

CelebrityNet: A Social Network Constructed from Large Scale Online Celebrity Images

Li-Jia Li, David A. Shamma, Xiangnan Kong, Sina Jafarpour, Roelof van Zwol, and Xuanhui Wang, Yahoo! Research

Photos are an important information carrier for implicit relationships. In this paper, we introduce an image based social network, called CelebrityNet, built from implicit relationships encoded in a collection of celebrity images. We analyze the social properties reflected in this image based social network and automatically infer communities among the celebrities. We demonstrate the interesting discoveries of the CelebrityNet. We particularly compare the inferred communities with human manually labeled ones and show quantitatively that the automatically detected communities are highly aligned with that of human interpretation. Inspired by the uniqueness of visual content and tag concepts within each community of the CelebrityNet, we further demonstrate that the constructed social network can serve as a knowledge base for high level visual recognition tasks. In particular, this social network is capable of significantly improving the performance of automatic image annotation and classification of unknown images.

Categories and Subject Descriptors: J.4 [SOCIAL AND BEHAVIORAL SCIENCES]: Sociology; E.1 [DATA]: Data Structures—*Graphs and Networks*; I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis—*Object Recognition*; I.2.10 [ARTIFICIAL INTELLIGENCE]: Vision and Scene Understanding—*Representations, data structures, and transforms*; I.5.4 [PATTERN RECOGNITION]: Applications—*Computer Vision*

General Terms: Human Factors

Additional Key Words and Phrases: Multimedia, Photos, Social Networks

ACM Reference Format:

Li, L.-J. and Shamma, D. A. and Kong, X. and Jafarpour, S. and Zwol, R. V. and Wang, X., 2014. CelebrityNet: A Social Network Constructed from Large Scale Online Celebrity Images *ACM TOMM* 9, 4, Article 39 (September 2014), 21 pages.

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

1. INTRODUCTION

People like photography. Since the invention of the 1901 Kodak Brownie, cameras have become more and more common place, capturing our personal life events, as well as, becoming a tool for conveying news. With the advent of digital photography, photo sharing for personal and professional applications underwent a speed up as photos no longer needed chemical development. In recent years, this was met with the growth of large-scale photo-sharing websites, research in image search, annotation and on-line organization. Undergoing rapid growth, it is projected that 10% of all photos ever

All authors were at Yahoo! Research when this work was performed.

Author's address: L.-J. Li, Yahoo!, 701 First Avenue, Sunnyvale, California, USA; email: li-jiali.vision@gmail.com; D. A. Shamma, Yahoo!, 475 Sansome Street, 9th Floor, San Francisco, CA, 94111, USA; email: aymans@acm.org; X. Kong, Worcester Polytechnic Institute; email: xkong@wpi.edu; S. Jafarpour, Turn Inc, email: sjafarpour@turn.com; R. van Zwol, Netflix; email: roelofvanzwol@mac.com; X. Wang, Facebook; email: xuanhui@gmail.com

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2014 ACM 0000-0000/2014/09-ART39 \$15.00

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

taken were taken in the past year.¹ A main feature of photo-sharing websites is explicit social connections and networks, where one can specify ‘friends’ to share and discuss photos. As the people and faces in these photos drive online engagement [Bakhshi et al. 2014], social computing research has taken focus as a research topic in computer science. However, the photos themselves are an important information carrier for implicit relationships—a relationship that has traditionally been neglected in social computing research as explicit ‘friend’ and ‘following’ relationships has taken favor.

Family members and friends like to travel together and take similar set of photos. Different generations of people have diverse preferences for objects, places and activities, which can be reflected in their photos. Celebrities participate in similar activities with a hoard of photographers capturing whatever they can for profit in the public media outlets. Recently, research combining the merits of both social network and media content has become an emerging research topic in multimedia and computer vision [Wang et al. 2010; Stone et al. 2010; Zhuang et al. 2011; Ding and Yilmaz 2010; Chen et al. 2012]. But very limited research has been done on constructing a social network of photo co-occurrence—inferring the implicit relationships of two or more people being in the same photo, captured at the same moment, at the same event, together. This implicit co-occurrence network is prevalent in large scale image collections; analysis of this phenomenon reflected is beneficial to many social network sites and other applications.

In this paper, we present CelebrityNet, an implicit social network constructed from the co-occurrence statistics of celebrities that appear in millions of professionally produced news images. CelebrityNet is used to automatically infer groups of celebrities sharing similar characteristics and present a quantitative analysis of the detected communities, based on comparison with manually defined communities. Our quantitative analysis demonstrates that the communities extracted automatically are highly aligned with the human notion of community. Further, there is a observable agreement of the images and tags, or semantic consistency, within each community. With the detected communities and their semantic consistency, we deploy the extracted communities to enhance the accuracy of automatic image annotation and classification. CelebrityNet encodes visual, semantic and social structural information, making it a valuable resource for developing structural learning algorithms which jointly models such information. We use a principle model to effectively incorporate such information for multimedia tasks. We explain the design rational of this model in details based on the observation from CelebrityNet. To coherently integrate the mutually beneficial information, we use a collective classification algorithm to learn this joint model inspired by the ICA algorithm [Lu and Getoor 2003]. Fundamentally different than pioneering social network algorithms [Stone et al. 2010; Wang et al. 2010; Zhuang et al. 2011; Chen et al. 2012], our approach serves as generic annotation or classification algorithms which are not limited to face recognition or friendship prediction. Experimental results confirm that CelebrityNet can be used as an important knowledge-base for applications such as classification and image annotation by providing informative tags for unseen people or images. Specifically, we make the following contributions in this paper:

- Automatically construct CelebrityNet, an implicit social network from large scale online images.
- Provide detailed analysis of the structure of the automatically constructed implicit social network and illustrate the influential leaders among the nodes.

¹<http://blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox> Accessed 10/2013.

- Detect communities within CelebrityNet and provide meaningful interpretation of the detected communities.
- Design rigorous quantitative experiment to evaluate the detected communities.
- Demonstrate effectiveness of CelebrityNet as an important complementary information source to traditional visual or textual data with significant performance improvement in multimedia tasks.

It is worth noting that our social network is developed and demonstrated on celebrity images. But it is not limited to celebrities, and can be extended to other information sources, such as people or objects appearing in photos of an online photo sharing website.

2. RELATED WORK

Over the last several years, with the emergence of large scale online structural data [Konstas et al. 2009; Brin and Page 1998], social network analysis has achieved substantial progress. While much of the research has been done on textual documents or hyper links, social network analysis on visual content has also made promising progress these years [Ding and Yilmaz 2010; Stone et al. 2010; Wang et al. 2010; Zhuang et al. 2011]. With the emergence of photo sharing websites such as Flickr and Facebook, pioneer research [Stone et al. 2008; Stone et al. 2010; Wang et al. 2010; Chen et al. 2012] has been conducted to tackle visual recognition problems by incorporating social network structure. For example, Wang et al. [2010] models the types of relationships based on face features such as face size ratio, age difference and gender distribution. Stone et al. [2010] leverages the social network structure to improve face recognition in a collection of face photos by using a MRF model. Both models focus on face recognition and analysis, which are not directly applicable to generic recognition tasks. Chen and her collaborators [Chen et al. 2012] propose an interesting approach for learning to classify family photos and predict different types of family relationships. Our approach tackles the challenging problems of generating generic annotation to photos while discovering more diverse social communities such as politician and athletes. Similar to [Kim et al. 2010], we construct our implicit social network by using the co-occurrence of faces in photos. Instead of manually labeling the appearance of people in a small set of photos (564 photos in total used in [Kim et al. 2010]), we rely on the person names appearing in the news article, which can be easily extracted or obtained from the tags of the photos in the news article. Therefore, we are able to apply our approach on millions of online photos. In addition, we take a few further steps to analyze the properties of this social network and explore the advantage of using it for high level visual recognition tasks.

Little has been done to construct an implicit social network from visual data, uncover the structure embedded and apply it for generic multimedia tasks such as image annotation and classification. In this paper, we explore large scale online photos to infer a social network and propose a model to harness the mutual information embedded in this social network. Visual recognition algorithms [Dalal and Triggs 2005; Felzenszwalb et al. 2007; Lin et al. 2011; Deng et al. 2010; Cao et al. 2009; Weston et al. 2010] have shown effectiveness in recognizing objects and classifying images. Most of them use only the visual content of images. Interesting research such as the multi-label classification approaches, [Read et al. 2009; Liu et al. 2010; Zha et al. 2008; Qi et al. 2007; Luo et al. 2013] further explore the correspondence among the tags related to the images and videos. Sophisticated algorithms are developed to model the relationship between the visual content and the semantic meaning of the visual content in these approaches. Our approach, on the other hand, takes the visual content, related semantic information and social network structure into account. We aim to emphasize

the impact of social network structure in visual tasks as a new knowledge resource for multimedia tasks such as community classification and image annotation.

3. CONSTRUCTION OF CELEBRITYNET

To build the CelebrityNet social network from a large collection of online celebrity photos, we first collect a rich set of photos, each of which is labeled with person(s) appearing in it, then we construct an edge list based on the co-occurrence of people in these photos, and finally we investigate the social network's mechanics.

