709.1526812.
[15] D.J. Crandall, L. Backstrom, D. Huttenlocher, J. Kleinberg, Mapping the world's photos, in: 18th International Conference on WWW, ACM, New York, NY, USA, 2009, pp. 35-44. http://dx.doi.org/http://doi.acm.org/10.1145/1810617.1810626.
[16] R. Crane, D. Sornette, Robust dynamic classes revealed by measuring the response function of a social system, Proc. Natl. Acad. Sci. 105 (October (41)) (2008) 15649–15653.
[17] B. Croft, D. Metzler, T. Strohman, Search Engines: Information Retrieval in Practice, 1st ed., Addison-Wesley, USA, 2009.
[18] H. Ding, G. Trajcevski, P. Scheuermann, X.Wang, E. Keogh, Querying and mining of time series data: experimental comparison of representations and distance measures, in: VLDB, vol. 1 (2), 2008, pp. 1542–1552.
[19] S. Diplaris, A. Sonnenbichler, T. Kaczanowski, P. Mylonas, A. Scherp, M. Janik, S. Papadopoulos, M. Ovelgonne, Y. Kompatsiaris, Emerging collective intelligence for personal, organisational and social use, Next Generation Data Technologies for Collective Computational Intelligence, Springer, Berlin Heidelberg (2011) 527–573. http://dx.doi.org/10.1007/978-1-4615-3626-0.
[20] S. Gammeter, L. Bossard, T. Quack, L.V. Van Gool, I know what you did last summer: object-level auto-annotation of holiday snaps, in: IEEE 12th International Conference on Computer Vision (ICCV), 2009.
[21] A. Gersho, R.M. Gray, Vector Quantization and Signal Compression, Springer, New York, US, 1992 http://dx.doi.org/10.1007/978-1-4615-3626-0.
[22] D. Gruhl, R. Guha, R. Kumar, J. Novak, A. Tomkins, The predictive power of online chatter, in: SIGKDD Conference on Knowledge Discovery and Data Mining, 2005.
[23] J. Hays, A.A. Efros, IM2GPS: estimating geographic information from a single image, in: CVPR, 2008.
[24] B.A. Huberman, D.M. Romero, F. Wu, Social networks that matter: Twitter under the microscope, First Monday 14 (January (1)) (2009).

Please cite this article as: E. Spyrou, P. Mylonas, Analyzing Flickr metadata to extract location-based information and semantically organize its photo content, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.104

