Chapter 18
Friends with Faces: How Social Networks Can Enhance Face Recognition and Vice Versa

Nikolaos Mavridis, Wajahat Kazmi, and Panos Toulis

Abstract The "friendship" relation, a social relation among individuals, is one of the primary relations modeled in some of the world's largest online social networking sites, such as "FaceBook." On the other hand, the "co-occurrence" relation, as a relation among faces appearing in pictures, is one that is easily detectable using modern face detection techniques. These two relations, though appearing in different realms (social vs. visual sensory), have a strong correlation: faces that co-occur in photos often belong to individuals who are friends. Using real-world data gathered from "Facebook," which were gathered as part of the "FaceBots" project, the world's first physical face-recognizing and conversing robot that can utilize and publish information on "Facebook" was established. We present here methods as well as results for utilizing this correlation in both directions. Both algorithms for utilizing knowledge of the social context for faster and better face recognition are given, as well as algorithms for estimating the friendship network of a number of individuals given photos containing their faces. The results are quite encouraging. In the primary example, doubling of the recognition accuracy as well as a sixfold improvement in speed is demonstrated. Various improvements, interesting statistics, as well as an empirical investigation leading to predictions of scalability to much bigger data sets are discussed.

18.1 Introduction

The work presented here was carried out as part of the "FaceBots" research project,[1] whose original purpose was to show that references to shared memories as well as to shared friends can enhance long-term human robot relationships. Towards that

N. Mavridis (✉) and W. Kazmi
Interactive Robots and Media Lab, United Arab Emirates University, 17551 Al Ain, United Arab Emirates
e-mail: irmluaeu@gmail.com; wajahat.kazmi@em-a.eu

P. Toulis
ISSEL laboratory, Aristotle University of Thessaloniki, Greece
e-mail: ptoulis@issel.ee.auth.gr

[1] We thank Microsoft External Research for providing seed funding for this project through its Human-Robot Interaction CFP.

A. Abraham et al. (eds.), *Computational Social Network Analysis*, Computer Communications and Networks, DOI 10.1007/978-1-84882-229-0_18,

purpose, a special physical robot was created, with face recognition (FR) as well as spoken natural-language dialogue capabilities, which was also equipped with an interaction as well as a social database [1]. Furthermore, this robot ("Sarah Mobileiro the FaceBot") is able to connect in real time to the FaceBook online networking web site, and thus was the world's first robot that was able to utilize and publish online social information. Sarah is able to perform training as well as recognition not only from its camera-derived photos, but also from online photos, posted on FaceBook, which might also contain tags. Thus, all the ingredients were there in order to explore the idea of trying to utilize social context towards better FR, as well as trying to find out who might be friends with whom on the basis of photos.

18.1.1 Background

Context-assisted visual recognition is a highly promising research area, and some attempts already exist, as we shall see. In contrast, very few attempts exist towards utilizing FR on images belonging to online social networking (SN) web sites (for example [2], without the utilization of context). On the other hand, as noted above, some methods for context-assisted object recognition have appeared in the literature recently: Torralba [3] provides an example of contextual priming for object recognition, based on holistic context representations, while Hoiem et al. [4] perform object detection by modeling the interdependence of objects, surface orientations, and camera viewpoint. However, none of these papers addresses the utilization of social context for FR; the only noteworthy exception is [5]. There is an important difference though between this paper and the methods we are presenting here as Stone et al. [5] only use the identity of the person contributing the photo to the online networking web site in order to enhance the recognition and the method only works if this is known. In contrast, our method does not require this information. It can be seeded by the social context created through postulated or recognized participants in the photo, and is much more flexible in that respect, and can thus be used also on photos with no submitting author information, arising anywhere on the Internet or live. Finally, it is worth noting that apart from the mutual benefits between FR and online networks, there exists a whole triangle of synergies between the two and interactive robotics, as demonstrated on Sarah the FaceBot, and discussed in [6]. And from a wider viewpoint, we hope for this chapter to also act as a concrete demonstration that very promising avenues exist at the crossroads between social networks and numerous other areas, many of which remain yet unexplored.

18.1.2 Overview

This chapter is structured around two sections: Section 18.2 addresses the enhancement of FR through social knowledge, whereas Sect. 18.3 discusses the acquisition of social knowledge through photos containing faces. In the first section, we introduce basic concepts, problems and classes of algorithms, describe the algorithms in detail, and provide thorough coverage of a real world empirical investigation of the performance of the proposed methods on a data set arising from the friends of our robot and their facebook profiles. We present interesting statistics, results, and conclusions, as well as a further investigation of the expected performance of the algorithm as the size of the friendship network grows, that is, a prediction of scalability-related issues. In Sect. 18.2, we discuss the converse problem of estimating friendship through photos, propose algorithms, and provide real-world results. A conclusion closes the chapter, together with appendices regarding notation and proofs.

18.2 Utilizing Social Context Towards Face Recognition

18.2.1 Basic Concepts

Our *purpose* here is to illustrate how social context can help towards enabling faster as well as more accurate FR in photos. We will do so by presenting a number of basic problems, algorithms and variations, and finally concrete results for a real-world example concerned with facebook photos accessible through a conversational interactive physical robot. The *salient idea* behind our illustration is simple to state: *co-occurrence* of faces in photos and *friendship* have a strong correlation between them. As we shall see, in our real-world experiments, two random tagged faces within a facebook photo had a probability of almost 80% of being faces of declared first-level friends within facebook. Thus, one can expect that knowledge of the friendship relationships among individuals can assist towards predicting co-occurrence of them in photos, and consequently towards better FR in photos with more than one faces.

From a higher-level viewpoint, one can conjecture *three realms* implicated in the setting of this discussion: first, a *social realm*, in which identities are entities, and friendship a relation; second, a *visual sensory realm*, of which faces are entities, and co-occurrence in images a relation; and third, a *physical realm*, in which bodies belong, with physical proximity being a relation. The frequent physical proximity of bodies of friends, as they engage in activities and interactions together, is imprinted in photos and correlates with co-occurrence of their faces; and thus photo co-occurrence (a sensory-domain relation among regions in images) correlates with friendship (a social relation among individuals).[2]

[2] When viewed from a slightly different viewpoint, here we have a *sensory grounding* not of a conceptual *entity* (as is often the case in language grounding research), but of a social-level *relation* among entities, in a manner similar to grounding ontologies. Also, one might conjecture that the actual grounding of the social-level relation of friendship, might start during development initially from a restricted tangible meaning: that of the bodies of two individuals often being close and interacting. This restricted meaning is later extended during development in order to include social-level attributes that might include co-operation, sincerity, etc.

Before we proceed, let us introduce some *basic notation* in order to clarify our exposition:

Identity (*Id*) An individual, which might or might not have a facebook Id, and whose name might or might not appear as a tag in a facebook photo.

Face (F) A region of an image corresponding to a human face, having ultimately been generated through the visual sensory effect of an underlying Identity.

Photo (Ph) An image potentially containing multiple faces, which might or might not be a photo available within facebook.

Tag (T) A string which has been entered by a facebook user in order to identify a face in a photo. This might or might not be equal to the facebook name of an identity, if the identity belongs to facebook.

Classifier (Cl) A black-box abstraction of a pattern recognizer, whose input is an image region (detected as having been a face in our case) and output a measure of likelihood of this face having been the sensible emission of an identity. Each classifier is trained through a training set consisting of faces, which are conjectured to belong to the target identity to be classified.

Friendship Relation (FR) A relationship among two identities.

Friendship Matrix (FM) A square matrix whose rows and columns are identities, and whose entries $FM_{i,j}$ are

$$\begin{array}{l} 1 \text{ (knowledge of friendship among identities } i \text{ and } j) \\ 0 \text{ (knowledge of non-friendship among } i \text{ and } j) \\ -1 \text{ (lack of knowledge about friendship of } i \text{ and } j) \end{array} \tag{18.1}$$

18.2.2 Basic Problems and Classes of Algorithms

Here, let us first introduce *three basic problems* that we will deal with in this section. In all three, our purpose is to recover the identities of the faces in the photo, but the number of available tagged faces differs:

Pr1) *Seeded Face-Rec* A photo is given, in which exactly one face is assumed to have a tag attached to it.

Pr2) *Unseeded Face-Rec* A photo is given, in which no faces are assumed to have a tag attached to them.

Pr3) *Multi-seed Face-Rec* A photo is given, in which more than one faces are assumed to have a tag attached to them.