3.1. Large Scale Online Celebrity Photos

We have collected a dataset of 8 million Images from Getty, a professional photographer image sharing website². Each image is provided with editorially labeled meta data which includes a list of concepts, the location where it was taken, person names appearing in the image, and a short explanation of the event. This dataset contains photos from early 2001 to late 2010. CelebrityNet is constructed by inferring the relationship among celebrities based on co-occurrence statistics in the photos. For this purpose, we sample the photos with people co-appear in the images. In total, we used 318,702 co-occurring pairs in photos to build the social network. In other words, people are connected if they are photographed together. Intuitively, two people who appear more often together in each other's photos are usually closely related.

3.2. Model Mechanics

Celebrities who are closely related to each other tend to attend same events and therefore appear in the same photos taken at those events. The presence of large scale celebrity photos allows us to accurately infer the relationship between the celebrities, and the strength of those relationship. We start by explaining the construction of CelebrityNet from which the celebrity relations are inferred.

Our social network is defined as a social graph $G = (V, E)$ for which the set of nodes V and edges E are:

- Each node $v \in V$ corresponds to a celebrity that appears in at least one photo of the photo collection.
- Each edge $e \in E$ connects two celebrities v_i, v_j if and only if there exists at least one photo in the collection in which v_i and v_j co-occur.
- Corresponding to each edge $e \in E$, there is a scalar number w_e that counts the number of photos in the collection in which both v_i and v_j appear.

In other words, edges represent the photo co-occurrence relationships between celebrities and the weight of each edge reflects the strength of the relationship. It follows from the construction of the graph that CelebrityNet is an undirected weighted graph.

4. OBSERVATIONS FROM CELEBRITYNET

CelebrityNet consists of 48,301 celebrities, and 113,158 edges connecting the nodes related to each other together in the social network. It has 5,343 connected components, where each connected component is a subgraph in which any two vertices are connected to each other but disconnected from other nodes outside the component. On average, each celebrity is connected to approximately 5 other celebrities. Number of nodes in the largest component is 32,927, where 98,294 edges connect the nodes together. The longest distance between two nodes is 24 steps. It is often important

²<http://www.gettyimages.com/> Accessed 1/2014.

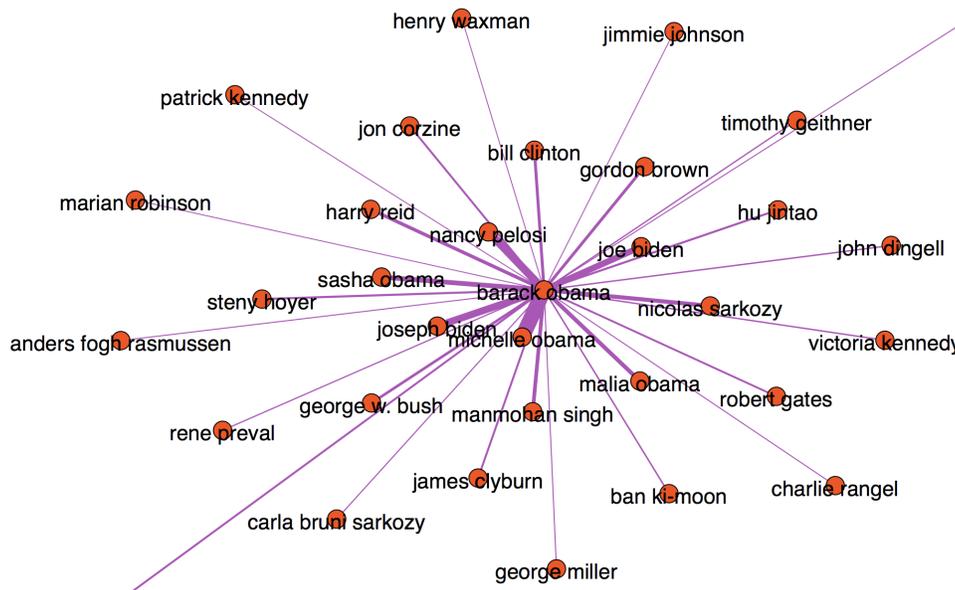


Fig. 1. Ego Network of Barack Obama. The ‘hub’, ‘Barack Obama’ is connected to all the other nodes. The width of each connection is proportional to the social strength between the two nodes.

to evaluate the people (nodes) with the largest numbers of direct connections, called hubs. Hubs usually have high impact in the social network. In CelebrityNet, ‘Barack Obama’ has the largest number of direct connections to 214 celebrities. Connection with the highest strength to ‘Barack Obama’ is ‘Michelle Obama’.

Now we turn to analyze the local behavior of CelebrityNet. To understand the local behavior, ego network plays a critical role. An ego network is a local network consists of the ‘focal’ node (ego) and all nodes to whom ego has a connection at some path length. We illustrate the ego network of ‘Barack Obama’ in Figure 1. Here, we only show the one-step neighbors of the ego node. The ego network includes the ego ‘Barack Obama’ and celebrities that are directly adjacent to him. For clarity of presentation, we only show people who co-occur more than 10 times. We can see that there are multiple communities which Barack Obama belongs to (e.g., family and politician) and we will discuss how to identify these communities in the next section.

Besides ‘Barack Obama’, there are many other important celebrities in CelebrityNet. In Figure 2, we illustrate the most ‘influential’ celebrities estimated by applying different centrality analysis criteria to CelebrityNet. These important celebrities are labeled with larger node size and assigned different colors than the rest in blue color. We demonstrated (see Figure 1) that many politician and their family members are directly related to President ‘Barack Obama’ as they often participate in same events together with the President. As expected, President ‘Barack Obama’ is the most prominent with the maximum standardized degree centrality:

$$C_D(v) = \frac{\text{degree}(v)}{g - 1}$$

where degree is the number of edges incident to the node v and g is the node group size, i.e., number of nodes in a graph [Wasserman and Faust 1994].

Closeness centrality [Beauchamp 1965] indicates how closely a node in a social network is connected to the others. In CelebrityNet, actor ‘Ashton Kutcher’ is assigned

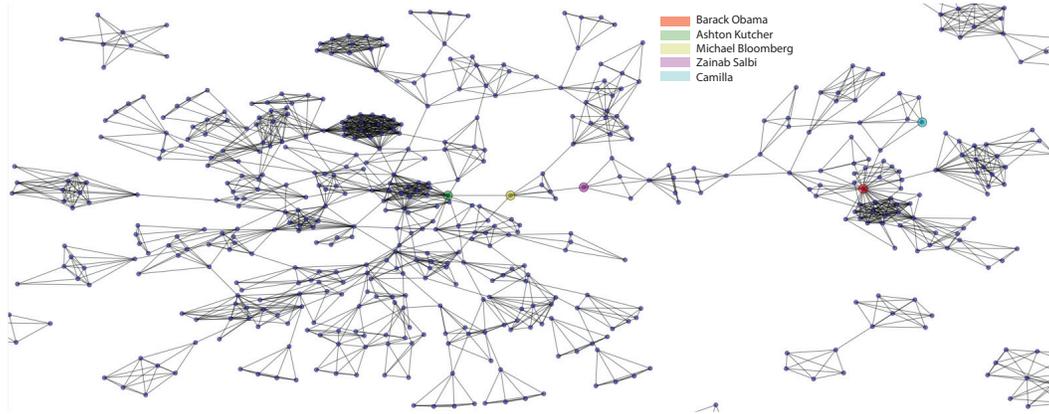


Fig. 2. The most ‘influential’ celebrities inferred by using different centrality analysis criteria. The celebrities with the highest degree centrality, closeness centrality and eigenvector centrality are: ‘Barack Obama’, ‘Ashton Kutcher’, and ‘Camilla’ respectively. The most prominent edge is between ‘Zainab Salbi’ and ‘Michael Bloomberg’ measured by using the between centrality criterion. For better clarity, we pruned out the nodes with few or weak connections.

with the largest standardized closeness centrality value as:

$$C_c(v) = \frac{g - 1}{\sum_{j=1}^g d(v, v_j)}$$

where $d(v, v_j)$ is the geodesic distance between node v and v_j . The geodesic distance is the number of lines of the optimal or most efficient connection between two nodes. From Figure 2, we can observe that ‘Ashton Kutcher’ is closely connected to a large number of celebrities in CelebrityNet. Most of them are connected to ‘Ashton Kutcher’ within a few links. The ego node ‘Ashton Kutcher’ is in fact closely connected to many actors, actresses and super models. It is not difficult to understand the importance of ‘Ashton Kutcher’ since CelebrityNet is constructed from online celebrity photos among which the majority are actors, actresses and super models. ‘Ashton Kutcher’, as a well known actor, producer and former fashion model, inevitably participates in numerous of events together with other Hollywood stars and has pictures taken in those events. Therefore, he naturally becomes the person who is closely connected to the largest number of celebrities in CelebrityNet.

Another critical measurement of the *importance* of a node in a social network is called eigenvector centrality [Gould 1967], defined as

$$C_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^n A_{ij} v_j$$

Here A_{ij} refers to the adjacent matrix. Unlike the previous two centrality measurements, when measuring the importance of a node, eigenvector centrality gives more credit to the node connected to the *influential* contacts. Here, ‘Camilla’ is estimated as the *influential leader* since she connects to many other influential leaders such as politician ‘Prince Charles’, ‘Stephen Harper’, and is only a few links away from ‘Barack Obama’ and ‘Ashton Kutcher’.

