Chapter 16
Leveraging Massive User Contributions for Knowledge Extraction

Spiros Nikolopoulos, Elisavet Chatzilari, Eirini Giannakidou,
Symeon Papadopoulos, Ioannis Kompatsiaris, and Athena Vakali

Abstract. The collective intelligence that emerges from the collaboration, competition, and co-ordination among individuals in social networks has opened up new opportunities for knowledge extraction. Valuable knowledge is stored and often "hidden" in massive user contributions, challenging researchers to find methods for leveraging these contributions and unfold this knowledge. In this chapter we investigate the problem of knowledge extraction from social media. We provide background information for knowledge extraction methods that operate on social media, and present three methods that use Flickr data to extract different types of knowledge namely, the community structure of tag-networks, the emerging trends and events in users tag activity, and the associations between image regions and

Spiros Nikolopoulos
Informatics & Telematics Institute, Thermi, Thessaloniki, Greece and School of Electronic Engineering and Computer Science - Queen Mary University of London
e-mail: nikolopo@iti.gr

Elisavet Chatzilari
Informatics & Telematics Institute, Thermi, Thessaloniki, Greece and Centre for Vision, Speech and Signal Processing University of Surrey Guildford, GU2 7XH, UK
e-mail: ehatzi@iti.gr

Eirini Giannakidou · Symeon Papadopoulos
Informatics & Telematics Institute, Thermi, Thessaloniki, Greece and Department of Computer Science, Aristotle University of Thessaloniki, Greece
e-mail: {igiannak,papadop}@iti.gr

Ioannis Kompatsiaris
Informatics & Telematics Institute, Thermi, Thessaloniki, Greece
e-mail: ikom@iti.gr

Athena Vakali
Department of Computer Science, Aristotle University of Thessaloniki, Greece
e-mail: avakali@csd.auth.gr

N. Bessis, F. Xhafa (Eds.): Next Generation Data Technologies for CCI, SCI 352, pp. 415–443.
springerlink.com

tags in user tagged images. Our evaluation results show that despite the noise existing in massive user contributions, efficient methods can be developed to mine the semantics emerging from these data and facilitate knowledge extraction.

1 Introduction

Content sharing through the Internet has become a common practice for the vast majority of web users. Due to the rapidly growing new communication technologies, a large number of people all over the planet can now work together in ways that were never before possible in the history of humanity. This user-driven approach is characterized by the fact that its structure and dynamics are similar to those of a complex system, yielding stable and knowledge-rich patterns after a specific usage period [9]. Combining the behavior preferences and ideas of massive users that are imprinted in collaborative data can result into novel insights and knowledge [47], often called Collective Intelligence. Analyzing such data will enable us to acquire a deep understanding of their inner structure, unfold the hidden knowledge and reveal new opportunities for the exploitation of collaborative data.

Collective Intelligence is mainly addressed in Web 2.0 applications, that have experienced an unprecedented information explosion. Social networks, like Facebook, Flickr, and Twitter, enable users to easily create and share information-rich and visually appealing content. Content is also often integrated or reused from other web pages, at the ease of a mouse click. The main vehicles for generating collaborative data in Social Networks are the Social Tagging Systems (STS), which are systems that enable their users to upload digital resources (e.g., bookmarks, pictures, blogs, etc) and annotate them with tags (i.e., freely chosen keywords). An established means of modeling the structure of Collaborative Tagging Systems is the folksonomy model [36] which encodes the associations among the different types of entities (i.e., users, resources and tags) in the form of a network. Based on this model a wide range of techniques have been developed for performing knowledge extraction from social networks.

Extracting the knowledge hidden in Social Networks can help tackle a variety of issues in different disciplines, such as content consumption (e.g., poor recall and precision), knowledge management (e.g., obsolescence, expertise), etc. Several analysis and extraction approaches are being developed towards extracting knowlege from social media. Community detection involves the analysis of a folksonomy with the goal of identifying communities, i.e., groups of objects (which are represented as nodes in the network) that are more densely connected to each other than with the rest of the objects on the network. Similarly, the incorporation of a temporal dimension into the analysis process reveals the macroscopic and microscopic views of tagging, highlights links between objects for specific time periods and, in general, lets us observe how the user tagging activity changes over time. Facilitating the learning process of image analysis models is another use of the knowledge extracted from leveraged user contributed content. All these approaches are motivated by the fact

that the intelligence emerging from the collaboration, competition and coordination among individuals is greater than the sum of the individuals' intelligence.

Numerous applications can be envisaged for exploiting the knowledge extracted from massive user contributions. It is common to derive community-based views of networks, i.e. networks of which the nodes correspond to the identified communities of the original networks and the edges to the relations between the communities. Such views are more succinct and informative than the original networks. It is for this reason that community detection has found applications in the field of recommendation systems [37, 51, 15, 45] , as well as for representing user profiles [1, 20]. Other applications that make use of the knowledge extracted from tag communities include sense disambiguation [2] and ontology evolution/population [51]. Despite the great potential of user contributed content as a source for knowledge extraction, there is a series of challenges involved in such an endeavor. First, the unprecendented growth of user content and associated metadata presents extreme scalability and efficiency challenges to knowledge discovery methods, which so far have been applicable in medium-to-large scale. In addition, the unconstrained nature of uploading and sharing such content has resulted in large amounts of spammy and noisy content and metadata, thus considerably compromising the quality of data to be analyzed. A related challenge stems from the fact that there is currently a variety of metadata associated with online content items; for instance, a photo can be described by a title, a set of tags, and GPS coordinates. However, not all photos consistently contain all of these metadata. Therefore, it is hard to devise sufficiently resilient knowledge discovery and content retrieval methods given that metadata is incomplete or of dubious quality.

Our main objective in this chapter is to demonstrate how the massive user contributions can be leveraged to facilitate the extraction of valuable knowledge. In order to extract the knowledge that is stored and often "hidden" in social data, various approaches have been employed. However, despite the active research efforts in this area, the full potential of Web 2.0 data has not been exploited yet, mainly due to the limitations mentioned earlier. In this chapter we contribute towards overcoming the aforementioned limitations and present three methods for extracting knowledge from Flickr data. A technique for detecting communities in folksonomy-derived tag networks, a time-aware user/tag co-clustering approach which groups together similar users and tags that are very "active" during the same time periods, and a technique that relies on user contributed content to guide the learning process of an object recognition detector. In all cases we use massive amounts of social data and exploit the semantics emerging from their collaborative nature to facilitate knowledge-related tasks.

The remaining of the chapter is organized as follows. In Section 2 we review the related literature with a special focus on the fields related with the presented methods. Sections 3, 4 and 5 are devoted in presenting our methods for extracting knowledge from flickr data and evaluating their results. Concluding remarks and avenues for future research are described in Section 6.

2 Related Work

There is a growing number of research efforts that attempt to exploit the dynamics of social tagging systems, exploit the Collective Intelligence that is fostered by this type of content and facilitate different types of applications. Here, we focus on three research directions that concern the methods to be presented in Sections 3, 4, and 5, respectively. That is, in the following we review the related literature in the fields of tag clustering, temporal tag analysis and using social media to facilitate image analysis. Specifically, emphasis is placed on: *i)* studying the tag clustering problem using *community detection* methods, *ii)* applying temporal analysis on social media for *event identification*, and, *iii)* combining tag and visual information from social media to assist image analysis algorithms.

2.1 Tag Clustering and Community Detection

The problem of tag clustering has recently attracted considerable research interest since it is a challenging task from a data mining perspective, but at the same time it also holds the potential for benefiting a variety of Information Retrieval (IR) applications due to the fact that tag clusters typically correspond to semantically related concepts or topics. For instance, tag clustering is considered important for extracting a hierarchical topic structure from a tagging system in order to improve content retrieval and browsing [7]. Similar conclusions are reached by [5] who point that the use of tags in their raw form limits the potential for content exploration and discovery in an information system; thus, there is a need for an additional level of organization through tag clustering. In [20], tag clusters are used as proxies for the interests of users. Using tag clusters instead of plain tags for profiling user interests proved beneficial for personalized content ranking. An additional application of tag clustering is presented in [2]. There, the clusters were used as a means of identifying the different contexts of use for a given tag, i.e., for sense disambiguation. It was shown that using the tag clusters results in improved performance compared to the use of external resources such as WordNet.

The methods used for performing tag clustering mainly adopt one of two approaches: (a) conventional clustering techniques, such as Hierarchical Agglomerative Clustering (HAC) [7, 20] and (b) Community detection methods [5, 49, 2]. HAC suffers from high complexity (quadratic to the number of tags to be clustered) and the need to set ad-hoc parameters (e.g. three parameters need to be set in the clustering scheme used in [20]). Community detection methods largely address the shortcomings of HAC since efficient implementations exist with a complexity of $O(Nlog(N))$ for finding the optimal grouping of N tags into communities. Furthermore, community detection methods rely on the measure of modularity [38] as a means to assess the quality of the derived cluster structure. Thus, modularity maximization methods do not require any user-defined parameters. However, a problem of modularity maximization methods, also pointed in [49] and confirmed by our experiments, is their tendency to produce clusters with a highly skewed size

distribution. This makes them unsuitable for the problem of tag clustering in the context of IR applications.