Regarding evaluation of the effectiveness of solutions to the above problems, we distinguish between *two types of real-world usage* of the system: *fully-automated recognition (AR)*, with no human intervention, as well as *semi-automated recognition (SAR)*, in which the system, instead of offering a unique solution regarding the postulated identity of a face, offers a number of alternatives, which are presented to a human, who finally picks up the one he or she thinks is correct. The requirements of classification accuracy for the latter case are more relaxed; although we require the postulated identity to be equal to the actual identity for the case of AR (i.e., only one guess is possible and it should be correct), we only require the postulated identity to belong to a small set of identities proposed by the system for the case of SAR. A practically tractable size for the set of identities proposed by a SAR system is here taken to be 10 photos, which can be glanced upon by a human and selected within 10 s or less. Thus, we will quantify performance using primarily two metrics:

Rank1 Accuracy (for the Case of AR) The percentage of times during which the postulated identity of a face by a classifier is correct.
Rank10 Accuracy (for the Case of SAR) The percentage of times during which the correct identity of a face belongs to the ten most probable identities as postulated by a classifier.

Towards the solution of the basic problems introduced above, we will later provide more details on the following four basic classes of algorithms:

Alg0 Face-by-face recognition without use of social context
Alg1 Whole-photo rec utilizing social context (Pr1, single seed id known)
Alg2 As Alg1 but for problem Pr2, i.e. no known seed identity
Alg3 As Alg1 for Pr3 (multiple seed identities known)

18.2.3 The Proposed Algorithms in Detail

For the *four basic classes of algorithms* that we have defined above, here we provide a detailed description and discuss possible *variations*.

18.2.3.1 Algorithm 0

A photo is given containing multiple faces, but no knowledge of any friendship matrix is assumed. Each face is subsequently passed through the set of available classifiers, and scores are recorded. The identity giving the best score is reported (for the rank1 case for AR), or the identities of the top ten scores (for the rank10 case for SAR). Summing up:

Input A photo *(Ph)* containing *n* faces $F1 \cdots Fn$.
Output A vector of postulated identities (for the AR version of Alg1):

$$[Id(F1) \cdots Id(Fn)] \tag{18.2}$$

or a vector of 10-D vectors of postulated best-10 identities (SAR version):

$$[\{Id1(F1), Id2(F1) \cdots Id10(F1)\} \cdots \{Id1(Fn), Id2(Fn) \cdots Id10(Fn)\}] \tag{18.3}$$

18.2.3.2 Algorithm 1

A seed is already given, as well as some knowledge of the friendship matrix *(FM)* is assumed. Then, the first-level friends of the Id of the seed are recovered from the corresponding column of the friendship matrix, in the form of a friendship vector:

$$FV = [FR(i,1), FR(i,2), FR(i,3) \cdots FR(i,m)] \tag{18.4}$$

where

$$FR(i,j) = \begin{cases} 1 \\ 0 \\ -1 & \text{(as explained in Sect. 18.2.1)} \end{cases} \tag{18.5}$$

At this stage, two possible variants of the algorithm exist:

***Alg1H**: (Hard biasing)* For each face F_i in the photo *Ph*, scores are taken only for those classifiers whose entries at the friendship vector are 1, i.e. only for the classifiers whose identities are the known friends of the seed. Then, the rank1 or rank10 best identities are chosen, among the classifiers of the first level friends of the seed (for AR and SAR, respectively).

***Alg1S**: (Soft biasing)* For each face F_i in the photo *Ph*, the scores at the output of all the known classifiers are taken, say *(S1$\cdots$Sm)*. Then, biasing is accomplished through a biasing vector added to the score vector; this vector *BV* is calculated as a function of the friendship vector *FV* in the following way:

$$BV(i) = \begin{cases} \alpha_1 & \text{if } FV(i) = 1 \\ 0 & \text{if FV(i)} = 0 \\ \alpha_{-1} & \text{if } FV(i) = -1 \end{cases} \tag{18.6}$$

Thus, the two constant parameters a1 and a-1 determine the relative contribution to biasing for the known friendship relationships, as well as the unknown relationships. The optimal values of these parameters can be determined empirically, as discussed later in this paper. Finally, the rank1 or rank10 best identities are chosen, among the classifiers of the first-level friends of the seed (for AR and SAR, respectively).

***Alg1TS:** (Biasing According to Training Set Size)* An extra term is added to the biasing vector, to account for variance in the training set sizes of different classifiers. As the classifiers are trained through the facebook photos which contain tags for the classifier's identity, there is considerable variance in the number of photos available for training for each identity (as quantified later in this paper). Identities with larger training sets generally result to classifiers with more reliable outputs; and the converse holds for those identities that have small training sets. The biasing term has the form:

$$BSV(i) = \beta log(size(Tr(Id_i))) \tag{18.7}$$

where

$$Tr(Id_i) = \text{the available training set for Identity } i \tag{18.8}$$

This extra biasing term is added to the social biasing vector BV.

Input A photo Ph containing n faces $F1 \cdots Fn$, as well as a seed (i: seed is face F_i, and $Id(F_i)$)
Output Rank1 and Rank10 scores, as described in Alg0.

18.2.3.3 Algorithm 2

As in Algorithm1, but with no seed. Thus, a seed should be selected, then Alg1 carried out, and then possible results evaluated and potentially a different seed reselected. Thus, here we distinguish three of the possible variations:

Alg2RS Here, out of the n faces in the picture, one is randomly selected to serve as the seed (alternatively, the first face always serves as the seed). Its postulated identity is given by choosing the identity of the classifier that has the highest score on this face, out of all existing classifiers. Then, Alg1 is run, i.e., the friendship vector is created, hard or soft biasing takes place, etc. The problem with this approach is that, as we shall see, any mistake in the identity of the seed might have devastating consequences for correct recognition on the rest of the faces of the photo.

Alg2BS Here, the seed is not randomly selected; all the faces are taken in turn, and classified as belonging to one of all existing classifiers. The face which has the biggest score is then selected as the seed (i.e., the one for whose identity we appear to be more certain).

Alg2PE Here, the seed is not randomly selected; all the faces are taken in turn, and classified as belonging to one of all existing classifiers. Then each of these faces is taken as a possible seed, and Alg1 is applied n times in total, giving n total whole-photo classifications. Then, one is chosen out of all these n whole-photo hypothesis, through maximization of a suitable "total match" metric. The metric chosen could be, for example, the sum of square of the scores of the chosen identities across all faces in the photo.

18.2.3.4 Algorithm 3

In the class of algorithms referred to as *Alg3*, multiple seed photos might exist at a given time. Thus, these extend upon *Alg2* and consequently *Alg1*, in the following respects:

Higher-Level Friendships and Mutual Friendships Once more than one seed photo is postulated, there exist multiple radii of social circles (first-level friends of seed 1, first-level friends of seed 2, second-level friends of seed 1, etc.) as well as of intersections of circles (mutual friends of seed 1 and seed 2, mutual friends of seed 1 and seed 3, etc.) that can be taken into account. Each one of these is taking into account, for example, through a different weighing parameter when adding a soft bias. As an example, for maximum two seeds, and maximum radius 1, we get a wealth of combinations of possibilities: $\{-1, 0, 1\} \times \{-1, 0, 1\}$, i.e. the set of all classifiers is partitioned into nine possible subsets according to each classifier identity's friendship relationship with the two seeds.

Ongoing Rebiasing Instead of just biasing initially, and then classifying all remaining faces at once, with ongoing rebiasing one can successively increase the number of seeds while classifying, by incorporating as a new seed each new face that has been classified with highest confidence. For example, we might start with problem 2; i.e. no known seed faces. Then, we might chose as a seed the face that was classified with highest confidence without social information (say seed 1); and bias through it. Then, we might again chose the new face that was classified with highest confidence (among the remaining yet unclassified faces – let us call it seed 2), and add this to the seed set. Now, at this stage we can perform mutual biasing from the two seeds, as described in the above paragraph. The next classified face with highest confidence will also be added to the constantly expanding seed set, and mutual biasing will be performed again, until no more unclassified faces remain.

Backtracking The primary problem of ongoing rebiasing is that any wrong choice in the postulated identity of the face might have destructive effects for the next faces, as through social biasing with the wrong seed it might avert their correct recognition. One possible solution for this is the introduction of backtracking; for each face, at each stage, multiple, say n, possible identities are kept (the *rankN* best solutions). Also, either at the end (whole photo classified), or at intermediate stages (k photos classified so far), an overall photo-so-far confidence metric is evaluated. If the overall confidence metric is low, then the identity choices for faces so far that have been made are partially retracted, and the next possible identity (for example, *rank2*) is considered for the faces in question, as well as their combinations.