Finally, an interesting question we would ask is “What is the most important link, or edge, in the social network?.” The importance of a link can be measured by the edge betweenness centrality:

$$C_b(e) = \sum_{v_i, v_j, i \neq j} \frac{n_{v_i, v_j}(e)}{n_{v_i, v_j}}$$

as proposed by Freeman [1977]. Here e , $n_{v_i, v_j}(e)$, n_{v_i, v_j} represents the edge of interest, the total number of shortest paths between node v_i and node v_j that pass the edge e , and the total number of shortest paths between node v_i and node v_j respectively. Edge betweenness centrality reflects the traffic fraction going through the edge of interest. The edge between ‘Michael Bloomberg’ and ‘Zainab Salbi’ is estimated as the most important edge by using the edge betweenness centrality. This edge connects the two largest subnetworks, one mostly consists of politician and majority in the other are actors, actresses and super models in CelebrityNet.

5. COMMUNITIES IN CELEBRITYNET

With CelebrityNet’s social graph, we wish to discover community information within the social network, where each community is a group of celebrities that share similar properties. More formally, given a social graph, a community is defined as a group of nodes that are more closely connected to each other than to other nodes in the network [González et al. 2007; Gao and Wong 2006]. Specifically, we are interested in overlapping communities. Overlapping communities can often be observed in the real world social groups. For example, a person can belong to different groups related to his/her profession, family, and friends groups simultaneously.

To detect the overlapping communities in CelebrityNet, we adopt Palla et al.’s and Derényi et al.’s Clique Percolation Method (CPM) [2005; 2005]³. The goal is to detect groups of nodes that are more densely connected to each other than to the rest of the social network. Palla et al. proposes to populate the communities from k -cliques, where the term k -clique corresponds to fully connected sub-graphs of k nodes. The CPM is built upon the observation that the internal edges of a community are likely to form cliques whereas inter-community edges are unlikely to form cliques.

For our community detection, we impose an explicit requirement that each community must have at least 4 members to avoid tiny communities. To detect the communities, we place a k -clique template on any k -clique of the CelebrityNet graph, and roll it to an adjacent k -clique by relocating one of its nodes and keeping the other $k - 1$ nodes fixed. Thus, the k -clique communities of a network are all those sub-graphs that can be fully explored by rolling a k -clique template over all adjacent k -cliques. Here, two k -cliques are adjacent if they share $k - 1$ nodes.

Figure 3 demonstrates an example of some detected overlapping communities that all share the Barack Obama node. We observe that Barack Obama is assigned to multiple communities, including his family (Michelle Obama, Malia Obama, Sasha Obama and Marian Robinson) and the governors (Joe Biden, John Dingell and Kathleen Sebelius etc.). This observation aligns well with the real world role of the President.

Besides politicians, many other diverse communities were identified such as super models, athletes, and famous chefs. These professions are a natural form of community and serve as a divider amongst celebrity types. For example, Figure 4 shows various celebrity professions groups in the network. One immediate question is whether the social interaction would influence the visual appearance and content related to each

³Note that [Leskovec et al. 2010] provide a thorough study of network community detection. However, we focus on analyzing the phenomenon reflected by possible communities in CelebrityNet and developing effective method to leverage the community structure. Inventing the best possible community detection algorithm is beyond the scope of this paper.

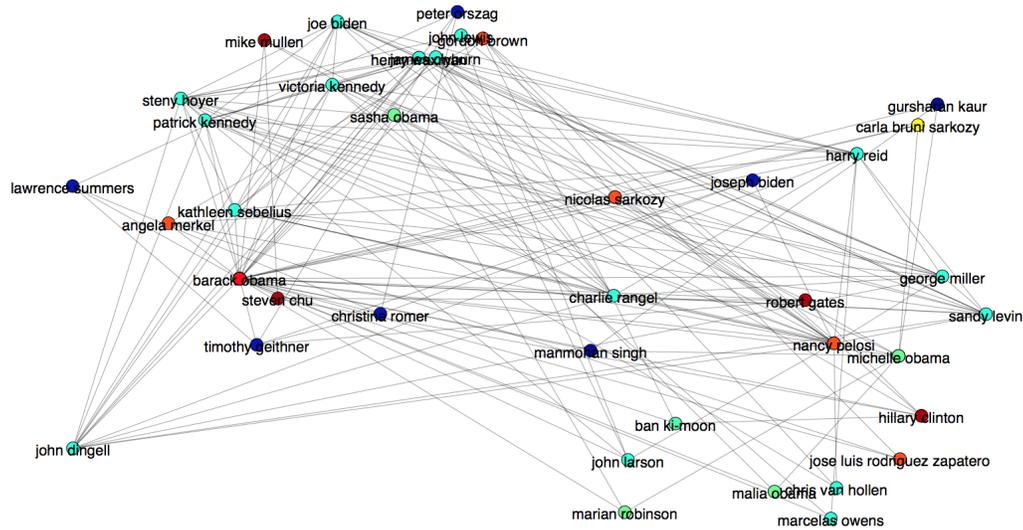


Fig. 3. Overlapping Community Example: Nodes belonging to different communities are labeled in different colors. Lines represent connection.

community. Would photos of people from the same society exhibit similar appearance and relate to similar descriptions?

Figure 5 illustrates examples of images of a few sample communities, as well as the tags assigned to them. As we can see from Figure 5, images belonging to a community often exhibit their unique property such as color, objects contained, scenery of the images, and the image tags. On the other hand, visual appearances and tags are much more diverse for images belonging to different communities. Communities indeed carry discriminative information about the visual and textual characteristics of the images that belong to them. In Section 7, we apply a principle model [Li et al. 2014] to incorporate this discriminative information for multimedia tasks such as image annotation and community classification.

6. COMMUNITY EVALUATION

In the previous section, we saw how the communities were automatically detected using the Clique Percolation Method. It is important to evaluate how the detected communities are aligned with the *human* definition of community. Toward this point, we evaluate how well the detected communities match the human perception. We recruited 12 editors to independently assign sets of celebrities into communities. By independently, we mean that the editors were unaware of the results of each other or the CPM.

We asked each editor to perform 27 distinct tests. Each test consisted of assigning a collection of different celebrities into communities, according to the following rules:

- Editors were presented with a set of images of people and their Wikipedia pages, and asked to group the people into communities based on common properties.
- Each celebrity depicted on the screen must be assigned to a community, but a person may belong to more than one community.
- A community should contain four or more people.

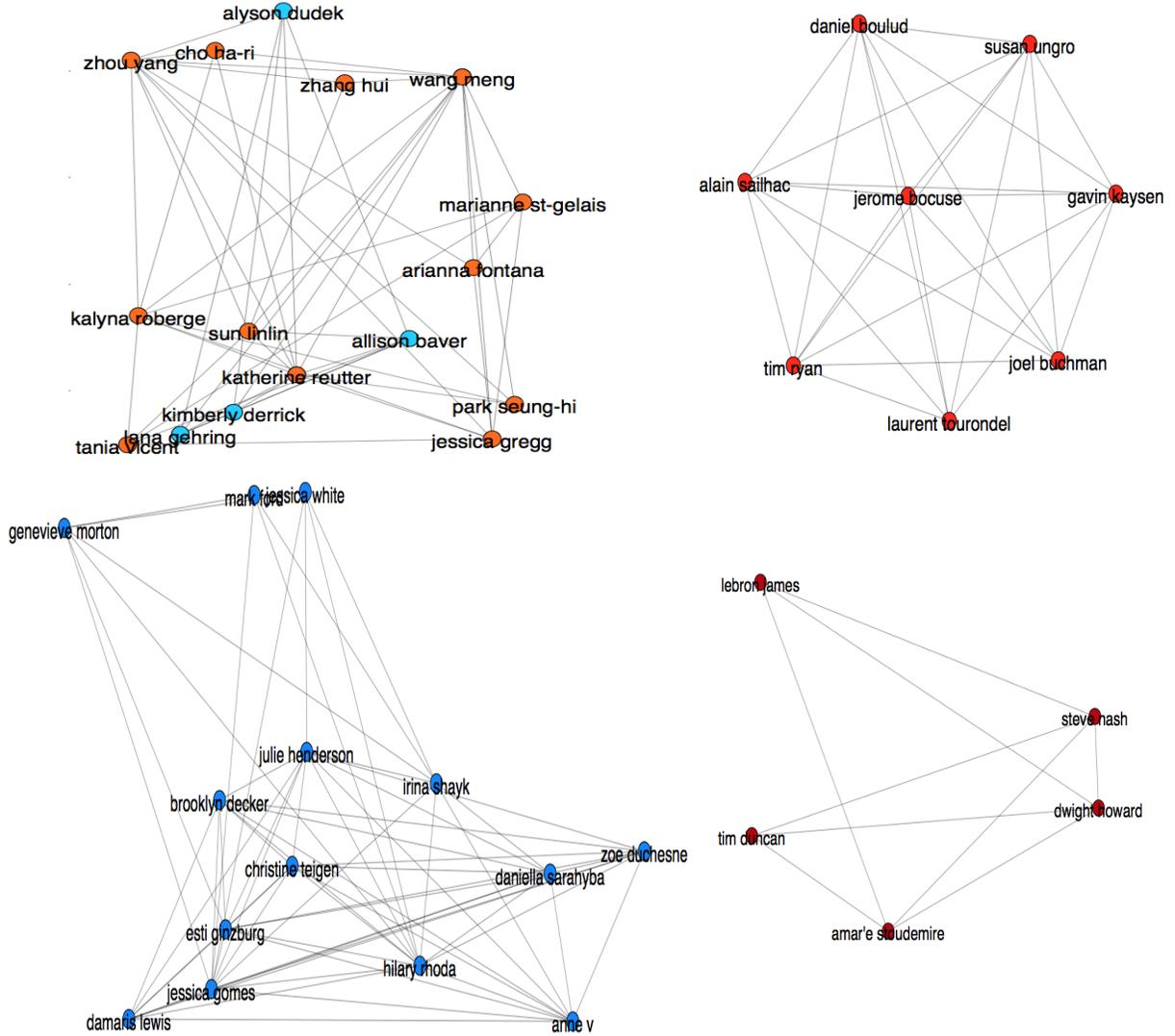


Fig. 4. Community Example: **Top Left:** A basketball player community. **Top Right:** A skater community. **Bottom Left:** A super model community. **Bottom Right:** A famous chef community. Nodes belonging to different communities are labeled in different colors in each subgraph.