2.2 Temporal Tag Analysis

Temporal analysis has been an active topic of research in many disciplines [12, 29]. In Social Tagging environments, where activities are performed in specific temporal contexts, such analysis can be used for extracting knowledge, such as dominant topics over specific time periods, emerging trends, and events that attract users' interest. More specifically, a number of researchers performed temporal tag analysis to locate coherent topics out of unstructured sets of tags in social media and identify "hot" topics that signify emerging trends. In [52] the authors use a statistical model [54], to discover tags that constitute "topics of interest" at particular timeframes. A trend detection measure is introduced in [26], which captures topic-specific trends at each timeframe and is based on the weight-spreading ranking of the PageRank algorithm [6]. The association of tags signified as topics or trends with specific users may be used for extracting user interests in personalized applications [30, 24].

A subdomain of topic detection research involves *event recognition*, that is the analysis of tags/time usage patterns along with geo-related information available in social media, to infer the event semantics of tags. The authors of [42] search for tags in Flickr that can be mapped to events by examining the tags' distribution over time and space. The intuition behind their method is that a tag describing an event usually occurs at a specific segment of time and is assigned on photos geo-tagged around a specific place (e.g., "olympics2008"). In order to capture events of different time scales, they introduce an approach that does not rely on a-priori defined timeframes, but searches for low-entropy clusters in the time usage distribution of a tag that are robust at many time scales. A set of similarity metrics for tracking social media content that is related to events and enable event-based browsing is presented in [4].

Furthermore, the potential of knowledge extraction from social media has been investigated by analyzing the dynamics of these systems and monitoring the activity over time. More specifically, Halpin et al. were the first that introduced the temporal dimension in tag distributions' analysis and presented results for tag dynamics over a dataset from del.icio.us, considering 35 distinct timeframes, [25]. The authors of [61] studied tag recurrence dynamics, by modeling a social media environment as a time-ordered series of posts. The analysis of dynamics of social media shows resemblance with those of complex systems, i.e., a consensus is built incrementally in a decentralized manner, proving, thus, that there is value in analyzing data and extracting knowledge from social media, since these environments are characterized by some kind of stability over time and use. Such techniques may be applied on tag prediction/suggestion approaches.

Finally, temporal analysis can also be used in many applications to illustrate tagging activity in social media with an explicit temporal dimension. In [16] the authors developed a browser-based application in which the user may navigate through

interesting tags of various timeframes in Flickr, at varying timescales. They grasp a tag's interestingness on a particular timeframe by counting its frequency in this timeframe over other timeframes. In order to achieve efficiency, they employ backend algorithms that pre-compute tag interestingness scores for varying sized timeframes. Russell presented a tool that visualizes the collective tagging activity on a resource over time, highlighting periods of stable and changing tagging patterns, [43]. The latter denote a change in users' awareness of the described resource.

2.3 Image Analysis Using Collaborative Data

The works combining user contributed tags with visual features are used to facilitate various tasks, such as image collection browsing and retrieval [3], tag-oriented clustering of photos [22], ranking the results of a video retrieval system [21], or even identifying photos that depict a certain object, location or event [28, 41]. Lately, considerable interest has also been placed on the potential of collaborative data to serve as the training samples for various image analysis tasks. The common objective of these approaches is to compensate for the loss in learning from weakly annotated and noisy training data, by exploiting the massive amount of available samples. Web 2.0 and collaborative tagging environments have further boosted this idea by making available plentiful user tagged data. From the perspective of exploring the trade-offs between analysis efficiency and the characteristics of the dataset, we can mention the works of [27, 13]. In [27] the authors explore the trade-offs in acquiring training data for image classification models through automated web search as opposed to human annotation. The authors set out to determine when and why search-based models manage to perform satisfactory and design a system for predicting the performance trade-off between annotation- and search-based models. In [13] the authors investigate both theoretically and empirically when effective learning is possible from ambiguously labeled images. They formulate the learning problem as partially-supervised multiclass classification and provide intuitive assumptions under which they expect learning to succeed.

Some indicative works that rely on the assumption that due to the common background that most users share, the majority of them tend to contribute similar tags when faced with similar type of visual content include [58, 53, 56]. In [58] the authors are based on social data to introduce the concept of Flickr distance. Flickr distance is a measure of the semantic relation between two concepts using their visual characteristics. The authors rely on the assumption that images about the same concept share similar appearance features and use images obtained from Flickr to represent a concept. The authors present some very interesting results demonstrating that collaborative tagging environments can serve as a valuable source for various computer vision tasks. In [53] the authors make the assumption that semantically related images usually include one or several common regions (objects) with similar visual features. Based on this assumption they build classifiers using as positive examples

the regions assigned to a cluster that is decided to be representative of the concept. They use multiple region-clusters per concept and eventually they construct an ensemble of classifiers. Similarly in [56] the authors investigate non-expensive ways to generate annotated training samples for building concept classifiers using supervised learning. The authors utilize clickthrough data logged by retrieval systems that consist of the queries submitted by the users, together with the images in the retrieval results, that these users selected to click on in response to their queries. The method is evaluated using global concept detectors and the conclusion that can be drawn from the experimental study is that although the automatically generated data cannot surpass the performance of the manually produced ones, combining both automatically and manually generated data consistently gives the best results.

The employment of clustering for mining images of objects has been also explored [28, 41]. In [28] the authors make use of user contributed photo collections and demonstrate a location-tag-vision-based approach for retrieving images of geography-related landmarks. They use clustering for detecting representative tags for landmarks, based on their location and time information. Subsequently, they combine this information with a vision-assisted process for presenting the user with a representative set of images. Eventually, the goal is to sample the formulated clusters with the most represenative images for the selected landmark. In [41] the authors are concerned with images that are found in user contributed collections and depict objects (such as touristic sights). The presented approach is based on geo-tagged photos and the task is to mine images containing objects in a fully unsupervised manner. The retrieved photos are clustered according to different modalities (including visual content and text labels) and Frequent Itemset Mining is applied on the tags associated with each cluster in order to assign cluster labels.

3 Tag Clustering through Community Detection in Tag Networks

The free nature of tagging (no constraints on the tags used, no requirement for expert users) has been responsible for the wide uptake of tagging in numerous web applications. At the same time, such lack of constraints with respect to tagging is the source of numerous annotation quality problems, such as spam, misspellings, and ambiguity of semantics. Coupled with the huge volume of tagging data, these problems compromise the performance (in terms of accuracy) of tag-based information retrieval applications. Given the above observation, tag clustering, i.e. the process of organizing tags in groups, such that tags of the same group are topically related to each other, can provide a powerful tool for addressing the annotation quality and large volume problems that are inherent in real-world tagging applications. Since tag clustering results in a form of semantic organization for the tags of the system, it can be seen as a knowledge organization process. Furthermore, since the extracted tag clusters correspond to meaningful concepts and topics, which are often non-obvious, tag clustering can also be seen as a knowledge extraction process.

There are several approaches for tackling tag clustering. Several works have made use of classic clustering schemes, such as *k*-means [22] and hierarchical agglomerative clustering [7, 20] to discover clusters of tags in folksonomies. According to them, tags are represented as vectors and the employed clustering scheme makes use of some similarity function (e.g. cosine similarity, inverse of Euclidean distance) in order to group tags into clusters. Such approaches suffer from two important limitations: (a) they are hard to scale, since they rely on all pairwise similarities/distances to be computed, (b) they need the number of clusters to be set a priori, which is usually not possible to estimate in real-world tagging systems.

Lately, tag clustering schemes have appeared [5, 49, 2] that are based on community detection in tag networks. Tag networks are very fast to build by use of tag co-occurrence analysis in the context of the tagged resources. Then, community detection methods identify sets of vertices in the networks that are more densely connected to each other than to the rest of the network. The majority of the aforementioned works rely on some modularity maximization scheme [38] in order to perform tag clustering. Modularity maximization methods are reasonably fast (given a network of size m there are methods with a complexity of $O(m \cdot \log m)$) and they do not require the number of clusters to be provided as a parameter. However, such methods suffer from the "gigantic" community problem, i.e. they tend to produce community structures consisting of one or few huge communities and numerous very small ones. In addition, they result in a tag partition, thus assigning every tag to some cluster even in the case that the tag is spam or of low quality.

To this end, we describe MultiSCAN, a new tag clustering scheme that is based on the notion of (μ,ε)-cores [60]. The proposed scheme results in a partial clustering of tags, and distinguishes between tags that belong to a cluster, tags that are associated with many clusters (hubs) and tags that should not be assigned to any cluster (outliers). Furthermore, the proposed scheme addresses an important issue present in the original SCAN scheme [60] that it extends. It does not require setting parameters μ and ε by conducting an efficient parameter space exploration process.

3.1 Description of MultiSCAN

The proposed scheme builds upon the notion of (μ, ε)-cores introduced in [60] and recapped in subsection 3.1.1. Subsequently, it conducts an efficient iterative search over the parameter space (μ, ε) in order to discover cores for different parameter values (subsection 3.1.2). In that way, it alleviates the user from the need of setting parameters μ and ε. An extended variant of this scheme is presented in [39]. The extended version contains an additional cluster expansion step that aims at attaching relevant tags to the extracted tag clusters. Here, we focus solely on the parameter exploration step to study in isolation its effect on the extracted cluster structure.

3.1.1 Core Set Discovery

The definition of (μ, ε)-cores is based on the concepts of *structural similarity*, ε-*neighborhood* and *direct structure reachability*.