18.2.4 Empirical Investigation

18.2.4.1 Data Acquisition

Aspects of the Facebots project [6] required accessing and processing, a large amount of data, contributed by people in the "Facebook" networking site. Generally, these include friendship relations, photos, news updates and also data generated through user-to-user communication (messages, chats, etc.). The ranging sensitivity of personal information is, in most cases, directly equivalent to its degree of accessibility, and this rule holds also on how much of this information, our robot was able to access. The idea of a robot crawling information pages on Facebook is quite an interesting one, and is tightly intertwined with issues regarding access and openness, and so this section is devoted to further details regarding how our information gathering was achieved, described at the programming level.

18.2.4.2 Facebook Site Mechanism and Security

Facebook [7] is a popular SN, which currently (2009), allows around 200 million people to interact with each other [8]. The site has been built with strict security mechanisms that protect users' data from unprivileged access, which in part accounts for its big popularity. In what follows, we provide with some information about the inner workings of Facebook as an application, that enabled us to build the first social networked robot. However, the reader should keep in mind, that this information became available to the authors only by means of experimentation and reverse engineering, and that it is due to constant changes. Nevertheless, it is still an outline of how related research efforts can be performed.

To start with, Facebook is itself very strict with accounts that seem suspicious for spamming or for other than personal use. As an example, it is not possible to create accounts with names containing the distinct words "spam" or "bot" or reporting an age less than 18. Communication with the site is being carried through SSL, with the familiar cookie-based authentication. The benefits and vulnerabilities of this scheme are well known and the reader can refer to [9].

Upon login to the SN site, the user's browser is receiving two important id's : the post-form Id and a channel Id, the former being an hexadecimal number and the latter being actually a host name, while both of them are being used for enabling the communication of the user with the web site. In specific, almost any POST action will require for the post-form Id (e.g., updating user status, sending messages, etc.), while the channel Id identifies a Facebook server which provides with all of the instant messaging functionality : updating online friends' list, sending instant(chat) message, receiving chat messages, and others.

All sensitive information and operations, such as messages exchanged or new friendship connections, are not available or accessible by any means. Provided that the intrinsic Facebook cookie-mechanism remains un-compromised, an automated software entity cannot access data that has not been published by their owners. In addition, a bot cannot perform any bulk activities, such as sending messages, or sending friendship requests, without being able to solve known computationally hard problems (e.g., captchas).

18.2.4.3 Facebots Data Access

What data does our robot actually have access to and what operations does it perform for the purposes of our research? In order to answer this question, we will present in brief, but in a technical manner, what and how our robot accesses the information needed.

Our robot, the first FaceBot [1], Sarah Mobilero, has currently 76 first-level friends in its Facebook account, and the following functions related to Facebook have been implemented:

login Logins into Facebook and retrieves basic information, such as robot's Id, post-form Id and channel Id

get(status, friends, posted_photos, joined_groups, status_updates) Gets a number of available information such as friend lists or status updates, all related to the robot's friends.
set(status, String STATUS) Sets the robot's status to STATUS
composeMessage(String MESSAGE, int FID) Composes a new message to the friend with a Facebook Id equal to FID
chat(int FID, St ring CHAT_MESSAGE) Sends the instant message CHAT_MESSAGE to user FID

The mechanism for obtaining data and performing the aforementioned actions is uniform and in fact it is described by the term "HTML scraping," [10] which has been a known technique for network programming. The structured format of new Web 2.0 applications, relying heavily on frameworks such as the CSS or JSON, makes them consumable with a reasonable effort, by regular expressions, without the use of further messaging protocols on top of HTML (e.g. Web Services/SOAP). The use of the Facebook API [11], was not considered due to latency problems and functionality limitations. The idea is to basically emulate ordinary user-browser sessions through the automated use of the underlying HTML GET/POST requests. This way, a software entity can replay the interactions with a browser, and parse their output using predefined regular expressions, to generate the desired data structures (e.g. friend lists, photo updates). This way, the robot is able to do what a normal user can do using a regular browser when interacting with Facebook.

18.2.5 A First Look at the Data Sets

In order to access a big pool of photos that were contributed by Sarah's friends or photos in which these friends were tagged, we used the following methodology: first, we acquired the first-level-friends set of Sarah. Then, for each first-level friend, we downloaded all photos in which he or she was tagged. After a purging process for discarding erroneous pictures, available photos summed up to a total of 7,597. This set was split into two different sets, one for training and one for testing, which contained 3,752 and 3,845 images, respectively. In order to extract faces from the photos, a Viola-Jones HAAR-based face detector [14] was used for detecting frontal faces only. Upon successful detection, a face was only accepted for training or testing if a compatible tag match was found. From the training set of photos, a total of 1,306 classifiers (based on Embedded HMM's [15]) were obtained out of which, for only 840 we could find social information and for the remaining 446 we could not (they did have their friends information private or the tag names did not correspond to a valid Facebook account). The testing set, after the face detection phase, produced a total of 5,258 total faces. Sarah was able to acquire the tags from these photos, each consisting of a pair of the following information:

Name of person tagged
Position of face in photo (X,Y coordinates in percentages of dimensions)

Along with the tag information, our robot, would try to explore the friendship relationships within the picture, even for people who were not her friends. This could be partially accomplished through a special Facebook AJAX call which is essentially equivalent to following the link "View Friends", which appears in the Facebook Search Results. By doing the same procedure iteratively, we were able to build a database of 2,637 people along with their friendship information, which accounted for the number of people that were reachable through the photos that the robot explored, and who had their friends list publicly open. The total size of the data, including the images, the classifiers' output and the social information, was around 640 MB fragmented in 15.920 files. These files were then used for experimenting with the algorithms described in Sect. 18.2.3.

18.2.5.1 A Closer Look at the Data Sets

The given name of the first Facebot that we have built is Sarah Mobileiro, which is also her online name. There exist three kinds of friends of the robot: first, those that have been physically encountered, but are not on facebook, second, those that have been physically encountered, and are also on facebook, and third, those that have not yet been physically encountered, but are facebook friends. It is also worth noting here that there is a highly dynamic nature in figures related to the network – facebook profiles are being added and retracted or become restricted every day; thus, here, we will chose to report approximate numbers, which are based on data gathered during three snapshots, in the last 6 months.

First-Level Friends The robot at this moment has 76 Facebook friends,[3] out of which 14 she has met physically, and has also acquired camera pictures of.[4] The robot also has another 80 friends who are not on Facebook, and also has camera pictures of them. The set of the first-level friends (direct friends) of the robot in Facebook is depicted in Fig. 18.1.

Higher-Level Friends and Mutual Friends Upon moving from the first-level friends to the second-level, i.e. the friends of the first-level friends of the robot who are not first-level friends, there is, as expected, a huge increase: the set FL2 (of friends with minimum distance 2) of Sarah the Facebot contains almost 14,000 members. By a simple division, one gets the figure of on average approximately 175 new second-level friends for each first-level friend. Of course, the average number of second-level friends corresponding to each first-level friend is higher (on the order of 210 as compared to 175, i.e. on average 35 friends are shared, i.e. approximately 15% of the friends are shared). This is due to the existence of mutual second-level friends between any two first-level friends. Also, it is worth noting that the variance

[3] The robot accepts friendships only from a selected circle at the moment.

[4] which we are not including in the experiments reported in this chapter. Interesting results regarding the transferability of training from camera- to facebook-photos and vice versa can be found in [6]. Also, results for hybrid training sets are included there.

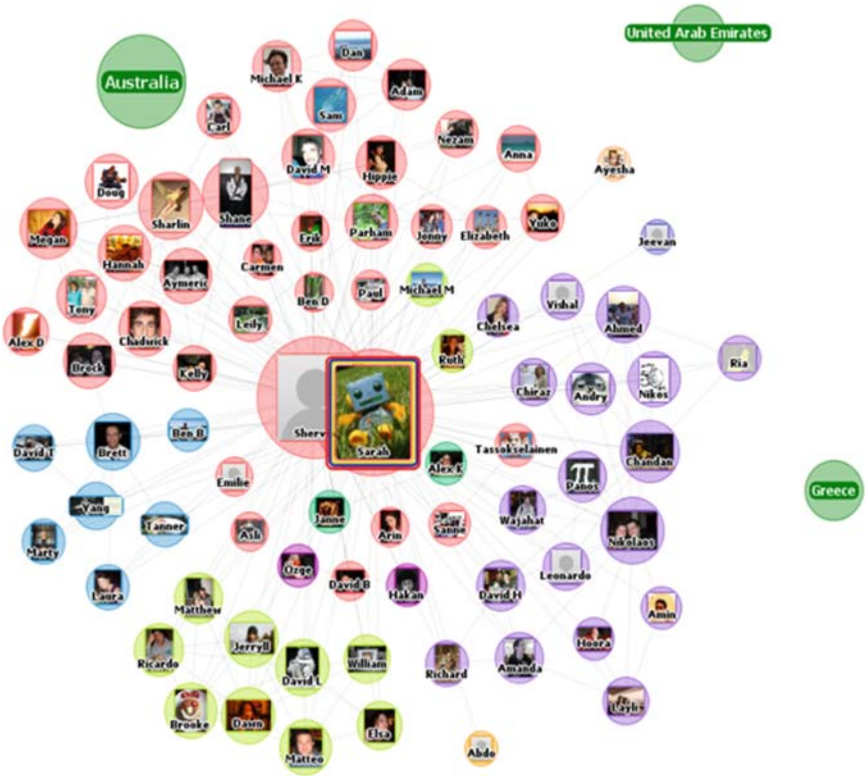

Fig. 18.1 A **Touchgraph** depiction of the 1st-level friends of our robot, 03/09

of the number of friends of each member of FL1 is quite high too almost 120 in this case.