We then measured the agreement of the communities provided by the editors and by the CPM. Specifically, we applied a modification of the Jaccard index to measure the agreement of communities generated from different sources. Let $X = \{x_1, \dots, x_n\}$ denote the set of n communities identified by an editor, and $Y = \{y_1, \dots, y_n\}$ denote the set of n communities identified by another editor or by our algorithm (remember that according to the rules, each x_i or y_i is either empty or has at least four members). The Jaccard [1901] index between the sets x_i and y_i is defined as [Tan et al. 2005]:

$$J(x_i, y_i) \doteq \frac{|x_i \cap y_i|}{|x_i \cup y_i|} \tag{1}$$

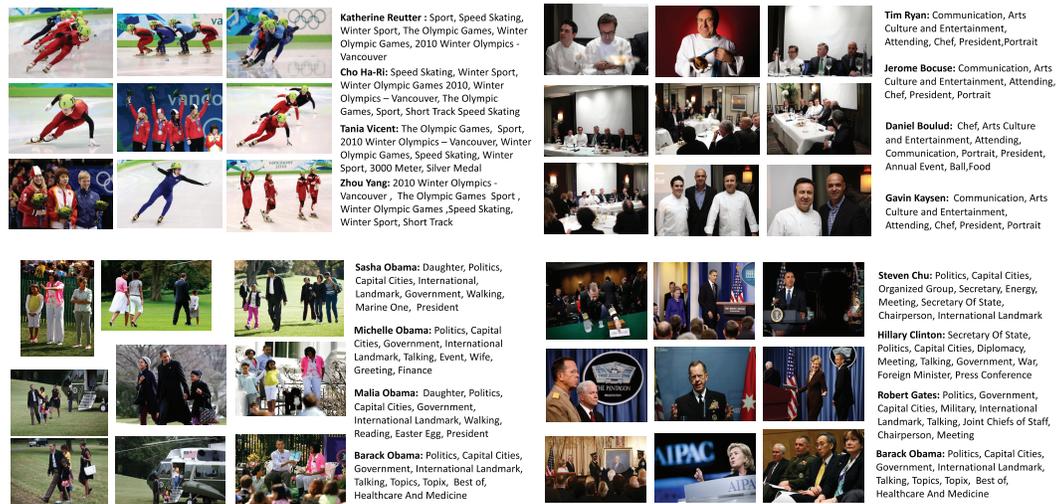


Fig. 5. Sampled images and their annotations of four communities. **Upper Left**: Sports Players (Speed Skating)(Photo credit from left to right, top to bottom: Streeter Lecka/Getty Images Sport, Streeter Lecka/Getty Images Sport, Matthew Stockman/Getty Images Sport, Streeter Lecka/Getty Images Sport, Kevork Djansezian/Getty Images Sport, Alex Livesey/Getty Images Sport, Matthew Stockman/Getty Images Sport, Matthew Stockman/Getty Images Sport, Matthew Stockman/Getty Images Sport) **Upper Right**: Chefs.(Photo credit from left to right, top to bottom: Neilson Barnard/Getty Images Entertainment, Scott Halleran/Getty Images Entertainment, Neilson Barnard/Getty Images Entertainment) **Bottom Left**: Obama Family.(Photo credit from left to right, top to bottom: Chip Somodevilla/Getty Images News, Mark Wilson/Getty Images News, Brendan Smialowski/Getty Images News, Mark Wilson/Getty Images News, Brendan Smialowski/Getty Images News, Mark Wilson/Getty Images News, Alex Wong/Getty Images News, Brendan Smialowski/Getty Images News, Chip Somodevilla/Getty Images News) **Bottom Right**: Politicians related to Obama.(Photo credit from left to right, top to bottom: Chip Somodevilla/Getty Images News, Chip Somodevilla/Getty Images News, Spencer Platt/Getty Images News, Chip Somodevilla/Getty Images News, Salah Malkawi/Getty Images News, Chip Somodevilla/Getty Images News, Win McNamee/Getty Images News, Brendan Smialowski/Getty Images News, Alex Wong/Getty Images News)

We compare the agreement between the community assignments X and Y by calculating the average Jaccard agreement between all community pair agreements x_i and y_i :

$$J(X, Y) \doteq \frac{1}{n} \sum_{i=1}^n J(x_i, y_i) \quad (2)$$

The Jaccard index $J(X, Y)$ is a measure of agreement between *ordered* community assignments X and Y . However, in our experiments there is no particular order in assigning people to communities. That is, one person can identify the Obama Family as the first community, and the World Leaders as the second community, whereas another editor may assign the World Leaders as the first community, and the Obama Family as the second community. To overcome this ordering difficulty, we measured the Jaccard agreement between X and all permutations of the elements of Y , and reported the largest such value. More precisely, let Π denote the set of all permutations of $\{1, \dots, n\}$. The Orderless Jaccard (OJ) index between the community assignments X and Y is

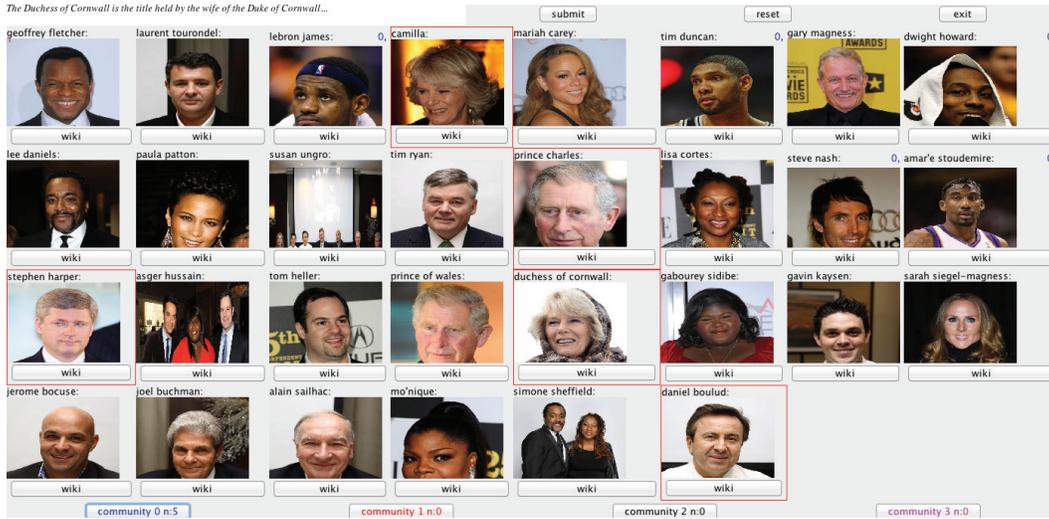


Fig. 6. A snapshot of the community assignment interface used by human editors to assign celebrities into different communities. We asked 12 editors to perform 27 tests independently using this interface.

then defined as

$$\text{OJ}(X, Y) \doteq \max_{\pi \in \Pi} \frac{1}{n} \sum_{i=1}^n J(x_i, y_{\pi_i}). \quad (3)$$

It follows from the definition of the Orderless Jaccard index that $\text{OJ}(X, Y)$ is always between 0 and 1. Moreover, larger $\text{OJ}(X, Y)$ indicates higher agreement between the community assignments X and Y .

Figure 6 shows a snapshot of one particular community detection test. Each test had 30 people and the editors were asked to assign them into at most 4 communities according to the above rules.

We compared the degree of agreement between each pair of editors and also between each editor and the CPM by calculating the all possible Orderless Jaccard indices for each of the 27 separate tests separately. We show in Figure 7 that the communities detected by the CPM aligns well with the human notion of communities. Figure 7(a) plots the mean and standard deviation of the Orderless Jaccard values between all Human-Human (circles) and Human-Algorithm (squares) pairs. As Figure 7(a) indicates, there are more difficult tests, where the degree of agreement between the editors is consistently low (e.g. test #7). On the other hand, in the simpler tests there are more agreements between different editors (e.g., test #6). In the more difficult tests there is also less agreement between the editors and the CPM, whereas in the easiest case the human and CPM assignments are consistently highly aligned. The agreement between our detected communities and the human subjects are close to the one among the human subjects. In contrast, Figure 7(b) plots the low similarity between the CPM community assignment and a random assignment of the celebrities into communities. In the random assignment, each person is independently assigned to each community with a weight proportional to the average number of people assigned to that community by human editors. Figure 7(c) shows the mean and standard deviation of all Jaccard similarities over all 27 tests. As it illustrates, the CPM average community assignment is much more similar to that by an editor, compared to a random community assignment.