Definition 1. *The* ***structural similarity*** σ *between two nodes v and w of a graph* $G = \{V, E\}$ *is defined as:*

$$\sigma(v,w) = \frac{|\Gamma(v) \cap \Gamma(w)|}{\sqrt{|\Gamma(v)| \cdot |\Gamma(w)|}} \tag{1}$$

where $\Gamma(v)$ *is the* structure *of node v, i.e., the set of nodes comprising the node itself and its neighbors:* $\Gamma(v) = \{w \in V | (v,w) \in E\} \cup \{v\}$.

Definition 2. *The* ε***-neighborhood*** *of a node is the subset of its structure containing only the nodes that are at least* ε*-similar with the node; in math notation:*

$$N_\varepsilon(v) = \{w \in \Gamma(v) | \sigma(v,w) \geq \varepsilon\} \tag{2}$$

Definition 3. *A vertex v is called a* (μ, ε)***-core*** *if its* ε*-neighborhood contains at least* μ *vertices:* $CORE_{\mu,\varepsilon}(v) \Leftrightarrow |N_\varepsilon(v)| \geq \mu$.

Definition 4. *A node is directly* ***structure reachable*** *from a* (μ, ε)*-core if it is at least* ε*-similar to it:* $DirReach_{\mu,\varepsilon}(v,w) \Leftrightarrow CORE_{\mu,\varepsilon}(v) \wedge w \in N_\varepsilon(v)$.

For each (μ, ε)-core identified in the network, a new community is built and adjacent nodes are attached to it provided they are directly reachable to it or reachable through a chain of nodes which are directly structure reachable to each other. The rest of the nodes are considered to be *hubs* or *outliers* depending on whether they are adjacent to more than one communities or not. An example of computing structural similarity values for the edges of a network and then identifying the underlying (μ, ε)-cores, hubs and outliers of the network is illustrated in Figure 1. This

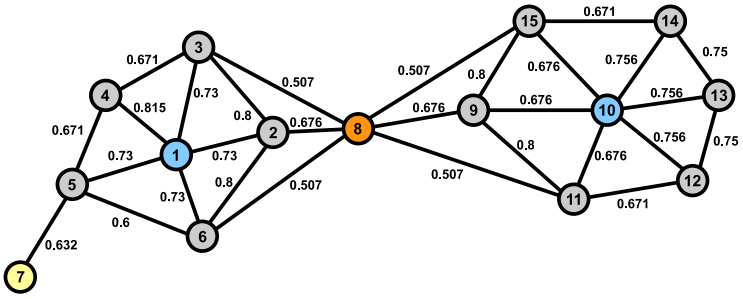


Fig. 1 Example of community structure in an artificial network. Nodes are labeled with successive numbers and edges are labeled with the structural similarity value between the nodes that they connect. Nodes 1 and 10 are (μ, ε)-cores with $\mu = 5$ and $\varepsilon = 0.65$. Nodes 2-6 are structure reachable from node 1 and nodes 9, 11-15 are structure reachable from node 10. Thus, two community seed sets have been identified: the first consisting of nodes 1-6 and the second consisting of nodes 9-15.

technique for identifying communities is computationally efficient since its complexity is $O(\overline{k} \cdot n)$ for a network of n nodes and average degree $\overline{k}$. Computing the structural similarity values of the m network edges introduces an additional $O(\overline{k} \cdot m)$ complexity in the community detection

3.1.2 Parameter Space Exploration

One issue that is not addressed in [60] pertains to the selection of parameters μ and ε. Setting a high value for ε (to a maximum value of 1.0) will render the core detection step very eclectic, i.e. few (μ, ε)-cores will be detected. Higher values for μ will also result in the detection of fewer cores (for instance, all nodes with degree lower than μ will be excluded from the core selection process). For that reason, we employ an iterative scheme, in which the community detection operation is carried out multiple times with different values of μ and ε so that a meaningful subspace of these two parameters is thoroughly explored and the respective (μ, ε)-cores are detected.

The exploration of the (μ, ε) parameter space is carried out as follows. We start by a very high value for both parameters. Since the maximum possible values for μ and ε are k_{max} (maximum degree on the graph) and 1.0 respectively, we start the parameter exploration by two values dependent on them (for instance, we may select $\mu_0 = 0.5 \cdot k_{max}$ and $\varepsilon_0 = 0.9$; the results of the algorithm are not very sensitive to this choice). We identify the respective (μ, ε) cores and associated communities and then relax the parameters in the following way. First, we reduce μ; if it falls below a certain threshold (e.g. $\mu_{min} = 4$), we then reduce ε by a small step (e.g. 0.05) and we reset $\mu = \mu_0$. When both μ and ε reach a small value ($\mu = \mu_{min}$ and $\varepsilon = \varepsilon_{min}$), we terminate the community detection process. This exploration path ensures that communities with very high internal density will be discovered first and subsequently less profound ones will also be detected. In order to speed up the parameter exploration process, we employ a logarithmic sampling strategy when moving along the μ parameter axis. The computational complexity of the proposed parameter scheme is a multiple of the original SCAN (excluding the structural similarity computation which is performed only once). The multiplicative factor is $C = s_\varepsilon \cdot s_\mu$, where s_ε is the number of samples along the ε axis ($\simeq 10$) and s_μ is the number of samples along the μ axis ($\simeq \log k_{max}$). This improves over the original proposal in [40], which requires k_{max} samples along the μ axis.

3.2 Evaluation of Tag Clustering

In order to evaluate the behavior of community detection in real-world tagging systems, we conduct a study comparing the performance of our method (MultiSCAN) against two competing community detection methods on two datasets coming from different tagging applications, namely BibSonomy and Flickr. The first of the two community detection methods under study is the well-known greedy modularity

Table 1 Folksonomy datasets used for evaluation

Dataset	#triplets	U	R	T	$\|V\|$	$\|E\|$	$\overline{k}$	$\overline{cc}$
BIBS-200K	234,403	1,185	64,119	12,216	11,949	236,791	39.63	0.6689
FLICKR-1M	927,473	5,463	123,585	27,969	27,521	693,412	50.39	0.8512

maximization scheme presented by Clauset, Newman and Moore (CNM) [11][1] and the second is the SCAN algorithm of [60], which constitutes the basis for MultiScan. The two datasets used for our study are described below.

BIBS-200K: BibSonomy is a social publication bookmarking application. The BibSonomy dataset was made available through the ECML PKDD Discovery Challenge 2009[2]. We used the "Post-Core" version of the dataset, which consists of a little more than 200,000 tag assignments (triplets) and hence the label "200K" was used as part of the dataset name.

FLICKR-1M: Flickr is a popular online photo sharing application. For our experiments, we used a focused subset of Flickr comprising approximately 120,000 images that were located within the city of Barcelona (by use of a geo-query). The number of tag assignments for this dataset approaches one million.

Starting from each dataset, we built a tag graph, considering an edge between any two tags that co-occur in the context of some resource. The raw graph contained a large component and several very small components and isolated nodes. For the experiments we used only the large component of each graph. Some basic statistics of the analyzed large components are presented in the right part of Table 1. The nodes of the three tag graphs appear to have a high clustering coefficient on average, which indicates the existence of community structure in them. We applied the three competing clustering schemes, CNM, SCAN and MultiSCAN, on the tag graphs and proceeded with the analysis of the derived communities. Since SCAN is parameter-dependent, we performed the clustering multiple times for many (μ, ε) combinations and selected the best solution.

We used the derived tag clusters for tag recommendation in order to quantify their effect on the IR performance of a cluster-based tag recommendation system. More specifically, we created a simple recommendation scheme, which, based on an input tag, uses the most frequent tags of its containing cluster to form the recommendation set. In case more than one tags are provided as input, the system produces one tag recommendation list (ranked by tag frequency) for each tag and then aggregates the ranked list by summing the tag frequencies of the tags belonging to more than one list. Although this recommendation implementation is very simple, it is suitable for benchmarking the utility of cluster structure since it is directly based on it.

[1] We used the publicly available implementation of this algorithm, which we downloaded from `http://www.cs.unm.edu/~aaron/research/fastmodularity.htm`

[2] `http://www.kde.cs.uni-kassel.de/ws/dc09`

The evaluation process was conducted as follows: Each tag dataset was divided into two sets, one used for training and the other used for testing. Based on the training set, the corresponding tag graph was built and the tag clusters based on the three competing methods were extracted. Then, by using the tag assignments of the test set, we quantified the extent to which the cluster structure found by use of the training set could help predict the tagging activities of users on the test set. For each test resource tagged with L tags, one tag was used as input to the tag recommendation algorithm and the rest $L-1$ were predicted. In that way, both the number of correctly predicted tags and the one of missed tags is known. In addition, a filtering step was applied on the tag assignments of the test set. Out of the test tag assignments, we removed the tags that (a) did not appear in the training set, since it would be impossible to recommend them and (b) were among the top 5% of the most frequent tags, since in that case recommending trivial tags (i.e., the most frequent within the dataset) would be enough to achieve high performance.