Number of Friends for Which We Can Create Classifiers All the above statistics are related to the social network of Sarah, at maximum distance two. Now, having briefly explored this, let us move on to the next question in sequence: How many of the first and second level friends of Sarah can we create classifiers for, towards FR?

The total number of tagged photos of the members of *FL1* and *FL2* which are directly accessible to Sarah is on the order of 11,000. This number arises as the sum of the number of tagged photos across each first-level friends tagged photos of second-level friends is not generally accessible due to visibility constraints. The distribution of the number of available tagged photos for the 76 first-level friends is in Fig. 18.2.

The average number of tagged photos per first-level friend is approximately 140, with a standard deviation of 180 easily explicable through the 4 outliers with more than 300 tagged photos. Thus, we expect to have a wide variety of training set sizes as numerous friends have only 1 photo available, while a significant number might have 100 or more.

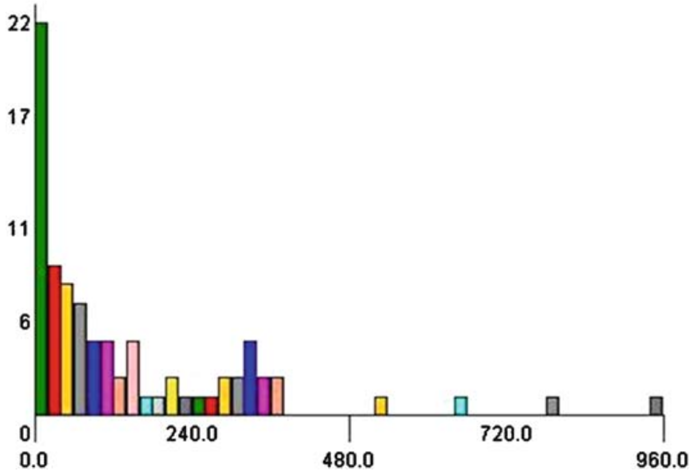

Fig. 18.2 Histogram of number of available tagged photos per first-level friend of the robot (these tagged to be potentially utilized as a training set)

Now, although as we mentioned there is a sum on the order of 11,000 photos when tagged photos are summed across the 76 first-level friends, not all of these are unique. Out of these, given the possibility of a photo having more than one person tagged, the number of unique photos is around 7,650, including 50 or so problematic images, leaving about 7,600 usable. Furthermore some of these will have only a single face tagged and some more than one face. Indeed, more than two thirds out of the 7,600 unique photos have more than one tagged face, as can be seen from the histogram of number of photos containing n tagged faces in Fig. 18.3. And now the question arises: for how many of the first- and second-level friends of the robot do we have adequate training sets to create classifiers out of? If we restrict ourselves to gathering training data through these tagged photos (the simplest and safest solution, as described in [6]), then we have at least one tagged photo for only approximately 3,600 out of the 14,000 or so first- and second-level friends of Sarah, i.e. roughly 25% of the union of first- and second-level friends.

Relationship of Co-occurrence with Friendship Now, let us provide a first quantification of the relationship between face co-occurrence and friendship, as manifested in our data set. Consider the following three questions:

Q1: Given a person A in a photo, what is the probability of any other person B in the photo being a first-level friend of A? Let us call this P1.

Q2: Given a person A in a photo, what is the probability of any other person B in the photo being a second level friend of A? Let us call this P2.

Q3: Given two persons A and B in a photo, what is the probability of any other person C in the photo being a mutual friend of A and B (where it is not necessary that A is a first level friend of B)? Let us call this P3.

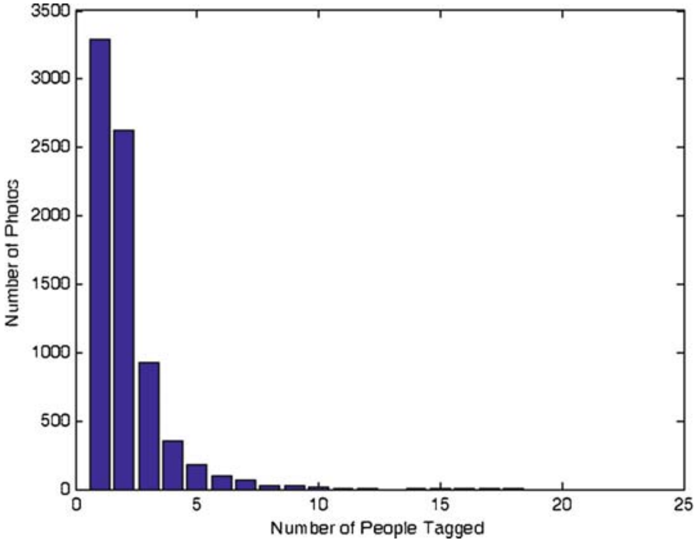

Fig. 18.3 Histogram of number of photos containing exactly *n* tagged faces

By examining all the approximately 5,000 unique tagged photos with more than one tagged face that Sarah has direct access to, we obtain the following estimates for the three above probabilities (measured across tagged photos with $>$ 1 face):

$$\mathrm{P1} = 0.785,\ \mathrm{P2} = 0.024 \text{ and } \mathrm{P3} = 0.278$$

Notice that P1 is strikingly high: almost 80% of any two faces in photos are first-level friends, and this very strong correlation underlies the high effectiveness of the incorporation of social context in our algorithms, which we will be illustrating by quantitative results in the next section. Finally, a very important point not discussed yet deals with the amount of overlap between the identities (people) included in the training set and having formed classifiers, and those tagged in the testing set. As mentioned above, there were approximately 1,300 classifiers and 1,400 unique tags in the testing photos; however, only approximately 400 people had classifiers and appeared in the testing photos, i.e., only roughly a third or so of the people appearing in the testing set we had classifiers for.

Demographics According to Friendship The demographics of the identities (people) are also quite interesting in their own right. As mentioned above, the intersection of training and testing set identities is approximately 400, and the union of the training and testing set identities is straightforward to calculate: $1{,}300 + 1{,}400 - 400 = 2{,}300$ people. These can be divided into five categories: those belonging to the first-level friends of the robot *(F1)*, those belonging to the second-level friends of the robot *(F2)*, those who are on facebook but not first- or second-level friends

(F4), and those who are not on facebook *(F5)*. Rough percentages of these categories within the union and intersection follow:

Union: F1 3%, *F2* 55%, *F3* 28%, *F4* 14%
Intersection: F1 16%, *F2* 69%, *F3* 10%, *F4* 5%

Now, having examined various interesting statistics regarding our data set, we will proceed to presenting results from our algorithms and comments.

18.2.6 Results and Commentary

Here we present results quantifying the performance of the previously described algorithms on our acquired data set. Later, in a separate section, experiments investigating the effect of training set size and consequently providing predictions of scalability are provided.