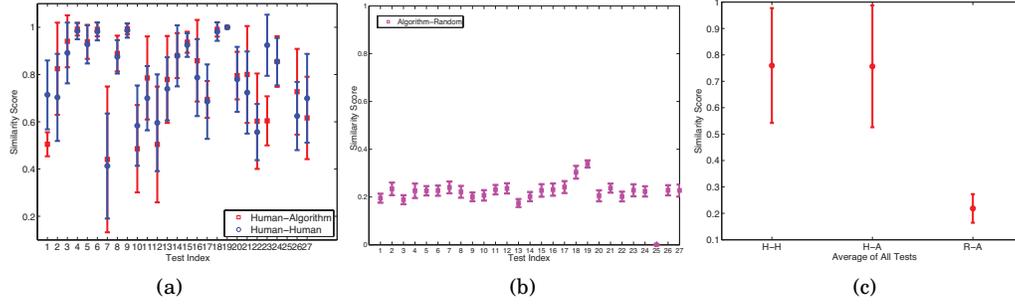


Fig. 7. The communities detected by algorithm are much more aligned with human community extraction compared to random community assignments. (a) Mean and standard deviation of all human-human and human-algorithm Orderless Jaccard similarities for each test. The tests with lower mean and higher standard deviation are the more difficult tests. (b) Orderless Jaccard similarity between the algorithm and a random assignment of the celebrities to communities for each test. The similarity scores are the average over 200 random trials. (c) Mean and standard deviation of all Jaccard similarities over all 27 tests.

7. A MULTI-MODAL MULTI-LABEL CLASSIFICATION MODEL

CelebrityNet can serve as a knowledge base for many potential applications such as image annotation or user recommendation. In this section, we evaluate the utility of CelebrityNet and the discovered communities on image annotation and community classification. As described in [Li et al. 2014], we formulate the image annotation and community classification tasks as multi-label classification problems. For each unknown test image, our goal is to provide a list of tags related to it. Similarly, each person in the social network will be assigned to one or multiple communities. We denote the multi-label dataset as $\mathcal{D}(\mathcal{X}, \mathcal{Y}) = \{(x_i, Y_i)\}_{i=1}^n$ and the network as $G(\mathcal{V}, \mathcal{E})$. Here $\mathcal{X} = \{x_i\}_{i=1}^n$, and $x_i \in \mathbb{R}^d$ is the feature vector of sample x_i in the d -dimensional input space. $\mathcal{Y} = \{Y_i\}_{i=1}^n$, where sample x_i refers to a sample image and $Y_i = (Y_i^1, \dots, Y_i^q)^\top \in \{0, 1\}^q$ denotes the multiple labels assigned to sample x_i . Let $\mathcal{C} = \{\ell_1, \dots, \ell_q\}$ be the set of q possible label concepts. In the network G , $\mathcal{V} = \{v_1, \dots, v_n\}$ is a set of nodes, which corresponds to the samples in \mathcal{D} . \mathcal{E} is the set of links/edges in $\mathcal{V} \times \mathcal{V}$. Assume that we have a training set $\mathcal{X}_{\mathcal{L}} \subset \mathcal{X}$ where the values $\mathcal{Y}_{\mathcal{L}}$ are known. Here \mathcal{L} denotes the index set for training data, $y_i = (y_i^1, \dots, y_i^q)^\top \in \{0, 1\}^q$ is a binary vector representing the observed label set assigned to sample x_i . $y_i^k = 1$ if the k -th label is in x_i 's label set.

Multi-label collective classification corresponds to the task of predicting the values of all $Y_i \in \mathcal{Y}_{\mathcal{U}}$ for the testing set collectively ($\mathcal{X}_{\mathcal{U}} = \mathcal{X} - \mathcal{X}_{\mathcal{L}}$), where the inference problem is to estimate $\Pr(\mathcal{Y}_{\mathcal{U}} | \mathcal{X}, \mathcal{Y}_{\mathcal{L}})$. Conventional supervised classification approaches usually has *i.i.d.* assumptions, i.e. the inference for each sample is independent from other samples, i.e. $\Pr(\mathcal{Y}_{\mathcal{U}} | \mathcal{X}, \mathcal{Y}_{\mathcal{L}}) \propto \prod_{i \in \mathcal{U}} \Pr(Y_i | x_i)$. Moreover, in multi-label classification, the simplest solution (i.e. one-vs-all) assumes that the inference of each label is also independent from other labels for an sample, i.e. $\Pr(Y_i | x_i) = \prod_{k=1}^q \Pr(Y_i^k | x_i)$. However, in many real-world classification tasks, there are complex dependencies not only among different samples but also among different labels.

To leverage the rich multi-modality information encoded in CelebrityNet, we explicitly consider three types of relationships in our model. We adopt the multi-kernel learning framework (MKL) [Vishwanathan et al. 2010] and build one kernel on each type of relationship. SVMs have been widely used for classification problems in recent years. Kernel selection has been an important factor of the performance of SVMs. Instead of

using one single kernel as the traditional SVM does, MKL incorporates multiple kernels and can learn a convex combination of these kernels (i.e., the kernel weights) simultaneously $\mathbf{K} = \sum_i \beta_i \mathbf{K}_i$. Specifically, we build three different kernels that can capture three different types of relationship in the data.

7.1. Content Relationship

Visual content of an image is often directly related to the tags and person(s) appear in it. The first type of relationships we consider is about the visual content features of the samples. Conventional image annotation approaches focus on using the image content features to build inference models. In order to capture the content/visual information of different samples, we build the *content kernel* based upon the input visual feature vector of different samples. $K_{content}(i, j) = \phi(\mathbf{x}_i, \mathbf{x}_j)$. Here, any conventional kernel function can be used for $\phi(\cdot, \cdot)$. Intuitively, the content kernel denotes the relationship that if two images share similar visual features, they are more likely to have similar labels.

7.2. Label Set Relationship

Different labels are correlated to each other in multi-label classification, thus should be predicted collectively. For example, in image annotation tasks, an image is more likely to have the tag ‘sports’ if the image has already been assigned with the tag ‘NBA’ or ‘basketball’. The image is less likely to be annotated as ‘sports’, if we already know the image has the label ‘academy awards’. Therefore, we explicitly model the label correlations within the label set of each sample as the second type of relationships. This label correlations can be learned during the training process to infer labels on unlabeled images during test.

Conventional multi-label classification approaches focus on exploiting such label correlations to improve the classification performances, which model $\Pr(Y_i^k | \mathbf{x}_i, \mathbf{Y}_i^{\{-k\}})$. $\mathbf{Y}_i^{\{-k\}}$ represents the vector of all the variables in the set $\{Y_i^p : p \neq k\}$. Hence, we have $\Pr(\mathbf{Y}_i | \mathbf{x}_i) \propto \prod_{k=1}^q \Pr(Y_i^k | \mathbf{x}_i, \mathbf{Y}_i^{\{-k\}})$. Based upon the above observation, we build the *label set kernel* encoding the correlations among different labels. $K_{labelset}(Y_i^k, Y_j^k) = \phi(\mathbf{Y}_i^{\{-k\}}, \mathbf{Y}_j^{\{-k\}})$. In our experiment, we simply use linear kernel to prove the concept, in which the function corresponds to the inner product of the two vectors. Intuitively, the label set kernel denotes the relationship that if two images share similar label sets, they are more likely to have similar values in any label variable.

7.3. Network Relationship

The label sets of related samples are usually inter-dependent in a network. For example, in our CelebrityNet network, the probability of an image having the label ‘politics’ should be higher if we already know the image contains the same people appearing in some other images with a label set of {‘government’, ‘politics’}. The third type of relationships we consider is the correlations among label sets of the related samples that are connected in the network. Conventional collective classification approaches focus on exploiting this type of dependencies to improve the classification performances, which models $\Pr(Y_i^k | \mathbf{x}_i, \mathbf{Y}_{j \in \mathcal{N}(i)})$. Here $\mathbf{Y}_{j \in \mathcal{N}(i)}$ denotes the set containing all vectors \mathbf{Y}_j ($\forall j \in \mathcal{N}(i)$), and $\mathcal{N}(i)$ denotes the index set of related samples to the i -th sample, i.e. the samples directly linked to the i -th sample. Hence, we will have $\Pr(\mathbf{Y}_i^k | \mathbf{X}) \propto \prod_{i \in \mathcal{U}} \Pr(Y_i^k | \mathbf{x}_i, \mathbf{Y}_{j \in \mathcal{N}(i)})$. Based upon the above observation, we build the *network kernel* encoding the correlations among related samples that are connected in the network as $K_{network}(Y_i^k, Y_j^k) = \phi(\mathbf{Y}_{l \in \mathcal{N}(i)}, \mathbf{Y}_{l \in \mathcal{N}(j)})$. Similarly, we use the linear kernel to represent the network kernel. Intuitively, the network kernel denotes the

Input:
 \mathcal{G} : a network, \mathcal{X} : attribute vectors for all instances.
 $\mathcal{Y}_{\mathcal{L}}$: label sets for the training instances, A : a base learner for multi-kernel learning model, T_{\max} : maximum # of iteration (default=10)

Training:
- Learn the MKL model f :
1. Construct q extended training sets $\forall 1 \leq k \leq q, \mathcal{D}_k = \{(\mathbf{x}_i^k, y_i^k)\}$ by converting each instance \mathbf{x}_i to \mathbf{x}_i^k as follows:
 $\mathbf{x}_i^k = (\mathbf{x}_i, \text{LabelSetFeature}(\ell_k, \mathbf{Y}_i), \text{NetworkFeature}(i, \mathcal{Y}_{\mathcal{L}}))$
2. Compute the corresponding kernels for each label: Φ , Φ_{Labelset} , and Φ_{network}
3. Calculate kernel weights and train MKL models on each label. Let $f_k = A(\mathcal{D}_k)$ be the MKL model trained on \mathcal{D}_k .