Table 2 presents a comparison between the Information Retrieval (IR) performance of tag recommendation when using the CNM, SCAN and MultiSCAN tag clusters. According to it, using the SCAN and MultiSCAN tag clusters results in significantly better tag recommendations than by use of CNM across both datasets. For instance, in the FLICKR-1M dataset, the MultiSCAN-based recommendation achieves five times more correct recommendations (R_{TP}) than the CNM-based one ($9,909$ compared to $2,074$). A large part of the CNM-based recommendation failure can be attributed to the few gigantic communities that dominate its community structure. Compared to the best run of SCAN, MultiSCAN performs better in terms of number of unique correct suggestions (U_{TP}) and $P@1$, but worse in terms of precision. In terms of F-measure, SCAN performs somewhat better in both datasets. Given the fact that SCAN requires parameter tuning to achieve this performance and that MultiSCAN provides more correct unique suggestions, we may conclude that the MultiSCAN tag cluster structure is more suitable for the task of tag recommendation.

There are several pertinent issues on the topic that have not been addressed here. First, the tag network creation step can be performed in different ways. Here, we used plain cooccurrence of tags in the context of some resource. There are other tag network creation approaches, such as vector-based tag similarities or tag-focused networks [2]. Depending on the employed tag network creation approach, the produced network will present different characteristics (e.g., edge density) that may affect the subsequent community detection process. An additional issue that we did not address pertains to the existence of multiple scales of community structure in a folksonomy. For instance, a division of a tag network into few large clusters would correspond to a high-level topic description (e.g. “sports”, “politics”, etc.), while a more fine-grained division would discover specific *micro-topics* (e.g. “firefox plugins”, “brownies recipe”). Instead, most community detection methods (including the one presented here) discover a single configuration of nodes into communities that is more “natural” given the properties of the tag network. The optimal scale of community structure depends on the information retrieval problem at hand. For

Table 2 IR performance of CNM, SCAN and MultiSCAN community structures in tag recommendation. The following notation is used: R_T denotes the number of correct tags according to the ground truth, R_{out} the number of tag suggestions made by the recommender, R_{TP} the number of correct suggestions, U_{TP} the number of unique correct suggestions, P, R, and F stand for precision, recall and F-measure respectively, and $P@1$, $P@5$ denote precision at one and five recommendations respectively.

	BIBS-200K			FLICKR-1M		
	CNM	SCAN	MultiSCAN	CNM	SCAN	MultiSCAN
R_T		15,344			57,206	
R_{out}	15,271	4,762	7,346	57,021	19,063	33,714
R_{TP}	377	**979**	2,545	2,074	9,781	**9,909**
U_{TP}	196	588	**705**	263	1,103	**1,437**
P (%)	2.47	**20.56**	13.10	4.46	**51.31**	29.39
R (%)	2.46	**6.38**	6.27	4.45	17.10	**17.32**
F (%)	2.46	**9.74**	8.48	4.46	**25.65**	21.80
$P@1$ (%)	2.54	2.97	**5.03**	1.89	5.03	**10.09**
$P@5$ (%)	2.39	**26.36**	19.94	3.04	**46.30**	34.09

instance, as was observed in our experimental study, the existence of large communities harms the performance of a cluster-based tag recommender.

4 Time-Aware Tag/User Co-clustering

The ability to capture emerging trends and dominant topics over time is another form of challenge that could be addressed by data mining approaches in social media content. Indeed, as more and more people tend to express themselves through tagging in social media environments on a daily basis, it can be drawn that monitoring these systems over time allows us to watch the evolution of community focus. Therefore, analysis of such content within its temporal context enables knowledge extraction regarding real world events, long-term or short term topics of interest, and trends. Difficulties arise, though, from the fact that the knowledge extracted from this kind of analysis is particularly sensitive to the time-scale used. For example, the tag `Olympics2008` does not appear to be an event at the hour or single day scale, but does exhibit distinctive temporal patterns at larger time scales. The approach presented in this section overcomes this concern by defining a time-aware co-clustering method that can be applied at multiple time-scales, τ.

4.1 The Proposed Framework

The knowledge extraction approach we propose here is based on the analysis of both users' and tags' activity patterns. The patterns are extracted from two sources of information: i) the meaning of the tags used, and, ii) the time period each activity

occurs. The intuition behind this decision is as follows. The social media associated with an event mainly exhibit similarity in terms of their tags and their time locality. Likewise, the users that are attracted by an event or trend tend to use tags related to this incident during the time it is happening. In this context, we follow the following assumption:

> *An **event** or **trend** can be tracked in a social media environment as a dense cluster that consists of two types of objects: related tags with frequent patterns of usage in a given period, and, many users that use these tags around the same period.*

To materialize this observation, a co-clustering method is utilized that employs the time locality similarity and yields a series of clusters, each of which contains a set of users together with a set of tags. Co-clustering is proposed as an approach which may be applied in grouping together elements from different datasets [14]. In our case, co-clustering is used to relate tags and users. In an effort for the clusters to better reflect user choices at particular time intervals, our approach examines tag-based similarity, as well. To examine tag-based similarity, we use the *Wu & Palmer* metric [59], which is based on WordNet to evaluate the similarity in meaning between two terms [18], as follows:

$$TagSim(u_x, t_y) = \max_{t_z} \frac{2 \times depth(LCS)}{[depth(t_z) + depth(t_y)]}, \tag{3}$$

$\forall t_z$ assigned by u_x, where $depth(t_x)$ is the maximum path length from the root to t_x and *LCS* is the least common subsumer of t_x and t_y.

To quantify the locality in the temporal patterns between a user and a tag at a given timescale τ, we divide the entire time period into I sequential timeframes of size τ and represent: i) each user as $u_x = [u_{x1}, u_{x2}, \ldots, u_{xI}]$, where u_{xj} is the number of tags user u_x has assigned during the timeframe j, and ii) each tag as $t_y = [t_{y1}, t_{y2}, \ldots, t_{yI}]$, where t_{yj} is the number of times the tag t_y has been used during the timeframe j. Then, we calculate the similarity between any two user or tag vectors, by taking their inner product:

$$TimSim(u_x, t_y) = \frac{\sum_{k=1}^{I} u_{ik} \cdot t_{jk}}{\sqrt{\sum_{k=1}^{I} u_{ik}^2 \cdot \sum_{k=1}^{I} t_{jk}^2}}, \tag{4}$$

Having calculated temporal and tag-based similarities between users and tags, we compute the dot product of *TagSim* and *TimSim* between any two objects, in order to get a matrix that embeds both kinds of similarities between users and tags:

$$Sim = TagSim \bullet TimSim, \tag{5}$$

Given *Sim*, we may proceed with the application of the co-clustering algorithm [14], in order to get clusters containing users and tags with similar patterns over time. The applied algorithm is based on the spectral clustering theory, as discussed in [23, 29], and relies on the eigenstructure of the similarity matrix, *Sim*, to partition users and tags into K disjoint clusters. The steps of the applied spectral clustering

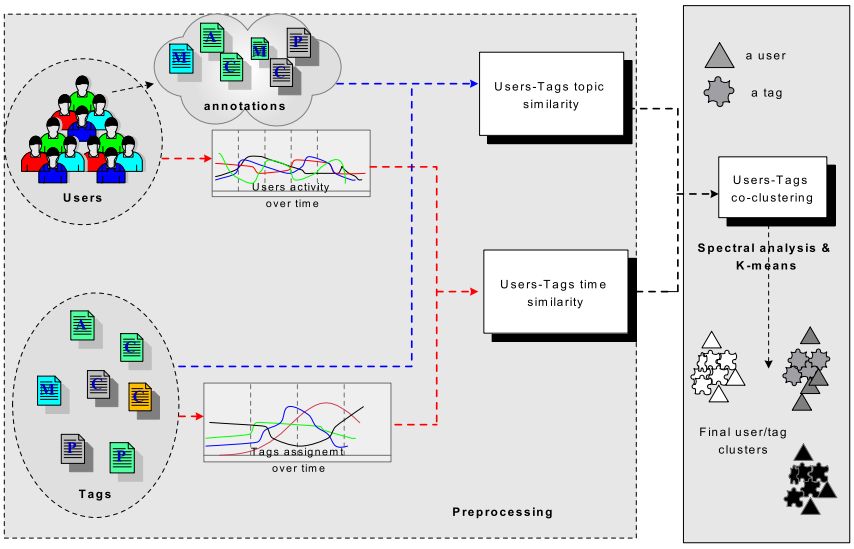


Fig. 2 The proposed time-aware co-clustering algorithm overview

algorithm, which are illustrated in Figure 2, are: i) normalization, ii) computation of eigenvalues, and iii) eigenvector clustering, using K-means.

4.2 Evaluation of Time-Aware User/Tag Co-clustering

We tested our method on a Flickr dataset of 1218 users, 6764 photos, and 2496 unique tags that span the time period from Sep. 2007 to Sep. 2008. To examine the method's applicability in tracking time-related happenings, e.g., events or trends, we used the following four seed tags that are associated with many real-world events, to create the dataset: `Olympics`, `earthquake`, `marriage`, `ancient greece`. The input parameters used are the cluster number K and the time scale τ.

First of all, we aim at studying the impact of the proposed similarity metric on capturing trends or events, in comparison with other similarity metrics. As suggested by the assumption presented in Section 4.1, we examine the compactness of the extracted clusters in terms of gathering together objects that have tag-based and temporal similarity. In order to check this, we performed a rough annotation of our dataset as follows. We assumed four thematic ranges in the dataset, each one associated with one seed tag. Then, we record the activity in time of both tags and users. We divide the time period in timeframes of duration τ. If the activity of an object in a timeframe is above a certain threshold ϑ, we assume that an event or trend is associated with this activity. Thus, a number of events or trends are generated. The value of the threshold ϑ at each time scale is defined empirically. Then, each object (i.e. user or tag) is assigned to the event or trend in which it was more active and had the closest proximity in time. Thus, a ground truth of our dataset is created.