Initial Comparison of Algorithms The plots of the recognition results for the three algorithms are presented in Figs. 18.4, 18.5 and 18.6, i.e. Alg0 without social info, and Alg1 and Alg2 with social info respectively. In each of these figures, there are two curves: one corresponding to the correct recognition percentage as a function of training set size and the other corresponding to participation in the Rank10 subset of the classifier, again as a function of training set size. It is obvious that the latter curve should always be above the former. Two linear fits are also presented above the curves. Correct recognition is in practice useful for an AR system; while Rank10

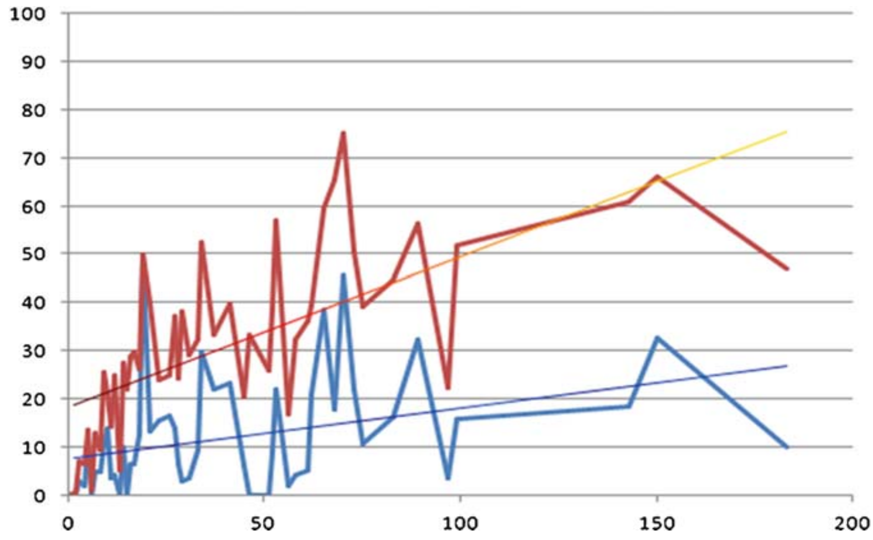

Fig. 18.4 Alg0 (no social info) Rank1 and Rank10 recognition accuracy, as a function of training set size

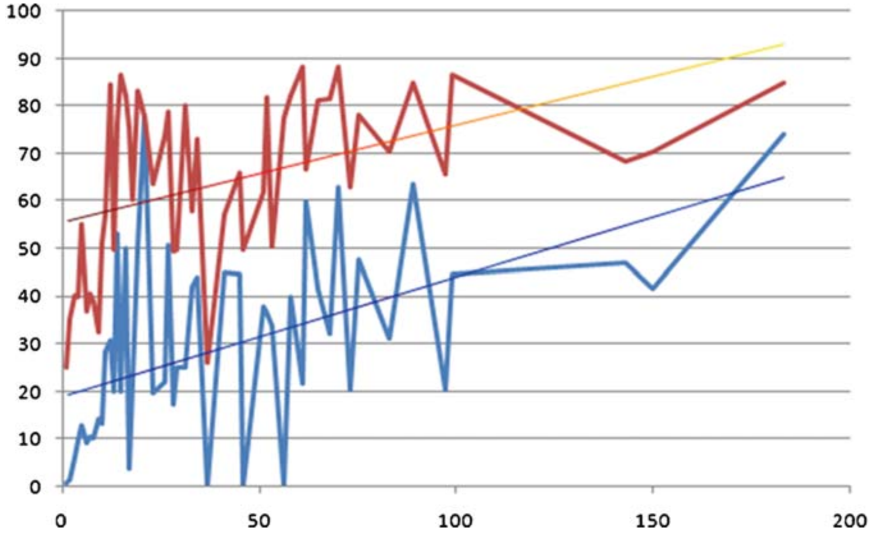

Fig. 18.5 Alg1 (social info, single seed known) Rank1 and Rank10 recognition accuracy, as a function of training set size

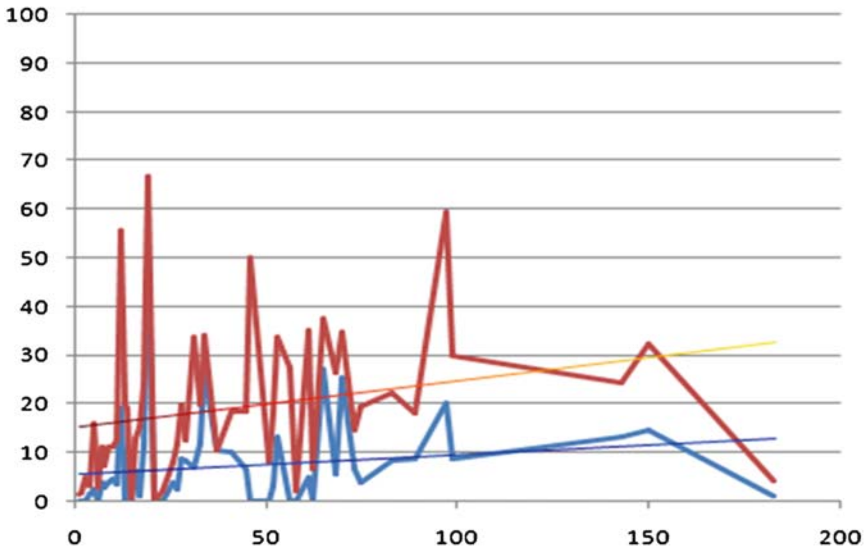

Fig. 18.6 Alg2 (social info, unseeded) Rank1 and Rank10 recognition accuracy, as a function of training set size

participation can be useful for an operator-assisted SAR system, where the Rank10 list is presented to an operator for selection.

The *first conclusion* to be reached by the figures is that clearly there is a *significant increase in recognition performance through the utilization of social information* (e.g., compare Figs. 18.4 and 18.5). In practice, without social info, classifiers made from training sets of size 1–50 or so were totally unusable for both AR as well as SAR; and only remotely helpful in the case of SAR in the case of larger sets (Fig. 18.4). However, with social info, one can start using SAR even with training sets of 10 or so, and can definitely use SAR with bigger sets and AR becomes useful with sets over 50. In quantitative terms, across all training set sizes, the Rank1 percentage is 11.5% with Alg0 and 20.3% with Alg1, while the Rank10 percentage grows from 30% to 52.4% (almost a twofold increase). If we restrict ourselves to only those training sets that have more than four photos, then Rank1 grows from 14.5% to 30%, and Rank10 from 38.5% to 64.4%.

The *second conclusion to be reached* is that although social information can really help, by comparing Fig. 18.6 with Fig. 18.5 or Fig. 18.4, it becomes clear that *a reliable seed is required* for this to take place. Alg2 (a very simple algorithm, multiple extensions of which exist as noted) just picks a face at random, calculates recognition scores for it and chooses the identity that has the highest score as its true identity, and then seeds Alg1 from this. However, if the seed is unreliable, then the social-context-driven boost cannot be so simply utilized. For Alg2, Rank1 and Rank10 percentages are of the order of 4.5% and 12.8% on average; which is even worse than Alg0. Things do not change with larger training sets, too. However, the other variants of Alg2 provide improvements, as we shall see later.

Finally, there is a *third conclusion* which is very important. The total testing time without utilizing social info is on the order of 23 s per face without parallelization. With social info, through the option of hard-restriction of the hypothesis space, this moves down to 4 s, i.e. a *sixfold improvement in recognition time*, quite important in real-time scenarios.

Thus, in conclusion: *social information helped us achieve a twofold increase in Rank1 and Rank10 accuracies, and has turned unusable results into usable ones. However, one should be very careful when seeding; an unreliable seed can revert the above situation, and a more complicated algorithm than Alg3 has to be used if no seed exists. Finally, social info can also enable a sevenfold speedup.*

18.2.6.1 Tuning Alphas and Betas

An investigation of the tuning of the parameters α_1, α_-1, and β, which appear in the soft versions of Algorithm 1, i.e. $Alg1S$ and $Alg1TS$. Initially, a number of values were hand-picked and tried. Then, a non-linear optimization was performed, using the *Nelder-Mead method* [13]. The resulting optimum was found to be at

$$(\alpha_1, \alpha_{-1}, \beta) = (25, 0, 0.3) \tag{18.9}$$

The interpretation of this result is the following: We found that performance was increasing with α_1 increasing; however, no further increase took place after $\alpha_1 = 25$

(no further decrease too). In essence, such a huge value of α_1 (our variance of classifier score output is much lower), practically makes all first-level friend classifier biased scores to be bigger even than the biggest non-first level friend classifier output (i.e., equivalent to hard biasing if more than 10 friends of the seed exist). On the other hand, α_{-1} was found to be optimally at zero; i.e. for the purpose of FR, no distinction needs to be made between a '0' in the friendship matrix (knowing that somebody is not a friend of the seed), and -1 (having no knowledge whether the person is a friend of the seed), hence both categories are equivalent when it comes to their treatment. Finally, it was found that biasing according to training set size could indeed improve results, as we had a non-zero value for the optimum β. The optimum beta for our classifier score statistics was approximately 0.3. For the optimal α, the increase in *Rank*1 accuracy by the optimal β was on the order of 1% (additive over the baseline of roughly 20%) and the increase of *Rank*10 accuracy roughly was double at 2% (additive over the 50% or so, baseline).