Bootstrap:
- Estimate the label sets, for $i \in \mathcal{U}$: produce an estimated values $\hat{\mathbf{Y}}_i$ for \mathbf{Y}_i as follows: $\hat{\mathbf{Y}}_i = f((\mathbf{x}_i, \mathbf{0}))$ using attributes only.

Iterative Inference:
- Repeat until convergence or #iteration $> T_{\max}$
1. Construct the extended testing instance by converting each instance \mathbf{x}_i to \mathbf{x}_i^k 's ($i \in \mathcal{U}$) as follows:
 $\mathbf{x}_i^k = (\mathbf{x}_i, \text{LabelsetFeature}(\ell_k, \hat{\mathbf{Y}}_i), \text{NetworkFeature}(i, \mathcal{Y}_{\mathcal{L}} \cup \{\hat{\mathbf{Y}}_i | i \in \mathcal{U}\}))$
2. Update the estimated value $\hat{\mathbf{Y}}_i$ for \mathbf{Y}_i on each testing instance ($i \in \mathcal{U}$) as follows: $\forall 1 \leq k \leq q, \hat{\mathbf{Y}}_i^k = f_k(\mathbf{x}_i^k)$.

Output:
 $\hat{\mathcal{Y}}_{\mathcal{U}} = (\hat{\mathbf{Y}}_1, \dots, \hat{\mathbf{Y}}_{n_u})$: the label sets of testing instances ($i \in \mathcal{U}$).

Fig. 8. The MKML algorithm

relationship that if the neighbors of the two images in the network share similar label sets, these two images are more likely to have similar label sets.

The general idea is as follows: we build one kernel on each type of the relations mentioned above, and then use MKL method to learn the weights of the multiple kernels (i.e., the importance of different kernels). We model the joint probability based upon the Markov property: if sample \mathbf{x}_i and \mathbf{x}_j are not directly connected in network G , the label set \mathbf{Y}_i is conditional independent from \mathbf{Y}_j given the label sets of all \mathbf{x}_i 's neighbors. The local conditional probability on label k can be modeled by a MKL learner with aforementioned kernels. The computation of these kernels depends on the predicted \mathbf{Y}_j ($j \in \mathcal{N}(i)$) and the predicted $\mathbf{Y}_i^{\{-k\}}$. Then, the joint probability can be approximated based on these local conditional probabilities by treating different labels as independent and the samples as *i.i.d.*. To simply demonstrate the effectiveness of our approach, we use linear kernels for all relations here.

Inspired by the ICA framework [Lu and Getoor 2003; McDowell et al. 2007], we design the following inference procedure of our MKML method as shown in Figure 8.

- (1) At the beginning of the inference, the label sets of all the unlabeled samples are unknown. The *bootstrap* step is used to assign an initial label set for each sample using the content feature of each sample. In our current implementation, we simply initialize the label set features and the network features for unlabeled samples with all zero vectors. Other strategies can also be used for bootstrapping, e.g, training SVM (single kernel) on training data using content feature only, and then we use these models to assign the initial label sets of unlabeled samples.
- (2) In the *iterative inference* step, we iteratively update the label set features/kernels and network features/kernels based upon the predictions of MKL models and update the prediction of MKL models using the newly updated kernels. The iterative process stops when the predictions of all MKL models are stabilized or a maximum number of iteration has been reached.

8. IMAGE ANNOTATION AND COMMUNITY CLASSIFICATION OVER CELEBRITYNET

As mentioned earlier in the paper, visual content, semantic information and the social network structure are mutually beneficial to each other. With the model introduced above, we evaluate the effectiveness of jointly modeling of multi-modality data including our implicit social network structure in multimedia tasks in this section. In test

time, only visual content of the images are provided. Label set features and network features are bootstrapped following the inference procedure described at the end of Section 7.

8.1. Compared Methods

We compare a set of methods exploring different types of information resource:

BSVM (binary SVM). This baseline method uses binary decomposition to train one classifier on each label separately, which is similar to [Boutell et al. 2004]. BSVM assumes all the labels and all instances are independent. It is based on visual content alone.

MKL (Multi-kernel learning). We directly apply multiple kernel learning algorithm on the joint information of the visual, semantic and social network without iterative inference steps.

KML (visual kernel + multi-label kernel). This baseline method trains one multi-kernel learner on each label, using two different kernels visual feature kernel and multi-label kernel. KML not only models the correspondence between the visual content and the tags, but also models the correlation among the tags.

MKICA (visual kernel + network kernel). In this baseline method, the multi-label dataset is first divided into multiple single-label datasets by one-vs-all binary decomposition. For each binary classification task, we use a multi-kernel version of ICA [Lu and Getoor 2003], as the base classification method. MKICA combines the social structure with visual modeling of the tags. However, it ignores the relationship among the tags.

MKML (Multi-kernel Multi-label Collective classification). Our proposed method for multi-label collective classification based upon multi-kernel learning, which jointly models the visual, semantic and social network information.

For a fair comparison, we use LibLinear [Fan et al. 2008] as the base classifier for BSVM and LibLinear MKL as the base learner for all the remaining methods. The maximum number of iterations in the methods KML, MKICA, and MKML are all set as 10 based on observation from the validation experiment.

8.2. Evaluation Metrics

We use the evaluation criteria proposed in [Ghamrawi and McCallum 2005; Kang et al. 2006; Liu et al. 2006] to verify the image annotation performance. Suppose a multi-label dataset \mathcal{D}_U contains n instances $(\mathbf{x}_i, \mathbf{Y}_i)$, where $\mathbf{Y}_i \in \{0, 1\}^q$ ($i = 1, \dots, n$). Denote $h(\mathbf{x}_i)$ as the predicted label set for \mathbf{x}_i by a multi-label classifier h , we have the harmonic mean of precision and recall as:

$$F1(h, \mathcal{D}_U) = \frac{2 \times \sum_{i=1}^n \|h(\mathbf{x}_i) \cap \mathbf{Y}_i\|_1}{\sum_{i=1}^n \|h(\mathbf{x}_i)\|_1 + \sum_{i=1}^n \|\mathbf{Y}_i\|_1}$$

The larger the F1 value, the better the performance.

All experiments are conducted on a machine with Intel Xeon™Quad-Core CPUs of 2.26 GHz and 24 GB RAM. We tested the performances on the following tasks:

- (1) image annotation task: we have 102,565 images with 159 frequent tags, each of which appears in at least 5% of the images. Each image can be annotated with a subset of these tags. On each image we extracted 5000 dimensional visual features in bag-of-words representation. We then randomly sample two thirds of the images as the training set, and use the remaining images as the test set.
- (2) community classification: we have 554 people in the dataset, where each person can be classified into a subset of 80 candidate communities. We randomly sample

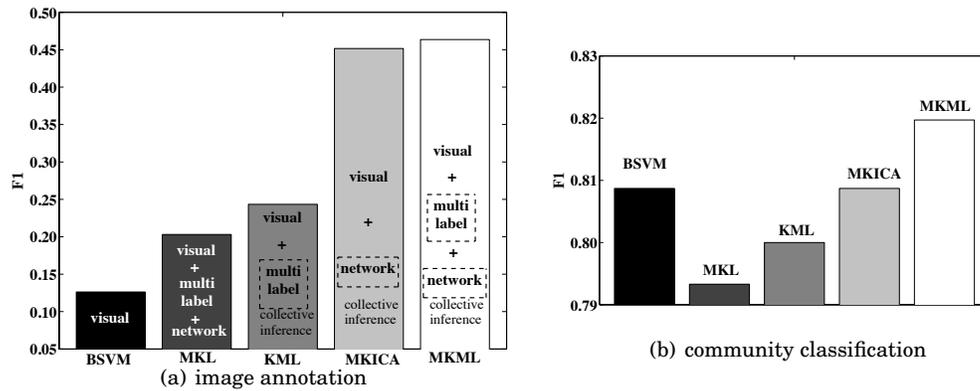


Fig. 9. Overall performances of the compared methods.

436 people into the training set, and use the remaining 118 people as the test set. For each person, we use the aggregated visual features of all his/her photos. Two persons are linked together if they appeared in at least one photo.

8.3. Results

In this section, we demonstrate results of two visual recognition tasks to show the advantage of jointly modeling visual content, semantic information and the social network structure. Specifically, image annotation task illustrates the potential of our approach for predicting semantic information based on visual content and the social network structure. Community classification of unknown person based on his/her set of photos and related tags demonstrates the possibility of using visual and semantic tags for social network structure prediction.