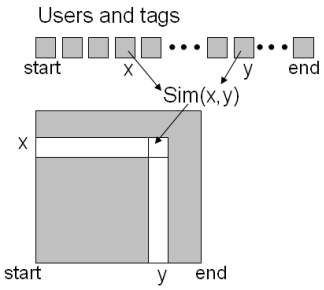
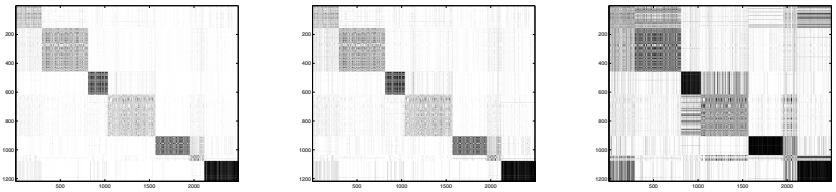


Fig. 3 Tag/user similarity matrix

Then, to evaluate the performance of the proposed similarity metric, we compute a similarity matrix for the 1218 users and 2496 tags using both tag-based and temporal similarity as described in Section 4. The matrix is filled by calculating the similarity between every pair $< user, tag >$. Specifically the (i, j) element of the matrix quantifies the similarity between the i^{th} and the j^{th} object, as depicted in Figure 3. Then, the matrix is reordered, so that objects that have been assigned to the same event or trend during the ground truth generation are contiguous (in rows and in columns). The darker the coloring of a cell (i, j) where $1 \leq i, j \leq |U| + |T|$ the more similar the objects at position (i, j) are. Thus, clusters appear as symmetrical dark squares across the main diagonal. A checkerboard pattern of the described similarity matrix across the main diagonal indicates good clustering, whereas grey rectangles across the diagonal and dark or grey areas outside the diagonal imply that the similarity metric used in the clustering process does not capture the objects assigned to the same trend or event in the same cluster.

In the same way we created a similarity matrix solely based on the temporal locality of objects and a similarity matrix solely based on the tag-based similarity. We conducted experiments for various values of τ. For each τ, we selected the value of K based on the ground truth generation. In Figure 4 we indicatively present the clustering outline for $K = 7$ and $\tau = 10$, in these three different cases. Particularly, the plot shown in (a) was extracted from the proposed similarity metric, while the plots in (b) and (c) were derived using the temporal and the tag-based similarity metric, respectively. It is obvious that the combination of both temporal and tag-based features in the similarity metric succeeds in finding more coherent clusters that according to our original assumption can be mapped to events or trends. The



(a) clustering based on temporal and tag-based features

(b) clustering based on tag-based features

(c) clustering based on temporal features

Fig. 4 Events or trends capturing(darker blocks indicate better capturing). All similarity matrices are ordered according to ground truth

coherence deteriorates in case we use only tag-based or temporal similarity between objects.

Next, we want to show that the proposed method is sensitive to various values of τ and performs knowledge extraction at various time scales. While the overall analysis on the entire dataset facilitates the extraction of massive activities, such as events or trends, the analysis at a user level allows the extraction of long-term or short-term user interests and the inclination of that user to follow certain trends or events. Figure 5 illustrates the tagging activity of three users during a yearly period (solid curves). The macroscopic tag clouds indicate each user's most frequent tags during the entire time period, while the microscopic tag clouds reflect each user's most frequent tags in specific timeframes. Given the little overlap in the users' macroscopic tag clouds and their differentiated tagging curves, one would expect that these three users would not be grouped together, if their similarity is evaluated

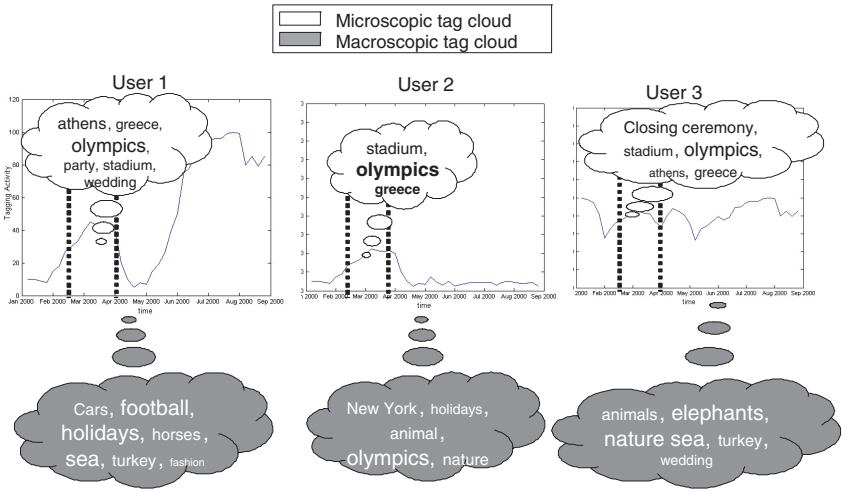
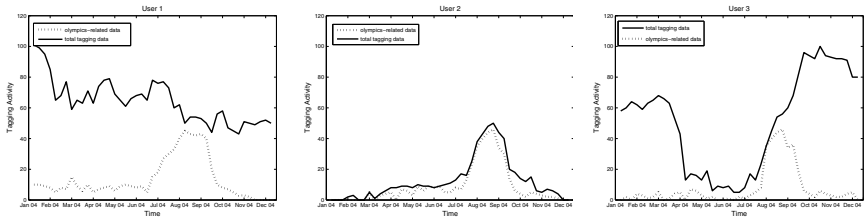


Fig. 5 Tag clouds of three users in an STS



(User1) (User2) (User3)

Fig. 6 Tagging activity of users over time

solely on the similarity of their associated tags. However, focusing on short-term views of the users' tagging activity and examining their microscopic tag clouds at monthly timescale, we observe that for the time interval highlighted with dotted lines in Figure 5 the similarity in these users' tags is very high, as they all use many "Olympics" related tags. This indicates a simultaneous preference by these users for this particular topic and at this specific time period. At the same time, this massive tagging preference may imply that a related event occurred at that period and attracted Olympics-friends to comment on it, through tags.

The users' similar Olympics-related tagging activity is highlighted by the dotted curves in Figure 6, which displays the usage of "Olympics" related tags of each user over time. We claim that such user groups, exhibiting both semantic and temporal cohesion, can only be extracted via a time-aware clustering method, which will examine user behavior at varying time scales. Each time scale selection reveals a different micro-view of users' interests that affects the current clustering, since the microscopic tag cloud of each user is likely to change as the selected time interval's length τ slides across the timeline.

To summarize, in this section, a technique was presented that performs temporal analysis on social media and is based on co-clustering tags and users by considering jointly their temporal and tag-based similarity. The extracted clusters may be used for event or trend recognition and for capturing users' interests at different timescales. An evaluation based on generated ground truth from a Flickr dataset demonstrates that the proposed framework performs better in tracking events or trends than other methods that consider solely tag-based or temporal locality. A number of applications can benefit from such a technique. For example, Olympics-related clusters can be exploited by a sports commercial advertising campaign or be embedded in an application, so that users receive personalized Olympics-related news (e.g., announcement of upcoming events).

5 Enhancing Image Analysis Using Collaborative Data

Semantic object detection is considered one of the most useful operations performed by the human visual system and constitutes an exciting problem for computer vision scientists. Due to its fundamental role in the detection process, many researchers have focused their efforts on trying to understand the mechanisms of learning and particularly the way that humans learn to recognize material, objects, and scenes from very few examples and without much effort. In this direction the authors of [31] make the hypothesis that, once a few categories have been learned with significant cost, some information may be abstracted from the process to make learning further categories more efficient. Based on this hypothesis, when learning new categories, they take advantage of the "general knowledge" extracted from previously learned categories by using it in the form of a prior probability density function in the space of model parameters. Similarly in [32] when images of new concepts are added to the visual analysis model, the computer only needs to learn from the new images.

What has been learned about previous concepts is stored in the form of profiling models, and the computer needs no re-training.

On the other hand in [55] the authors claim that with the availability of overwhelming amounts of data, many problems can be solved without resorting to sophisticated algorithms. The authors mention the example of Google's "Did you mean" tool, which corrects errors in search queries by memorizing billions of query-answer pairs and suggesting the one closest to the user query. In their paper the authors present a visual analog to this tool using a large dataset of 79 million images and a non-parametric approach for image annotation that is based on nearest neighbor matching. Additionally, the authors of [8] employ multiple instance learning to learn models from images labeled as containing the semantic concept of interest, but without indication of which image regions are observations of that concept. Similarly in [17] object recognition is viewed as machine translation that uses expectation maximization in order to learn how to map visual objects (blobs) to concept labels. In all cases, the authors are trying to discover a scalable (in terms of the number of concepts) and effortless (in terms of the necessary annotation) way to teach the machine how to recognize visual objects the way a human does. Motivated by the same objective, in this work we investigate whether the knowledge aggregated in social tagging systems by the collaboration of web users can help in this direction.