Summing Up To achieve optimal recognition for *Alg*1, it is enough to first perform hard biasing and then to just add a training-set-size biasing term to the remaining classifiers, after having optimized for a value of β (for our classifiers this was $\beta = 0.3$). This extra term has a noticeable, however, not significantly large effect on recognition accuracy.

Effect of Random versus Best Seed A quantitative comparison of Alg2RS and Alg2BS was carried out. It was found that the latter, which was chosing as the seed the face that we had most confidence regarding its identity, was superior, giving an increase of 2% or so for *Rank*1, and 5% or so for *Rank*10 accuracy. Still, however, the overall performance of the algorithm was quite prohibitive.

Effect of Overall Match Metric Initial experiments on cycling around all possible seeds (i.e., the *Rank*1 postulated identities of each of the faces in the photos), performing seeded classification, evaluating the overall confidence of the resulting solution and selecting as seed, the one that gave maximum overall confidence took place. For a simple sum square metric, it was found that the results were worse than *Alg*2*BS* (a priori choose best seed) and close to random seed results.

18.2.7 Predicting Scalability

In this section, we attempt to explore the relation of recognition accuracy with the size of the available friendship network. In order to achieve this, our method, which we are going to present formally in what follows, was to randomly create subsets of various sizes, of the robot's friends and then repeat the recognition process based on data derived from this subset.

At this stage, we will start using the extra notation which is introduced in Appendix A.

Using these extra notations, we generate randomly a subset of Sarah's friends of predefined size, denoted by F', in each run of our experiments. This subset is then used to create new photo sets denoted by P', $P'tr$ and $P'te$ for total photos, the training photos and the testing photos, respectively. Finally, from these sets, only a subset of the original DFN can be created, denoted by DFN', and fewer classifiers can be built denoted by C'.

Formally, these sets, satisfy the following relations:

$$F' < F, \text{ random subset of size-}n \tag{18.10}$$

$$P' = \{p \in P | tags(p) \cap F' \neq 0\} \tag{18.11}$$

$$P'tr = \{p \in Ptr | tags(p) \cap F' \neq 0\} \tag{18.12}$$

$$P'te = \{p \in Pte | tags(p) \cap F' \neq 0\} \tag{18.13}$$

$$DFN' = \{t, t \in tags(p) \cap Fr(t) \neq 0, p \in P'tr\} \tag{18.14}$$

$$C' = \{U tags(p) | p \in P'tr\} \tag{18.15}$$

All these equations are straightforward and actually answer to the question of what portion of any data set would be accessible provided that the friends set was actually equal to F'. It is important to note that, every classifier in C' is not trained with the exact same set of photos, with which the same classifier in C was trained with, or mathematically the respective sets $P'tr_i$ and Ptr_i are not identically equal for every I in C'. Remember that the i-set of a set of photos Po, denoted by Po_i is defined by

$$Po_i = \{p \in Po | I \in tags(p)\} \tag{18.16}$$

However, the following lemma holds.

Lemma 18.1.

$$P'tr_i = Ptr_i, \forall I \in C' \cap F' \tag{18.17}$$

A proof of this lemma can be found in Appendix B.

18.2.7.1 The Testbed

For the purposes of the research presented in this section, we devised a testbed in which, random subsets of various sizes were created for the complete set of friends of our robot (denoted as F). This yielded new sets of photos for training and testing and new sets for DFN and for the classifiers (P', $P'tr$, $P'te$, DFN', C'), using the process described in Eqs. 18.10 to 18.15. In specific, the subset size was ranging from 0 to $|F|$, incremented by 2 in each run, taking four samples for each. Actually F' is subset of F, so it is actually a $|F'|$-combination of F, meaning that there are exactly $\frac{N!}{(N-k)!k!}$, if we denote $|F|$ by N and $|F'|$ by k. The sample size we take is small, but from the one hand, a single run of the testbed is computationally expensive (25 s and 50 Mbs for each sample), which puts a significant constraint on how many samples we can actually take, and from the other hand taking almost 160 samples in total is enough to perform a quick analysis of how the accuracy is affected by the total number of people in the social information seed and finally check how our social algorithms scale to the amount of social information at hand.

In the Table in Fig. 18.7 the sizes of the aforementioned sets are being presented for subset size that are multiples of 10. It is easy to see that all metrics quickly converge to a linear relation with the size of the friend s' subset (F'). A linear regression on these values might yield the following results

$$|P'| = 79.|F'| + 451 \tag{18.18}$$

$$|P'tr| = 39.|F'| + 231 \tag{18.19}$$

$$|P'te| = 23.|F'| + 134 \tag{18.20}$$

$$|DNF'| = 18.|F'| + 102 \tag{18.21}$$

$$|C'| = 13.|F'| + 90 \tag{18.22}$$

The above equations do reveal some interesting attributes of the Facebook data set we are analyzing, from which we distinguish the most important to be the relation of the *DFN* and the set P of photos as compared to the friends subset size. All other metrics (testing/training sets and classifiers) are in fact dependent on these sets.

We begin by taking into consideration the case in which a new person joins our friendship network and try to examine its effect on the photos that we can access (P') and the identities we can be aware of (DFN'). First of all, for every new person coming into a social friendship network, new resources are added, such as 79 new photos, 19 new tags (from these photos) and 13 more classifiers can be built for our system. These numbers are referring to mean values (Fig. 18.7) and might vary greatly depending on the social attributes of the new person added. In our data set, the big majority of our photos contained only one face, which seems also to be true for the entire Facebook data set based on our experience. This in part explains the

Number of Friends (F')	Set of photos (P')	Set of training photos ($P'tr$)	Set of testing photos ($P'te$)	Discoverable friendship network (DFN')	New classifiers(C')
10	1241	621	364	282	220
20	1848	919	505	466	355
30	2828	1406	820	706	526
40	3701	1834	1079	869	643
50	4692	2326	1342	1158	861
60	5188	2576	1510	1167	863
70	6133	3034	1801	1412	1049

Fig. 18.7 Effect of increasing the number of first-level friends

relatively large number of photos added for each new person added in the social circle, since these photos most probably contain only the face of this person, so they were previously unknown.

The total number of distinct tags of these new photos, add new identities in the DFN. The total number of these new identities obtained from the tags are mostly affected by two factors:

1. The likelihood of this person being a friend of someone in our initial network, and the transitivity of this network
2. The type of images that the new person has available in his profile

For example, if the new person is already a friend of someone in our initial network and the transitivity of this network is high, the new person will most probably have many of his friends already in it. As a consequence, there is again a high probability that his or most of his photos have already been available from through those friends. In addition, the number of faces in a photo greatly depends on the type of the photo (e.g., personal, friends, events, etc.) which in turn defines how many new identities can be inserted into *DFN*.

From our results, this number is approximately 19, which means that for any new person, 19 more people are discovered whose social information (just the list of 1st level friends) can be accessed. A simple analysis reveals that we get approximately one new identity for almost four new pictures added in the photo set. This ratio is yet another testament on the fact that the personal photos (photos with one face) is the prevailing type of our photo data set.

18.2.7.2 Recognition Performance and Training Set Size

Another important aspect is how recognition performance changes according the size of the friendship network that we use as a basis. For example, currently our robot has around 80 friends; what would be the expected performance of the methods described above if she had 1,000 or 5,000? An initial layman's thought might be that we might not be able to make strong estimations of what this relation might be, since accuracy improvement using social information is largely dependent on the type of photos that were used both in the training and the testing set (e.g., applying social information on a testing set with personal photos only, i.e. photos with one face, would be useless).

Nevertheless, the results of the previous section indicate a strong and stable correlation among the various data sets that are the focus of our research, which can be used to a certain extent for obtaining some insight over the extent of how useful the social information-based algorithms can be, for bigger numbers of first-level friends and larger sets in general. For this, one has to first identify the circumstances under which, applying social algorithms on the testing photos, provide the best results.

A very important observation to be made at this stage is the following: in Alg0 (w/o social ctxt), the larger the set of classifiers we have, the smaller we expect our recognition performance to be, as there are more possibilities for false identification.