In Figure 9(a), we make the following observations:

- (1) The visual content based approach BSV achieves reasonably good performance in image annotation⁴, indicating strong correlation between the visual content and the tags.
- (2) Learning the correlation among tags is helpful, reflected by the substantial improvement of KML over BSV. This improvement is understandable: a photo with tag ‘NBA’ usually has ‘basketball’ in the tag list as well.
- (3) Methods MKICA and MKML significantly outperform the other methods indicating that incorporating the social network structure is especially useful. From the analysis of social network in Section 5, we learn that images and tags belonging to the same person and same community are very characteristic. Therefore, modeling the social network structure naturally improve the tag prediction performance of unknown images.
- (4) The significant performance improvement of MKML over the traditional MKL is largely attributed to the power of iterative prediction and error correction in our method.
- (5) Finally, jointly modeling the visual, semantic and social structure (MKML) provides additional improvement over combining visual and semantic information (KML), demonstrating the effectiveness of social network structure.

In Figure 9(b), we show the community classification results of different algorithms. In this experiment, tag correlation (KML) is not as useful as it is in the image anno-

⁴Random approach achieves only 0.03 by using the F1 measure.

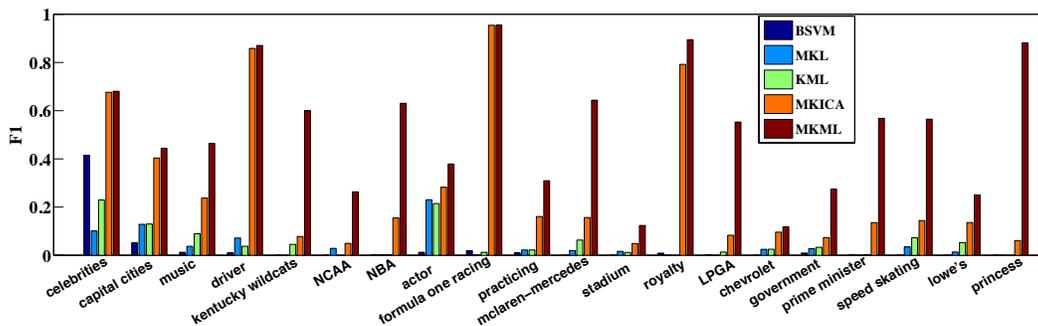


Fig. 10. F1 scores on example labels in image annotation task.

tation task. This is interpretable: as long as we know the person is related to the tag ‘NBA’, we can already do a good community class prediction without knowing other tags. On the other hand, if we know whom the unknown person is connected to, it is fairly easy to predict his/her community. This naturally leads to the good performance of social network based algorithms.

To provide more details of the annotation result, we show the F1 scores of example labels in Figure 10. While we observe similar pattern as in Figure 9(a) with clear advantage of the social network based algorithms over the other methods, the social network based algorithms usually perform much better on specific labels with social meaning such as ‘kentucky wildcats’ and ‘royalty’. Such social meaning can not be inferred from the visual content. The observation aligns well with our motivation of incorporating social network structure as a source of complimentary information for high level visual recognition tasks.

Finally, we show example results of image annotation in Figure 11 and Figure 12. Visual only method provides conservative prediction of common tags correlated to the visual content. Incorporating social network upon the visual and semantic modeling enables the algorithms to be more accurate in image annotation. MKML further improves the tag annotation by exploring the tag correlation upon jointly modeling the three sources of information. For example, in the 3rd picture in Figure 11, the tags ‘princess’ and ‘Spanish royalty’ can only be inferred correctly by combining information from social networks and correlations with other tags (such as ‘royalty’).

9. CONCLUSIONS

In this paper, we proposed to construct a social network based upon large scale online images. Our social network reflects the relationship and interaction among the celebrities. To analyze the group behavior of celebrities, we conducted community detection by using the clique percolation method. The detected communities agree to a large extent with the human judgement of the potential communities in the social network. We demonstrated that images and tags in communities exhibit uniqueness in each community and diverseness over different communities, which inspired us to use a principled algorithm to jointly model the visual content, semantic information and relationship structure for a few multimedia tasks. We show striking results on improving the image annotation and classification performance by using our proposed mechanism. We demonstrate that social networks and communities automatically generated from large scale image dataset could be very useful for high level visual tasks. This could be further extend to community based social networks like Flickr and Picasa. With the advancement of visual feature research especially the recent advances in deep learning for feature learning, visual classification has achieved significant improvement in

Test Image	Annotation
	<p>Ground Truth: capital cities, event, international landmark, politics, government, healthcare and medicine</p> <p>BSVM: arts culture and entertainment, celebrities, portrait, politics</p> <p>MKL: arts culture and entertainment, politics, television show</p> <p>KML: arts culture and entertainment, celebrities, attending, capital cities, government, politics, international landmark, television show</p> <p>MKICA: portrait, international landmark, politics</p> <p>MKML: capital cities, event, international landmark, politics, government</p>
	<p>Ground Truth: arts culture and entertainment, attending, capital cities, royalty, spanish royalty</p> <p>BSVM: arts culture and entertainment, celebrities</p> <p>MKL: arts culture and entertainment, capital cities, attending</p> <p>KML: arts culture and entertainment, capital cities, attending</p> <p>MKICA: capital cities, politics, attending, capital cities</p> <p>MKML: arts culture and entertainment, attending, capital cities, royalty, spanish royalty</p>
	<p>Ground Truth: arts culture and entertainment, royalty, spanish royalty, visit, princess</p> <p>BSVM: arts culture and entertainment, sport</p> <p>MKL: arts culture and entertainment, sport</p> <p>KML: arts culture and entertainment, sport, politics</p> <p>MKICA: capital cities, politics, royalty</p> <p>MKML: arts culture and entertainment, attending, royalty, spanish royalty, princess, capital cities</p>
	<p>Ground Truth: capital cities, meeting, politics, government, prime minister, british culture, diplomacy, conference</p> <p>BSVM: arts culture and entertainment, politics</p> <p>MKL: arts culture and entertainment, capital cities, politics</p> <p>KML: arts culture and entertainment, celebrities, capital cities, attending, actor, premiere, politics</p> <p>MKICA: capital cities, politics, prime minister, british culture</p> <p>MKML: capital cities, politics, prime minister, british culture, diplomacy</p>
	<p>Ground Truth: arts culture and entertainment, celebrities, actor, film industry, actress, academy awards, movie, the kodak theatre</p> <p>BSVM: arts culture and entertainment</p> <p>MKL: arts culture and entertainment, celebrities</p> <p>KML: arts culture and entertainment, celebrities, capital cities, hold, music, sport, politics, film industry, driver, actress</p> <p>MKICA: arts culture and entertainment, celebrities, film industry</p> <p>MKML: arts culture and entertainment, celebrities, film industry, actor, premiere, movie</p>
	<p>Ground Truth: arts culture and entertainment, celebrities, hold, music, singer, grammy awards</p> <p>BSVM: arts culture and entertainment, celebrities</p> <p>MKL: arts culture and entertainment, celebrities, attending, music</p> <p>KML: arts culture and entertainment, celebrities, attending, hold, music, sport, actor, motorized sport, film festival, movie, politics</p> <p>MKICA: arts culture and entertainment, music</p> <p>MKML: arts culture and entertainment, celebrities, hold, music, grammy awards</p>

Fig. 11. Annotation Examples of different algorithms. BSVM, MKL, KML, MKICA, MKML represent binary SVM, traditional multi-kernel learning method, method built upon visual kernel + multi-label kernel, method built upon visual kernel + network kernel, Multi-kernel Multi-label Collective classification method respectively. Tags correctly recognized by the methods are highlighted in green. Incorrect tags are in purple. (Photo credit from top to bottom: Chip Somodevilla/Getty Images News, Chip Somodevilla/Getty Images Entertainment, Carlos Alvarez/Getty Images Entertainment, Daniel Berehulak/Getty Images News, Carlo Allegri/Getty Images Entertainment, Frazer Harrison/Getty Images Entertainment)

ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 9, No. 4, Article 39, Publication date: September 2014.

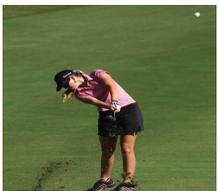
Test Image	Annotation
	<p>Ground Truth: sport, basketball - men's college, ball, people, duke blue devils, stadium SVM: sport, basketball - men's college, ball, duke blue devils MKL: sport, basketball - men's college, ball KML: sport, basketball - men's college, ball, people, making a basket, taking a shot - sport MKICA: sport, basketball - men's college, basketball, duke blue devils MKML: sport, basketball, basketball - men's college, ball, duke blue devils, stadium</p>
	<p>Ground Truth: sport, golf, lgpa BSVM: arts culture and entertainment, celebrities, sport MKL: arts culture and entertainment KML: arts culture and entertainment, sport, politics MKICA: sport, golf MKML: sport, golf, lgpa</p>
	<p>Ground Truth: sport, motorized sport, formula one racing, practicing, qualification round, driving BSVM: celebrities, sport, motorized sport MKL: arts culture and entertainment, celebrities, sport KML: arts culture and entertainment, celebrities, attending, movie, hold, sport, film industry, fashion, formula one racing, chevrolet MKICA: sport, motorized sport, formula one racing, ferrari MKML: sport, motorized sport, formula one racing, qualification round, driving, ferrari</p>
	<p>Ground Truth: sport, lead, winter sport, speed skating BSVM: arts culture and entertainment, celebrities, sport, film industry MKL: arts culture and entertainment, actor KML: arts culture and entertainment, sport, politics MKICA: sport, winter sport MKML: sport, winter sport, speed skating</p>
	<p>Ground Truth: sport, basketball - nba pro, nba, making a basket BSVM: celebrities, sport MKL: sport, making a basket KML: arts culture and entertainment, sport, making a basket, politics MKICA: sport, basketball - nba pro, ncaa college conference team, nba, making a basket, stadium, duke blue devils MKML: sport, basketball - nba pro, nba, making a basket</p>
	<p>Ground Truth: sport, hitting, taking a shot - sport, competition round, golf, lgpa BSVM: award, sport, motorized sport, stock car racing, driver MKL: arts culture and entertainment, celebrities, hold, sport KML: arts culture and entertainment, sport, hold, award, motorized sport, driver, radio, television show, theatrical performance MKICA: sport, taking a shot - sport, golf, lgpa, hole MKML: sport, hitting, taking a shot - sport, competition round, golf, lgpa, hole</p>