[25] B. Jansen, M. Zhang, K. Sobel, A. Chowdury, Twitter power: Tweets as electronic word of mouth, J. Am. Soc. Inf. Sci. Technol. 60 (11) (2009) 2169–2188.
[26] A. Java, X. Song, T. Finin, B. Tseng, Why we twitter: understanding microblogging usage and communities, in: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, 2007, pp. 56–65.
[27] M. Joshi, D. Das, K. Gimpel, N.A. Smith, Movie reviews and revenues: an experiment in text regression, in: NAACL-HLT, 2010.
[28] Y. Kalantidis, G. Tolias, Y. Avrithis, M. Phinikettos, E. Spyrou, P. Mylonas, S. Kollias, VIRaL: visual image retrieval and localization, Multimedia Tools Appl. 51 (2) (2011) 1–38.
[29] E. Kalogerakis, O. Vesselova, J. Hays, A.A. Efros, A. Hertzmann, Image sequence geolocation with human travel priors, in: ICCV, 2009.
[30] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ, USA, March 2005, http://dx.doi.org/10.1002/9780470316801.
[31] L. Kennedy, M. Naaman, S. Ahern, R. Nair, T. Rattenbury, How Flickr helps us make sense of the world: context and content in community-contributed media collections, ACM Multimedia 3 (2007) 631–640.
[32] A. Krause, J. Leskovec, C. Guestrin, Data association for topic intensity tracking, in: ICML 06, 2006.
[33] M. Kumar, N.R. Patel, J. Woo, Clustering seasonality patterns in the presence of errors, in: KDD 02, 2002.
[34] T. Kurashima, T. Iwata, G. Irie, K. Fujimura, Travel route recommendation using geotags in photo sharing sites, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10, 2010, pp. 579–588.
[35] S.S. Lee, D. Won, D. McLeod, Tag-geotag correlation in social networks, in: ACM Workshop on Search in Social Media, 2008.
[36] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, M. Hurst, Cascading behavior in large blog graphs, in: SDM 07, 2007.
[37] V.I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, Sov. Phys. Dokl. 10 (1966) 707–710.
[38] X. Li, C.G.M. Snoek, M. Worring, Learning tag relevance by neighbor voting for social image retrieval, in: 1st ACM International Conference on Multimedia Information Retrieval, 2008.
[39] Y. Liu, J. Bian, E. Agichtein, Predicting information seeker satisfaction in community question answering, in: SIGIR'08 Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2008, pp. 483–490.
[40] R.D. Malmgren, D.B. Stouffer, A.E. Motter, L.A.N. Amaral, A Poissonian explanation for heavy tails in e-mail communication, Proc. Natl. Acad. Sci. 105 (47) (2008) 18153–18158.
[41] M. Mathioudakis, N. Koudas, TwitterMonitor: trend detection over the Twitter stream, in: Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, 2010, pp. 1155–1158.
[42] Q. Mei, C. Liu, H. Su, C. Zhai, A probabilistic approach to spatiotemporal theme pattern mining on weblogs, WWW 06 (2006).
[43] J.L. Michail, J. Lin, M. Vlachos, E. Keogh, D. Gunopulos, Iterative incremental clustering of time series, in: EDBT, 2004.
[44] P.A. Moëllic, J.E. Haugeard, G. Pitel, Image clustering based on a shared nearest neighbors approach for tagged collections, in: International Conference on Content-based Image and Video Retrieval, 2008.
[45] E. Moxley, J. Kleban, J. Xu, B.S. Manjunath, Not all tags are created equal: learning Flickr tag semantics for global annotation, in: IEEE International Conference on Multimedia and Expo (ICME), 2009.
[46] W. Nie, X. Wang, Y.L. Zhao, Y. Gao, Y. Su, T.S. Chua, Venue semantics: multimedia topic modeling of social media contents, in: Proceedings of PCM, 2013, pp. 574–585.
[47] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comput. Vis. 42 (3) (2001) 145–175.
[48] T. Quack, B. Leibe, L. Van Gool, World-scale mining of objects and events from community photo collections, in: International Conference on Content-based Image and Video Retrieval (CIVR), 2008.
[49] T. Rattenbury, N. Good, M. Naaman, Towards automatic extraction of event and place semantics from Flickr tags, in: 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2007.
[50] P. Serdyukov, V. Murdock, R. Van Zwol, Placing Flickr photos on a map, in: 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2009.
[51] R. Sharda, D. Delen, Predicting box-office success of motion pictures with neural networks, Expert Syst. Appl. 30 (2006) 243–254.
[52] K. Stepanyan, G. Gkotsis, V. Banos, A.I. Cristea, M. Joy, A hybrid approach for spotting, disambiguating and annotating places in user-generated text, in: WWW Companion, ACM, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2013, pp. 83-84.
[53] Y. Sun, H. Fan, M. Bakillah, A. Zipf, Road-based travel recommendation using geo-tagged images, Comput. Environ. Urban Syst. (2013). Available online 13 September 2013, ISSN 0198-9715, http://dx.doi.org/10.1016/j.compenvurbsys.2013.07.006.
[54] G. Szabo, B.A. Huberman, Predicting the popularity of online content, ArXiv e-prints, November 2008.
[55] M. Tipping, B. Schölkopf, A kernel approach for vector quantization with guaranteed distortion bounds, Artif. Intell. Stat. (2001) 129–134.
[56] P. Venetis, G. Koutrika, H. Garcia-Molina, On the selection of tags for tag clouds, in: 4th ACM International Conference on Web Search and Data Mining, 2011.
[57] T. Warren Liao, Clustering of time series data a survey, Pattern Recognit. 38 (11) (2005) 1857–1874.
[58] F. Wu, B.A. Huberman, Novelty and collective attention, Proc. Natl. Acad. Sci. 104 (45) (2007) 17599–17601.
[59] J. Yang, J. Leskovec, Patterns of temporal variation in online media, in: ACM International Conference on Web Search and Data Minig (WSDM), Hong Kong, China, 2011.
[60] S. Yardi, S.A. Golder, M.J. Brzozowski, Blogging at work and the corporate attention economy, in: CHI 09, 2009.
[61] M. Ye, D. Shou, W.C. Lee, P. Yin, K. Janowicz, On the semantic annotation of places in location-based social networks, in: 17th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, ACM, New York, NY, USA, 2011, pp. 520-528, http://dx.doi.org/http://doi.acm.org/10.1145/2020408.2020491.
[62] Y.T. Zheng, S. Yan, Z.-J. Zha, Y. Li, X. Zhou, T.-S. Chua, R. Jain, GPSView: a scenic driving route planner, ACM Trans. Multimedia Comput. Commun. Appl. 9 (February (1)) (2013).
[63] Y. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua, H. Neven, Tour the world: building a web-scale landmark recognition engine, in: CVPR, 2009.
[64] C. Zigkolis, Y. Kompatsiaris, A. Vakali, Information analysis in mobile social networks for added-value services, in: W3C Workshop on the Future of Social Networking, 2009.

Evaggelos Spyrou was born in Athens in 1979. He received his Diploma in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in 2003, specializing in Telecommunications. He was with Image, Video and Multimedia Systems Laboratory (IVML) of NTUA from 2004 to 2010, working as a researcher and developer in many National and European R&D projects. In 2009 he completed his PhD thesis in the field of Semantic Image and Video Analysis. In 2011 he worked as an adjunct lecturer at the University of West Macedonia (UOWM). Since 2011, he works as a scientific associate at the Technological Educational Institute (T.E.I) of Central Greece. His current research interests lie in the areas of semantic multimedia analysis, indexing and retrieval, low-level feature extraction and modeling, visual context modeling, multimedia content representation, neural networks and intelligent systems. He has published research articles in 8 international journals and book chapters and also in 35 international conferences and workshops.

Phivos Mylonas was born in Athens in 1978. He received his Diploma in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in 2001, his Master of Science (M.Sc.) in Advanced Information Systems from the National & Kapodestrian University of Athens (UoA) in 2003 and earned his Ph.D. degree (with distinction) at the former University (NTUA) in 2008. He is currently a senior researcher by the Image, Video and Multimedia Laboratory, School of Electrical and Computer Engineering, Department of Computer Science of the National Technical University of Athens, Greece and has been appointed as an assistant professor by the Department of Informatics of the Ionian University, Greece. His research interests include content-based information retrieval, visual context representation and analysis, knowledge-assisted multimedia analysis, issues related to multimedia personalization, user adaptation, user modeling and profiling. He has published articles in 32 international journals and book chapters, he is the author of 60 papers in international conferences and workshops, he has edited 10 books and is a guest editor of 8 international journals, he is a reviewer for 25 international journals and has been actively involved in the organization of 53 international conferences and workshops. He is a member of the Technical Chamber of Greece since 2001, a member of the Hellenic Association of Mechanical & Electrical Engineers since 2002 and a member of W3C since 2009, whereas he is a past member of IEEE (1999–2010) and ACM (2001–2010).