This is indeed the case, as can be verified by the lower curves in Fig. 18.8. On the other hand, for Alg1 (with social ctxt), one should notice that in the hard-bias case we are effectively restricting our hypothesis space (number of effective classifiers) to the first-level friends of the seed face (and not to the first-level friends of the robot, and all the classifiers that are created through our process of getting their tagged photos and using other tags too). That is to make things clearer: as the number of first-level friends of our robot increases, so does the number of classifiers that it can create. However, if using Alg1 for recognition, the effective hypothesis space of classifiers does not grow; because it depends on the number of friends of the seed faces, and not of the robot. That is even if the number of friends of the robot increases dramatically, in which case the recognition performance of Alg0 (w/o social ctxt) will fall, we do not expect the same to be the case for Alg1 (w social ctxt), given that its performance depends on the average number of friends of the seed faces, and not on the actual number of friends of the robot. Thus, as the number of friends of the robot increases dramatically, *we expect Alg 1 to sustain its performance*; and this seems to be, at first glance at least, the case in Fig. 18.8.

In conclusion: We saw how various quantities related to our problem vary as a function of the number of first-level friends of our robot. Many of these grow linearly, as we have seen. Recognition performance without social information (Alg0) falls as this number grows, as expected. However, recognition performance with social information (Alg1) is expected to *remain stable* after a while, due to the argument given above: i.e., the effective constraining of the hypothesis space to size equal to the average number of first-level friends of the seed. This, at first sight,

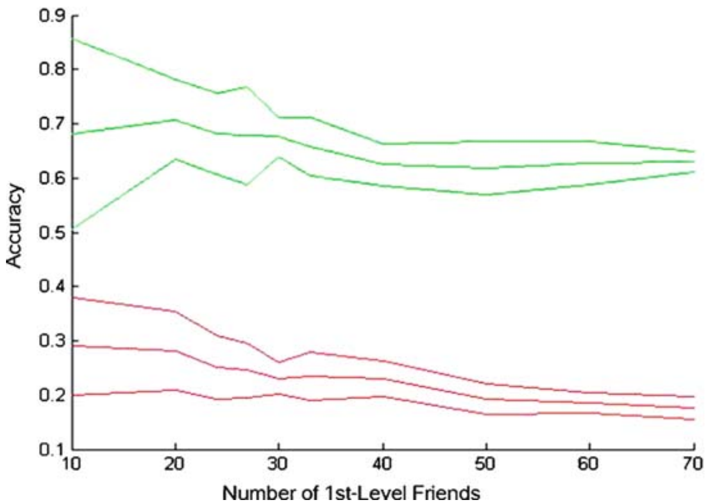

Fig. 18.8 Recognition performance as a function of increasing number of first-level friends of the robot, estimated using the multiple subsets technique described here. *Upper curves*: Alg1 (with social ctxt). *Lower curves*: Alg0 (no social ctxt). The *three curves* in each group correspond to the mean value as well as mean+std and mean−std.

seems to also agree with the results of our empirical investigation, and is a very encouraging result, promising very good scalability of our method.

18.3 Using Co-occurrence in Photos Towards Estimation of The Friendship Network

18.3.1 Introduction and Problem Setting

After having seen how social context (and more specifically knowledge of the friendship relation among individuals) can help enhance FR, through exploitation of the correlation of friendship with face co-occurrence in photos, now we will discuss the inverse problem, i.e., how we can estimate the friendship relations of a number of individuals, by having many photos of them.[5] Again, the key is obviously the correlation of co-occurrence of faces in photos with friendship. More precisely, we define the following problem:

Pr4 Given a number of photos containing tagged faces, estimate the friendship matrix ($\hat{FM}$) of a number of individuals. Metrics for comparing the success of different approaches usually measure the differences between the estimated ($\hat{FM}$) and original friendship matrix (*FM*). One possibility is to employ signal detection theoretic metrics and investigate false positive rates, sensitivities, confusion matrices, and ROC curves. This is the approach being followed here.

18.3.2 Proposed Algorithms

We propose a basic class of algorithms for solving ***Pr4***.

18.3.2.1 Algorithm 4

Input A set of tagged photos
Output A friendship matrix estimate

First, we create a co-occurrence count matrix *CM*, in the following manner, starting from a zero matrix: for each photo in the input set, take all possible pairs of faces (F_i, F_j), including the case for $i = j$, and increase the count in the co-occurrence matrix in $CM(i, j)$. At the end of the process, the resulting symmetric matrix will have the number of occurrences of an identity's face in the diagonal of the matrix,

[5] To our knowledge, Hiroshi Ishiguro first mentioned a similar problem for the case of a robot observing people [12].

and the number of co-occurrences of two identities i *and* j in the off-diagonal element $CM(i, j)$. Now, the co-occurrence matrix has to be converted to an estimate of the friendship matrix. The following two approaches are proposed, and results will be presented below.

Alg4T Here, simple thresholding is employed. The friendship matrix is to be filled with $\{1,\ 0\ or\ -1\}$, corresponding to the three cases (are friends, are not friends, do not know). If our ultimate purpose is to later reuse the matrix for social-context assisted FR, then as we saw above in the results section for the algorithm2, we do not need to differentiate between the cases of ($-1 = donot\,know$) and ($0 = not\,friends$), because for the case of optimal recognition results the weight a_{-1} is zero. Thus, we move across the diagonal of the co-occurrence matrix CM and if the diagonal element is zero, we fill row i and column j of $\hat{FM}$ with -1. If it is non-zero, then we copy the corresponding row and column from CM to $\hat{FM}$ after transferring through the following rule:

R1

$$\hat{FM}(i,j) = \begin{cases} -1 & \text{if } \dfrac{CM(i,j)}{CM(i,i){\cdot}CM(j,j)} \le Th1 \\ 1 & \text{if } \dfrac{CM(i,j)}{CM(i,i){\cdot}CM(j,j)} > Th1 \end{cases} \tag{18.23}$$

If we are interested in also creating a friendship matrix containing not only 1 and -1 but zeros, then a variation of the above rule could be:

R2

$$\hat{FM}(i,j) = \begin{cases} -1 \text{ if } CM(i,i)\cdot CM(j,j) < Th2^{6} \\ 0 \text{ if } \left\{\dfrac{CM(i,j)}{CM(i,i)\cdot CM(j,j)} \le Th1\right\} \wedge \{CM(i,i)\cdot CM(j,j) > Th2\} \\ 1 \text{ if } \left\{\dfrac{CM(i,j)}{CM(i,i)\cdot CM(j,j)} > Th1\right\} \wedge \{CM(i,i)\cdot CM(j,j) > Th2\} \end{cases} \tag{18.24}$$

The appropriate thresholds are chosen through signal detection theory, given a criterion for choosing the operating point.

Alg4TE Here we employ transitive extensions (TE) instead of thresholding. The underlying idea that when the set of photo observations is small, if *Id1* and *Id2* appear in a photo, and *Id2* and *Id3* appear in another, then chances are they are all first-level friends with each other. Thus, we quantize the co-occurrence matrix *CM* to contain only 0 (no co-occurrence) and 1 (for non-zero co-occurrence count). Then, we multiply *CM* with itself *m* times, and quantize again. The appropriate value of *m* is again determined through a signal detection theoretic criterion. We

[6] Not enough observations for adequate knowledge.

expect to reach transitive closure (TC) after a specific m, and after this no further changes arise if we further multiply and quantize.

18.3.3 Results

Results of the two cases of *R1* (with and without TEs) will be described here. In our experiments, the data set as described in Sect. 18.2.5.1 was used and we compared our resulting estimated friendship matrix $\hat{FM}$ against the original friendship matrix *FM* described in Sect. 18.2.1. In more detail, we were able to reach a total 2,637 people (following the tags accompanying the facebook photos) through the friends of our robot Sarah's facebook profile. Hence the dimensions of *CM*, *FM* and $\hat{FM}$ matrices are the same (i.e., 2637×2637). As already discussed in Sect. 18.3.1, in R1, we do not differentiate between relation levels -1 and *0*, the resultant confusion matrix reduces to Fig. 18.9. Receiver Operating Characteristics (ROCs: True Positive Rate versus False Positive Rate) are presented in Figs. 18.10 and 18.11.

When *FM* is constructed using R1 without transitive extensions, the ROC of Fig.18.10 is obtained.

Using R1 with TEs, TPR (True Positive Rate) and FPR (False Positive Rate) stabilize after nine iterations ($m = 9$). By looking at the graph in Fig. 18.11, one can observe that with TEs, the TPR almost saturates at 23%, meanwhile the FPR gradually grows. On the other hand, in the case without TEs, the TPR and FPR remain almost constant at 11% and 0% (Fig. 18.10 which are both lower than the former. Therefore, it is quite obvious that using simple TEs increases the TPR rate by approximately double meanwhile the corresponding increase in FPR remains insignificant. Thus, we have seen how with computationally very less expensive methods that are easy to implement one can easily recover a significant amount of the friendship network from photos; and one can chose among many possible operating points depending on the tolerance of different FP and FN.