Fig. 12. Annotation Examples of different algorithms. BSVM, MKL, KML, MKICA, MKML represent binary SVM, traditional multi-kernel learning method, method built upon visual kernel + multi-label kernel, method built upon visual kernel + network kernel, Multi-kernel Multi-label Collective classification method respectively. Tags correctly recognized by the methods are highlighted in green. Incorrect tags are in purple. (Photo credit from top to bottom: Streeter Lecka/Getty Images Sport, Stephen Dunn/Getty Images Sport, Clive Rose/Getty Images Sport, Matthew Stockman/Getty Images Sport, Gregory Shamus/Getty Images Sport, Darren Carroll/Getty Images Sport)

the past few years. However, many classes including small, metallic, see-through and highly varies scenes [Russakovsky et al. 2014] still remains extremely challenging for visual based approaches such as the successful deep learning approach [Krizhevsky et al. 2012]. Our social network information serves as a complementary knowledge base to the visual content and semantic information, which is promising to improve upon the challenging ones together with advanced visual representation such as that learned by using deep learning approaches. In the future, we would also like to use our social network together with image and tag properties of an unknown person to predict the connections of him/her to the rest of the social network. As well as, there is the potential of expansion of using our social network in other high level recognition tasks such as image retrieval.

REFERENCES

- S. Bakhshi, D. A. Shamma, and E. Gilbert. 2014. Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, ACM, Toronto, Canada. To appear. 2
- M.A. Beauchamp. 1965. An improved index of centrality. *Behavioral Science* 10, 2 (1965), 161–163. 5
- M. R. Boutell, J. Luo, X. Shen, and C. M. Brown. 2004. Learning multi-label scene classification. *Pattern Recognition* 37, 9 (2004), 1757–1771. 15
- S. Brin and L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems* 30, 1 (1998), 107–117. 3
- L. Cao, J. Yu, J. Luo, and T. S. Huang. 2009. Enhancing Semantic and Geographic Annotation of Web Images via Logistic Canonical Correlation Regression. In *Proceedings of the 17th ACM International Conference on Multimedia (MM '09)*. ACM, New York, NY, USA, 125–134. DOI: <http://dx.doi.org/10.1145/1631272.1631292> 3
- Y-Y Chen, W. H. Hsu, and H-Y M. Liao. 2012. Discovering Informative Social Subgraphs and Predicting Pairwise Relationships from Group Photos. In *Proceedings of the 20th ACM International Conference on Multimedia (MM '12)*. ACM, New York, NY, USA, 669–678. DOI: <http://dx.doi.org/10.1145/2393347.2393439> 2, 3
- N. Dalal and B. Triggs. 2005. Histograms of oriented gradients for human detection. In *CVPR*. 886. 3
- J. Deng, A. Berg, K. Li, and L. Fei-Fei. 2010. What does classifying more than 10,000 image categories tell us? *ECCV (2010)*, 71–84. 3
- I. Derényi, G. Palla, and T. Vicsek. 2005. Clique Percolation in Random Networks. *Phys. Rev. Lett.* 94 (2005). 7
- L. Ding and A. Yilmaz. 2010. Learning relations among movie characters: A social network perspective. *ECCV (2010)*, 410–423. 2, 3
- Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. *LIBLINEAR: A library for large linear classification*. 15
- P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. 2007. Object Detection with Discriminatively Trained Part Based Models. *JAIR* 29 (2007). 3
- L.C. Freeman. 1977. A set of measures of centrality based on betweenness. *Sociometry (1977)*, 35–41.
- W. Gao and K. Wong. 2006. Natural document clustering by clique percolation in random graphs. In *Proceedings of the Third Asia conference on Information Retrieval Technology*. Springer, Singapore, 119–131. 7
- N. Ghamrawi and A. McCallum. 2005. Collective multi-label classification. In *CIKM*. Bremen, Germany, 195–200. 15
- M. C. González, H. J. Herrmann, J. Kertész, and T. Vicsek. 2007. Community structure and ethnic preferences in school friendship networks. *Phys. A* 379, 1 (2007), 307–316. 7
- P.R. Gould. 1967. On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers (1967)*, 53–86. 6
- P. Jaccard. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles (1901)*.
- F. Kang, R. Jin, and R. Sukthankar. 2006. Correlated Label Propagation with Application to Multi-label Learning. In *CVPR*. New York, NY, 1719–1726. 15
- H-N Kim, J-G Jung, and A. El Saddik. 2010. Associative Face Co-occurrence Networks for Recommending Friends in Social Networks. In *Proceedings of Second ACM SIGMM Workshop on Social Media (WSM '10)*. ACM, New York, NY, USA, 27–32. DOI: <http://dx.doi.org/10.1145/1878151.1878160> 3

- I. Konstas, V. Stathopoulos, and J.M. Jose. 2009. On social networks and collaborative recommendation. In *SIGIR*. ACM, 195–202. 3
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105. 20
- J. Leskovec, K.J. Lang, and M. Mahoney. 2010. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*. ACM, 631–640. 7
- L-J. Li, X. Kong, and P. S. Yu. 2014. Visual Recognition by Exploiting Latent Social Links in Image Collections. In *MultiMedia Modeling*. Springer, 121–132. 8, 12
- Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang. 2011. Large-scale image classification: fast feature extraction and svm training. In *CVPR*. 1689–1696. 3
- X. Liu, Z. Shi, Z. Li, X. Wang, and Z. Shi. 2010. Sorted label classifier chains for learning images with multi-label. In *ACM MM*. 3
- Y. Liu, R. Jin, and L. Yang. 2006. Semi-supervised multi-label learning by constrained non-negative matrix factorization. In *AAAI*. Boston, MA, 421–426. 15
- Q. Lu and L. Getoor. 2003. Link-based classification. In *ICML*. 2, 14, 15
- Yong Luo, Dacheng Tao, Bo Geng, Chao Xu, and Stephen J Maybank. 2013. Manifold regularized multitask learning for semi-supervised multilabel image classification. *Image Processing, IEEE Transactions on* 22, 2 (2013), 523–536. 3
- L. K. McDowell, K. M. Gupta, and D. W. Aha. 2007. Cautious inference in collective classification. In *AAAI*. Vancouver, Canada, 596–601. 14
- G. Palla, I. Derényi, I. Farkas, and T. Vicsek. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 7043 (2005), 814–818. 7
- Guo-Jun Qi, Xian-Sheng Hua, Yong Rui, Jinhui Tang, Tao Mei, and Hong-Jiang Zhang. 2007. Correlative multi-label video annotation. In *Proceedings of the 15th international conference on Multimedia*. ACM, 17–26. 3
- J. Read, B. Pfahringer, G. Holmes, and E. Frank. 2009. Classifier Chains for Multi-label Classification. In *ECML/PKDD*. 3
- Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2014. ImageNet Large Scale Visual Recognition Challenge. (2014). 20
- Z. Stone, T. Zickler, and T. Darrell. 2008. Autotagging Facebook: Social network context improves photo annotation. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*. 1–8. DOI: <http://dx.doi.org/10.1109/CVPRW.2008.4562956> 3
- Z. Stone, T. Zickler, and T. Darrell. 2010. Toward large-scale face recognition using social network context. *Proc. IEEE* 98, 8 (2010), 1408–1415. 2, 3
- Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2005. *Introduction to Data Mining*. Addison Wesley. 9
- S. V. N. Vishwanathan, Z. Sun, and N. Theera-Ampornpant. 2010. Multiple Kernel Learning and the SMO Algorithm. In *NIPS*. Vancouver, B.C., Canada, 2361–2369. 12
- G. Wang, A. Gallagher, J. Luo, and D. Forsyth. 2010. Seeing people in social context: Recognizing people and social relationships. In *ECCV*. Springer, Crete, Greece, 169–182. 2, 3
- S. Wasserman and K. Faust. 1994. *Social network analysis: Methods and applications*. (1994). 5
- J. Weston, S. Bengio, and N. Usunier. 2010. Large scale image annotation: Learning to rank with joint word-image embeddings. *Machine learning* 81, 1 (2010), 21–35. 3
- Zheng-Jun Zha, Xian-Sheng Hua, Tao Mei, Jingdong Wang, G-J Qi, and Zengfu Wang. 2008. Joint multi-label multi-instance learning for image classification. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 1–8. 3
- J. Zhuang, T. Mei, S. Hoi, X. Hua, and S. Li. 2011. Modeling social strength in social media community via kernel-based learning. In *ACM MM*. 2, 3