While model parameters can be estimated more efficiently from strongly annotated samples, such samples are very expensive to obtain. On the contrary, weakly annotated samples can be found in large quantities especially from social media sources. Social Tagging systems such as Flickr accommodate image corpora populated with hundreds of user tagged images on a daily basis. Motivated by this fact, our work aims at combining the advantages of both strongly supervised (learn model parameters more efficiently) and weakly supervised (learn from samples obtained at low cost) methods, by allowing the strongly supervised methods to learn from training samples that are found in collaborative tagging environments. Specifically, drawing from a large pool of weakly annotated images, our goal is to benefit from the knowledge aggregated in social tagging systems, in order to automatically determine a set of image regions that can be associated with a certain tag.

5.1 Framework Description

The proposed framework for leveraging social media to train object detection models is depicted in Figure. 7. The analysis components of the framework are: tag-based image selection, image segmentation, extraction of visual features from image regions, region-based clustering using their visual features and supervised learning of object detection models using strongly annotated samples.

More specifically, given an object c that we wish to train a detector for, our method starts from a large collection of user tagged images and performs the following actions. Images are selected based on their tag information in order to formulate image group(s) that correspond to thematic entities. Given the tendency of social tagging systems to formulate knowledge patterns that reflect the way content is

perceived by the web users [34], tag-based image selection is expected to identify these patterns and create image group(s) emphasizing on a certain object. By emphasizing we refer to the case where the majority of the images within a group depict a certain object and that the linguistic description of that object can be obtained from the most frequently appearing tag (see Section 5.2 for more details). Subsequently, region-based clustering is performed on all images belonging to the image group that emphasizes on object c, that have been pre-segmented by an automatic segmentation algorithm. During region-based clustering the image regions are represented by their visual features and each of the generated clusters contains visually similar regions. Since the majority of the images within the selected group depicts instances of the desired object c, we anticipate that the majority of regions representing the object of interest will be gathered in the most populated cluster, pushing all irrelevant regions to the other clusters. Eventually, we use as positive samples the visual features extracted from the regions belonging to the most populated cluster, to train in a supervised manner a model detecting the object c.

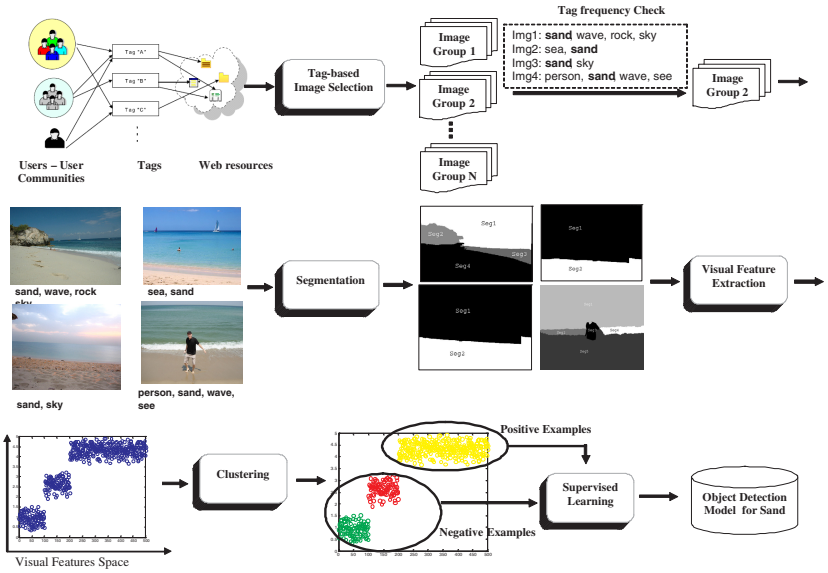


Fig. 7 Leveraging a set of user tagged images to train a model for detecting the object *sand*.

5.2 *Analysis Components*

Tag-based image selection: Refers to the techniques used to select images from a large dataset (S) of arbitrary content, based on their tag information. We employ one of the following three approaches based on the associated annotations:

1. **Keyword-based selection:** This approach is used for selecting images from strongly annotated datasets. In order to create $S^c \subset S$ we need only to select the images that are labeled with the name of the object c.

2. **Flickr groups:** Flickr groups are virtual places hosted in collaborative tagging environments that allow social users to share content on a certain topic. In this case, S^c is created by taking all images contained in a Flickr group titled with the name of the object c. From here on we will refer to those images as roughly-annotated images.
3. **SEMSOC:** SEMSOC [22, 23] is applied by our framework on weakly annotated images (i.e., images that have been tagged by humans in the context of a collaborative tagging environment, but no rigid annotations have been provided) in order to create semantically consistent groups of images. In order to obtain the image group S^c that emphasizes on object c, we select the SEMSOC-generated group S^{ci} where its most frequent tag relates with c.

Segmentation: Segmentation is applied on all images in S^c with the aim to extract the spatial masks of visually meaningful regions. In our work we have used a K-means with connectivity constraint algorithm as described in [35]. The output of this algorithm for an image I_q is a set of segments $R_{I_q} = \{r_i^{I_q}, i = 1,\ldots,m\}$, which roughly correspond to meaningful objects.

Visual Descriptors: In order to describe visually the segmented regions we have employed the following: a) the Harris-Laplace detector and a dense sampling approach for determining the interest points, b) the SIFT descriptor as proposed by Lowe [33] in order to describe each interest point using a 128-dimensional feature vector, and c) the bag-of-words model initially proposed in [50] in order to obtain a fixed-length feature vector for each region. The feature extraction process is similar to the one described in [44] with the important difference that in our case descriptors are extracted to represent each of the identified image segments, rather than the whole image. Thus, $\forall r_i^{I_q} \in R_{I_q}$ and $\forall I_q \in S^c$ a 300-dimensional feature vector $f(r_i^{I_q})$ is extracted.

Clustering: For performing feature-based region clustering we applied the affinity propagation clustering algorithm [19] on all extracted feature vectors $f(r_i^{I_q})$, $\forall r_i^{I_q} \in R_{Iq}$ and $\forall I_q \in S^c$. This is an algorithm that takes as input the measures of similarity between pairs of data points and exchanges messages between data points, until a high-quality set of centers and corresponding clusters is found.

Learning Model Parameters: Support Vector Machines (SVMs) [46] were chosen for generating the object detection models due to their ability in smoothly generalizing and coping efficiently with high-dimensionality pattern recognition problems. All feature vectors assigned to the most populated of the created clusters were used as positive examples for training a binary classifier. Negative examples were chosen arbitrarily from the remaining dataset. For training the object detection models we have used the libSVM library [10]. The radial basis function(RBF) kernel was used to map the samples into a higher dimensional space. In order to find the optimal parameters for the RBF kernel (C and γ) we performed 10-fold cross validation (i.e., divide the training set into 10 subsets of equal size and evaluate the performance using each time one of the subsets for testing and the remaining 9 for training). A "grid-search" on the exhaustive range of C and γ parameters provides us with

various pairs of (C,γ) values. These pairs are evaluated using cross-validation and the one with the best cross-validation accuracy is selected.

5.3 Evaluation of Object Detection Models

The goal of our experimental study is to compare the quality of object models trained using samples leveraged by the proposed framework, against the models trained using manually provided, strongly annotated samples. To carry out our experiments we have relied on three different types of datasets. The first type includes the strongly annotated datasets constructed by asking people to provide region detail annotations of images pre-segmented with the automatic segmentation algorithm of Section 5.2. For this case we have used a collection of 536 images S^B from the *Seaside* domain annotated in our lab. The second type refers to the roughly-annotated datasets like the ones formed in Flickr groups. In order to create a dataset of this type S^G, for each object of interest, we have downloaded 500 member images from a Flickr group that is titled with a name related to the name of the object. The third type refers to the weakly annotated datasets like the ones that can be collected freely from the collaborative tagging environments. For this case, we have crawled 3000 images S^{F3K} from Flickr using the wget[3] utility and Flickr API facilities. Moreover, in order to investigate the impact of the dataset size on the robustness of the generated models we have also crawled from Flickr another dataset consisting 10000 images S^{F10K}. Depending on the annotation type we use the tag-based selection approaches presented in Section 5.2 to construct the necessary image groups S^c.

In order to compare the efficiency of the models generated using training samples with different annotation type (i.e., strongly, roughly, weakly), we need a set of objects that are common in all three types of datasets. For this reason after examining the contents of S^B, reviewing the availability of groups in Flickr and applying SEMSOC on S^{F3K} and S^{F10K}, we determined four object categories C^{bench}={**sky, sea, vegetation, person**}. These objects exhibited significant presence in all different datasets and served as benchmarks for comparing the quality of the different models. The factor limiting the number of benchmarking objects is on the one hand the need to have strongly annotated images for these objects and on the other hand the un-supervised nature of SEMSOC that restricts the eligible objects to the ones emphasized by the generated image groups. C^{bench} is the maximum set of objects shared between all different dataset types. For each object $c_i \in C^{bench}$ one model was trained using the strong annotations of S^B, one model was trained using the roughly-annotated images contained in S^G, and two models were trained using the weak annotations of S^{F3K} and S^{F10K}, respectively. In order to evaluate the performance of these models, we test them using a subset (i.e., 268 images) of the strongly annotated dataset $S^B_{test} \subset S^B$, not used during training. F-Measure was used for measuring the efficiency of the models.

By looking at the bar diagram of Figure 8, we derive the following conclusions: a) Model parameters are estimated more efficiently when trained with strongly

[3] wget: http://www.gnu.org/software/wget

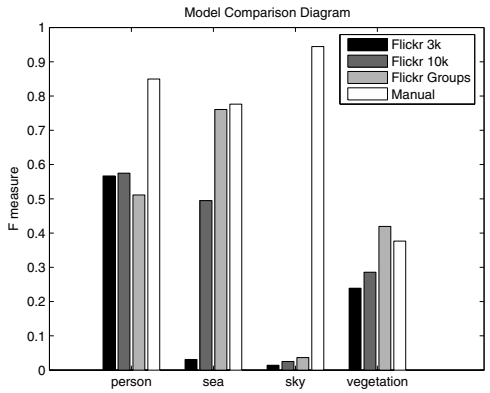


Fig. 8 Performance comparison between four object recognition models that are learned using samples of different annotation quality (i.e., strongly, roughly and weakly)

annotated samples, since in three out of four cases they outperform the other models and sometimes by a significant amount (e.g., sky, person). b) Flickr groups can serve as a less costly alternative for learning the model parameters, since using the roughly-annotated samples we get comparable and sometimes even better (e.g., vegetation) performance than manually trained models, while requiring considerable less effort to collect the training samples. c) The models learned from weakly annotated samples are usually inferior to the other cases, especially in cases where the proposed approach for leveraging the data has failed in selecting the appropriate cluster (e.g., *sea* and *sky*).

One drawback of Flickr groups derives from the fact that since they are essentially virtual places they are not guaranteed to constantly increase their size and therefore provide larger datasets that could potentially increase the efficiency of the developed models. This is why we also employ SEMSOC for constructing the necessary images sets. SEMSOC is an unsupervised selection procedure that operates directly on image tags and its goal is to provide a set of images the majority of which depict the object of interest. Naturally the image sets generated by SEMSOC are not of the same quality as those obtained from Flickr groups. However, the motivation for using SEMSOC is that it can potentially produce considerably larger image sets. Given that in Flickr groups the user needs to classify an image in one of the existing groups (or create a new group), the total number of positive samples that can be extracted from the images of a Flickr group, has an upper limit on the total number of images that have been included in this group by the users. On the other hand, the total number of positive samples that can be obtained by SEMSOC in principle, is only limited by the total number of images that are uploaded on the entire Flickr repository and depict the object of interest. However, given that collaborative tagging environments like Flickr are growing rapidly, we can accept that SEMSOC will manage to produce arbitrary large image sets. In this respect, in our experiment

Table 3 Comparing with existing methods in object recognition

	Building	Grass	Tree	Cow	Sheep	Sky	Aeroplane	Water	Face	Car	Bicycle	Flower	Sign	Bird	Book	Chair	Road	Cat	Dog	Body	Boat	Average
Textonboost [48]	62	**98**	**86**	58	50	83	60	**53**	**74**	63	**75**	63	**35**	19	**92**	15	**86**	54	19	**62**	7	**58**
PLSA-MRF/I [57]	45	64	71	**75**	**74**	**86**	**81**	47	1	**73**	55	**88**	6	6	63	18	80	27	26	55	8	50
Prop. Framework	**87**	9	65	45	45	14	29	**53**	56	12	**75**	**88**	27	**30**	25	**50**	44	**59**	**71**	29	**41**	45

we also examine how the efficiency of the developed models is affected by the size of the image set that has been used to obtain their training samples.

From the bar diagram of Figure 8 it is clear that when using the S^{F10K} the incorporation of more indicative examples into the training set improves the generalization ability of the generated models in all four cases. However, in the case of object *sea* we note also a drastic improvement of the model's efficiency. This is attributed to the fact that the increment of the dataset size alleviates the error introduced by the employed algorithms (i.e., segmentation, feature extraction, clustering) and allows the proposed method to select the appropriate cluster for training the model. On the other hand, in the case of object *sky* it seems that the correct cluster is still missed despite the use of a larger dataset. In this case the size of the dataset should grow even larger in order to compensate for the aforementioned error and select the appropriate cluster.

In order to compare our framework with existing methods we used the publicly available MSRC dataset[4] consisting of 591 images. In order to train the models, for each of the 21 objects, we have downloaded 500 member images from a Flickr group that is titled with a name related to the name of the object. We compare the region label annotations that were automatically acquired by our framework using Flickr groups with the patch level annotations of the approach proposed Verbeek and Triggs [57] and the ones obtained from Textonboost [48]. The the classification rates per object for each method are shown in Table 3. Looking at the individual objects we can see that despite the low cost for annotation our method yields the best performance in 9 out of 21 cases, compared to 7 out of 21 for the PLSA-MRF/I and 8 out of 21 for the Textonboost (note that in three cases Water, Flower, Bicycle the classification rates are identical for two different methods). On average, the accuracy obtained from our approach (45%) is inferior to the one obtained from PLSA-MRF/I (50%) which is again inferior to the accuracy obtained from Textonboost (58%). This is in accordance with our expectation since the performance scores obtained by the three methods are ranked proportionally to the amount of annotation effort required to train their models. Based on the above we can claim that the significant gain in effort that we achieve by leveraging social media to obtain the necessary training samples, compensates for the limited loss in performance that we suffer when compared with state of the art object recognition systems.

In this Section we have shown that the collective knowledge encoded in the user contributed content can be successfully used to remove the need for close human supervision when training object detectors. The experimental results have

[4] http://research.microsoft.com/vision/cambridge/recognition

demonstrated that although the performance of the detectors trained using leveraged social media is inferior to the one achieved by manually trained detectors, there are cases where the gain in effort compensates for the small loss in performance. In addition we have seen that by increasing the number of utilized images we manage to improve the performance of the generated detectors, advocating the potential of social media to facilitate the creation of reliable and effective object detectors. Finally, despite the fact that there will always be strong dependence between the discriminative power of the employed feature space and the efficiency of the proposed approach in selecting the appropriate set of training samples, our experimental study has shown that we can maximize the probability of success by using large volumes of user contributed content.

6 Conclusions

In this chapter we have demonstrated how massive user contributions can be leveraged to extract valuable knowledge. The community structure of tag networks, the emerging trends and events in users tag activity, as well as the associations between image regions and tags in user tagged images, all form different types of knowledge that was made possible to extract due to the collaborative and massive nature of the data. It is true that with the abundant availability of social data on the Web, analysis can now use the information coming both from the content itself, the social context and the emergent social dynamics. Although noisy and of questionable reliability, user contributions exhibit noise reduction properties when considered massively, given that they encode the collective knowledge of multiple users. Thus, the common objective among all methods performing knowledge extraction on social media, is to exploit those noise reduction properties and capture the knowledge provided by multiple users.

Our review on the methods performing knowledge extraction from massive user contributions has resulted in the following observations. Unsupervised approaches constitute the main vehicle for extracting the statistical patterns of the data. Either through the use of clustering, co-clustering or community detection techniques, we have noticed the tendency of keeping human intervention to a minimum and favoring algorithms that are able to extract all necessary parameters (e.g., the number of clusters) by pre-processing the available data. This tendency is basically motivated by the need to process a huge amount of data, which renders impractical schemes that require supervision. This tendency is further explained by the fact that the effectiveness of the methods extracting knowledge from social media is tightly bound to the amount of data that need to be processed. Given that the knowledge-rich patterns encoded in the data become stable and thus "visible" only after a specific usage period, many are the cases where the proposed approaches are unable to produce meaningful results, unless applied on large scale datasets. This is the reason why scalability constitutes an important requirement for such methods.

As avenues for future research we can identify the tendency of scientific efforts to optimally combine the information carried by the different modalities hosted by

social networks (i.e., images, tags, friendship links, etc). Being different in nature and heterogeneous in representation, this information should be analyzed by appropriately designed methods in order to become exploitable under a certain task. Finally, as a particularly challenging objective we also identify the potential of employing all those knowledge extraction methods, for automating the process of making the content contributed by users part of the Linked Open Data (LOD) cloud.

Acknowledgements. This work was sponsored by the European Commission as part of the Information Society Technologies (IST) programme under grant agreement n215453 - WeKnowIt and the contract FP7-248984 GLOCAL.

References

1. Ching-man, Gibbins, N., Yeung, N.S.A.: A study of user profile generation from folksonomies. In: SWKM (2008)
2. Au Yeung, C.m., Gibbins, N., Shadbolt, N.: Contextualising tags in collaborative tagging systems. In: HT 2009: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, pp. 251–260. ACM, New York (2009)
3. Aurnhammer, M., Hanappe, P., Steels, L.: Augmenting navigation for collaborative tagging with emergent semantics. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 58–71. Springer, Heidelberg (2006)
4. Becker, H., Naaman, M., Gravano, L.: Learning similarity metrics for event identification in social media. In: WSDM 2010, pp. 291–300. ACM, New York (2010)
5. Begelman, G.: Automated tag clustering: Improving search and exploration in the tag space. In: Proc. of the Collaborative Web Tagging Workshop at WWW 2006 (2006)
6. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30, 107–117 (1998)
7. Brooks, C.H., Montanez, N.: Improved annotation of the blogosphere via autotagging and hierarchical clustering. In: WWW 2006, pp. 625–632. ACM, New York (2006)
8. Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.: Supervised learning of semantic classes for image annotation and retrieval. IEEE Trans. Pattern Anal. Mach. Intell. 29(3), 394–410 (2007)
9. Cattuto, C.: Collaborative tagging as a complex system. talk given at internationl school on semiotic dynamics. In: Language and Complexity, Erice (2005)
10. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), Software, available at `http://www.csie.ntu.edu.tw/cjlin/libsvm`
11. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E 70(6), 066,111 (2004)
12. Cooper, M., Foote, J., Girgensohn, A., Wilcox, L.: Temporal event clustering for digital photo collections. ACM Trans. Multimedia Comput. Commun. Appl. 1(3), 269–288 (2005)
13. Cour, T., Sapp, B., Jordan, C., Taskar, B.: Learning from ambiguously labeled images. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2009 (2009)
14. Dhillon, I.S.: Co-clustering documents and words using bipartite spectral graph partitioning. In: Proceedings of KDD 2001, San Francisco, California, pp. 269–274 (2001)

15. Diederich, J., Iofciu, T.: Finding communities of practice from user profiles based on folksonomies. In: Proceedings of the 1st International Workshop on Building Technology Enhanced Learning Solutions for Communities of Practice, TEL-CoPs 2006 (2006)
16. Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., Tomkins, A.: Visualizing tags over time. In: Proceedings of WWW 2006, pp. 193–202. ACM, Edinburgh (2006)
17. Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002)
18. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database (Language, Speech, and Communication). The MIT Press, Cambridge (1998)
19. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science 315, 972–976 (2007), http://www.psi.toronto.edu/affinitypropagation
20. Gemmell, J., Shepitsen, A., Mobasher, B., Burke, R.: Personalizing navigation in folksonomies using hierarchical tag clustering. In: Song, I.-Y., Eder, J., Nguyen, T.M. (eds.) DaWaK 2008. LNCS, vol. 5182, pp. 196–205. Springer, Heidelberg (2008)
21. Ghosh, H., Poornachander, P., Mallik, A., Chaudhury, S.: Learning ontology for personalized video retrieval. In: MS 2007: Workshop on Multimedia Information Retrieval on The Many Faces of Multimedia Semantics, pp. 39–46. ACM, New York (2007)
22. Giannakidou, E., Kompatsiaris, I., Vakali, A.: Semsoc: Semantic, social and content-based clustering in multimedia collaborative tagging systems. In: ICSC, pp. 128–135 (2008)
23. Giannakidou, E., Koutsonikola, V.A., Vakali, A., Kompatsiaris, Y.: Co-clustering tags and social data sources. In: WAIM, pp. 317–324 (2008)
24. Giannakidou, E., Koutsonikola, V.A., Vakali, A., Kompatsiaris, Y.: Exploring temporal aspects in user-tag co-clustering. In: Special session: Interactive Multimedia in Social Networks, WIAMIS (2010)
25. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: Proceedings of WWW 2007, pp. 211–220. ACM, New York (2007)
26. Hotho, A., Ja"schke, R., Schmitz, C., Stumme, G.: Trend detection in folksonomies. In: Avrithis, Y., Kompatsiaris, Y., Staab, S., O'Connor, N.E. (eds.) SAMT 2006. LNCS, vol. 4306, pp. 56–70. Springer, Heidelberg (2006)
27. Kennedy, L.S., Chang, S.F., Kozintsev, I.: To search or to label?: predicting the performance of search-based automatic image classifiers. In: Multimedia Information Retrieval, pp. 249–258 (2006)
28. Kennedy, L.S., Naaman, M., Ahern, S., Nair, R., Rattenbury, T.: How flickr helps us make sense of the world: context and content in community-contributed media collections. ACM Multimedia, 631–640 (2007)
29. Koutsonikola, V.A., Petridou, S., Vakali, A., Hacid, H., Benatallah, B.: Correlating time-related data sources with co-clustering. In: Bailey, J., Maier, D., Schewe, K.-D., Thalheim, B., Wang, X.S. (eds.) WISE 2008. LNCS, vol. 5175, pp. 264–279. Springer, Heidelberg (2008)
30. Koutsonikola, V., Vakali, A., Giannakidou, E., Kompatsiaris, I.: Clustering of social tagging system users: A topic and time based approach. In: Vossen, G., Long, D.D.E., Yu, J.X. (eds.) WISE 2009. LNCS, vol. 5802, pp. 75–86. Springer, Heidelberg (2009)
31. Li, F.F., Fergus, R., Perona, P.: One-shot learning of object categories. IEEE Trans. Pattern Anal. Mach. Intell. 28(4), 594–611 (2006)
32. Li, J., Wang, J.Z.: Real-time computerized annotation of pictures. IEEE Trans. Pattern Anal. Mach. Intell. 30(6), 985–1002 (2008), http://dx.doi.org/10.1109/TPAMI.2007.70847

33. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
34. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Ht06, tagging paper, taxonomy, flickr, academic article, to read. In: Hypertext, pp. 31–40 (2006)
35. Mezaris, V., Kompatsiaris, I., Strintzis, M.G.: Still image segmentation tools for object-based multimedia applications. IJPRAI 18(4), 701–725 (2004)
36. Mika, P.: Ontologies are us: A unified model of social networks and semantics. Web Semant 5(1), 5–15 (2007), http://dx.doi.org/10.1016/j.websem.2006.11.002
37. Nanopoulos, A., Gabriel, H.H., Spiliopoulou, M.: Spectral clustering in social-tagging systems. In: Vossen, G., Long, D.D.E., Yu, J.X. (eds.) WISE 2009. LNCS, vol. 5802, pp. 87–100. Springer, Heidelberg (2009)
38. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69(2), 026,113 (2004)
39. Papadopoulos, S., Kompatsiaris, Y., Vakali, A.: A graph-based clustering scheme for identifying related tags in folksonomies. In: Bach Pedersen, T., Mohania, M.K., Tjoa, A.M. (eds.) DAWAK 2010. LNCS, vol. 6263, pp. 65–76. Springer, Heidelberg (2010)
40. Papadopoulos, S., Vakali, A., Kompatsiaris, Y.: Community detection in collaborative tagging systems. In: Pardede, E. (ed.) Community-Built Database: Research and Development. Springer, Heidelberg (2010)
41. Quack, T., Leibe, B., Gool, L.J.V.: World-scale mining of objects and events from community photo collections. In: CIVR, pp. 47–56 (2008)
42. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: SIGIR 2007: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 103–110. ACM, New York (2007)
43. Russell, T.: Cloudalicious: Folksonomy over time. In: Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 364–364. ACM, Chapel Hill, NC, USA (2006)
44. van de Sande, K., Gevers, T., Snoek, C.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 99(1) (5555)
45. Schifanella, R., Barrat, A., Cattuto, C., Markines, B., Menczer, F.: Folks in folksonomies: social link prediction from shared metadata. In: WSDM 2010: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 271–280. ACM, New York (2010)
46. Scholkopf, B., Smola, A., Williamson, R., Bartlett, P.: New support vector algorithms. Neural Networks 22, 1083–1121 (2000)
47. Segaran, T.: Programming Collective Intelligence. O'Reilly Media Inc., Sebastopol (2007)
48. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: *textonBoost*: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006)
49. Simpson, E.: Clustering tags in enterprise and web folksonomies. HP Labs Techincal Reports (2008), http://www.hpl.hp.com/techreports/2008/HPL-2008-18.html
50. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: ICCV 2003: Proceedings of the Ninth IEEE International Conference on Computer Vision, p. 1470. IEEE Computer Society, Washington, DC, USA (2003)

51. Specia, L., Motta, E.: Integrating folksonomies with the semantic web. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 624–639. Springer, Heidelberg (2007)
52. Sun, A., Zeng, D., Li, H., Zheng, X.: Discovering trends in collaborative tagging systems. In: Yang, C.C., Chen, H., Chau, M., Chang, K., Lang, S.-D., Chen, P.S., Hsieh, R., Zeng, D., Wang, F.-Y., Carley, K.M., Mao, W., Zhan, J. (eds.) ISI Workshops 2008. LNCS, vol. 5075, pp. 377–383. Springer, Heidelberg (2008)
53. Sun, Y., Shimada, S., Taniguchi, Y., Kojima, A.: A novel region-based approach to visual concept modeling using web images. In: ACM Multimedia, pp. 635–638 (2008)
54. Swan, R., Allan, J.: Extracting significant time varying features from text. In: Proceedings of the Eighth International Conference on Information and Knowledge Management, pp. 38–45 (1999)
55. Torralba, A., Fergus, R., Freeman, W.T.: 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 1958–1970 (2008), http://doi.ieeecomputersociety.org/10.1109/TPAMI.2008.128
56. Tsikrika, T., Diou, C., de Vries, A.P., Delopoulos, A.: Image annotation using clickthrough data. In: 8th ACM International Conference on Image and Video Retrieval, Santorini, Greece (2009)
57. Verbeek, J.J., Triggs, B.: Region classification with markov field aspect models. In: CVPR (2007)
58. Wu, L., Hua, X.S., Yu, N., Ma, W.Y., Li, S.: Flickr distance. ACM Multimedia, 31–40 (2008)
59. Wu, Z., Palmer, M.: Verm semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, New Mexiko, USA, pp. 133–138 (1994)
60. Xu, X., Yuruk, N., Feng, Z., Schweiger, T.A.J.: Scan: a structural clustering algorithm for networks. In: KDD 2007: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 824–833. ACM, New York (2007)
61. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. Int. J. Comput. Vision 73(2), 213–238 (2007), http://dx.doi.org/10.1007/s11263-006-9794-4