	−1	0	1
−1	TN	TN	FP
0	TN	TN	FP
1	FN	FN	TP

Fig. 18.9 Confusion Matrix for R1

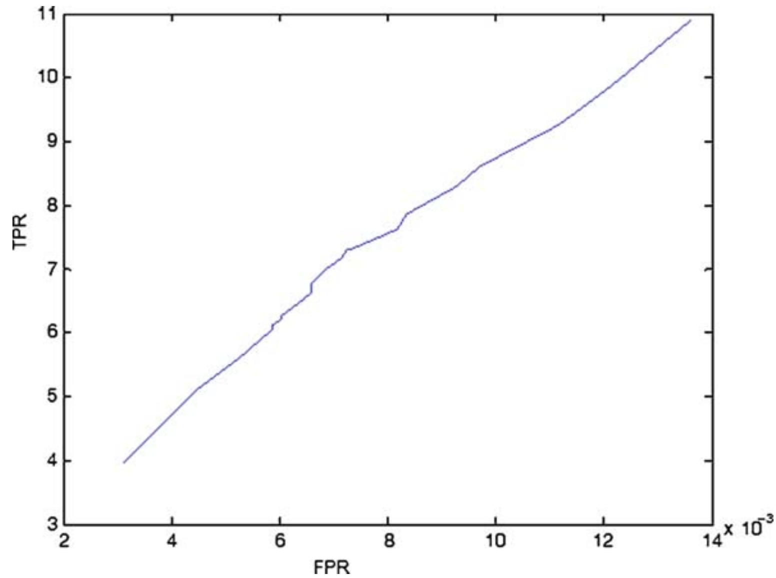

Fig. 18.10 ROC: TPR × 100 (Y) vs FPR × 100 (X) for 50 iterations (Th1=0:0.01:0.5), no TE

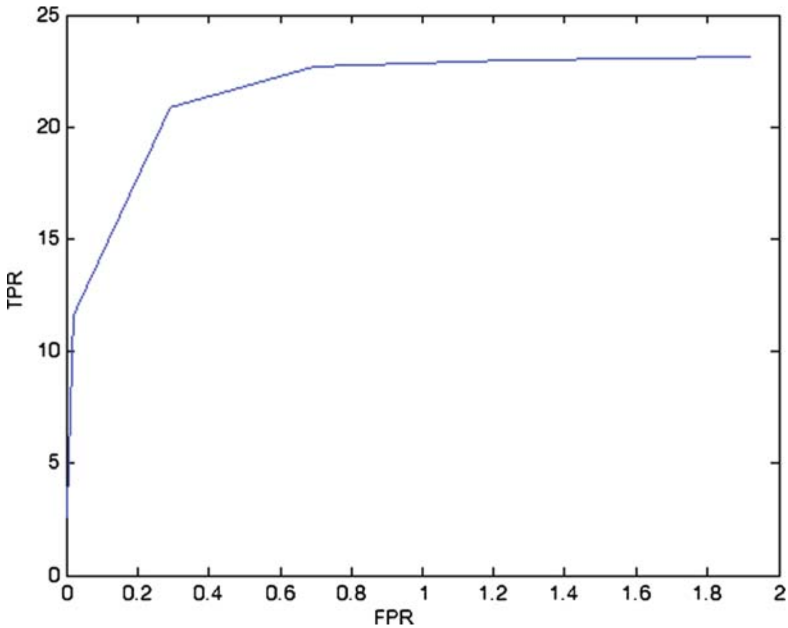

Fig. 18.11 ROC: TPR × 100 (Y) vs FPR × 100 (X) for 50 iterations (Th1=0:0.01:0.5), TC at $m = 9$

18.4 Conclusion

In this chapter, we have discussed proposed algorithms and demonstrated through real-world results *how knowledge of the "friendship" social relation can help create faster and better face recognition, and how sets of photos with recognized faces can help estimate the friendship relationships existing among individuals*. This was possible, because the two relations of friendship (among individuals) and co-occurrence (of faces in photos), though appearing in different realms (social vs. visual sensory), have a strong correlation: faces that co-occur in photos often belong to individuals that are friends. Using real-world data gathered from "facebook", which were gathered as part of the "FaceBots" project, the world's first physical face-recognizing and conversing robot that can utilize and publish information on "Facebook", we presented novel methods as well as results for utilizing this correlation in both directions. The results were quite encouraging: in our primary example, we were able to demonstrate doubling of the recognition accuracy as well as a sixfold improvement in speed. Various improvements, interesting statistics, as well as an empirical investigation leading to predictions of scalability to much bigger data sets were also discussed. Finally, we hope that apart from the specifics presented here, this chapter has also acted as a concrete demonstration that very promising avenues exist at the crossroads between social networks and numerous other areas, many of which are open to future exploration.

Appendix A

Formal Notation Here we introduce a notational system in order to formalize our basic concepts and enable a succinct and precise exposition of the rationale behind using social algorithms. This system should be considered complimentary to the notation introduced in Sect. 18.2.1 which was more a verbose explanation of the algorithms. In order to be able to enumerate identities, we are assigning to an identity, a unique positive integer (not necessarily the same with the facebook identity, for those identities that are on facebook). If the maximum assigned *id* is denoted by M, we can define a total set of our assigned *ids* as:

$$D = \{1 \ldots M\} \tag{18.25}$$

Next, we denote the set of all tagged facebook photos as P_b.

$$P_b = \{\text{Total set of facebook photos}\} \tag{18.26}$$

We assume that there is a function which maps a photo to the set of id's of its corresponding tags:

$$tags() = P_b \overset{maps\ to}{\rightarrow} D^k \tag{18.27}$$

We also assume that there exists a procedure *Fr*:

$$Fr(x) : D^k \overset{maps\ to}{\rightarrow} D^k = \{\text{ids of friends of } x\} \tag{18.28}$$

For convenience, we assume $x \in Fr(x)$. Based on Fr we can define:

$$FR = \text{Friendship Relationship} = D \times D \overset{maps\ to}{\rightarrow} \{-1, 0, 1\} \tag{18.29}$$

for which it holds

$$FR(i, j) = \begin{cases} 1, & \text{iff } i \in Fr(j) \text{ or } j \in Fr(i) \\ 0, & \text{iff } i \notin Fr(j) \text{ and } Fr(j) \neq 0 \text{ or } j \notin Fr(i) \text{ and } Fr(i) \neq 0 \\ -1, & \text{otherwise} \end{cases} \tag{18.30}$$

Notice that $FR(i, j) = -1$, only when we do not have social information for neither i nor j. Also notice that $FR(i, j) = FR(j, i)$, i.e., FR is a symmetric relation and also non-transitive. The FR relation is the same as presented in Sect. 18.2.1. We further define notation to account for our Robot's Id, for the first-level friends of an Id, and for the available tagged photos containing a set of ids:

$$s = \text{Our Robot's Id} \tag{18.31}$$

$$F = Fr(s) \tag{18.32}$$

$$P = \{p \in P_b | tags(p) \cap F \neq 0\} \tag{18.33}$$

where the member p of P is the same as Ph in Sect. 18.2.1. Equation 18.32 provides a set of our robot's friends. Equation 18.33 is the set of all photos that are accessible by our robot through its friend network (F). By dichotomizing this, we are able to build two sets of photos, for training and testing Ptr, Pte, respectively. A classifier (denoted as Cl in Sect. 18.2.1 is an HMM (Hidden Markov Model) of the facial characteristics of an identity. Therefore, the set of classifiers can be mapped on the set of ids (D)

$$C = \{\text{set of classifier ids derived from the training set } \forall p \in Ptr\} \tag{18.34}$$

We then proceed by defining the discoverable friendship network DFN as the total set of tagged ids on every photo of the training set:

$$DFN = \{t | \forall p \in Ptr, t \in tags(p) \text{ and } Fr(t) \neq 0\} \tag{18.35}$$

We then create a new set which will be the basis of our social information:

$$D_e = \text{extended id set} = \{C \cup DFN\} \tag{18.36}$$

Finally, we then define the friendship matrix (already introduced in Sect. 18.2.1) which holds information for friendship relationship among all accessible ids (as available in D_e):

$$FM = \text{Friendship matrix} = D_e \times D_e \to \{-1, 0, 1\} \tag{18.37}$$

for which

$$FM(i, j) = FR(i, j),\ \forall (i, j) \in Fe \tag{18.38}$$

the tuple (F, P, DFN, FM, C) defines completely an instance of the problem we are investigating. This completes the set of notational machinery required.

Appendix B

Lemma 18.2. *The training sets of the classifiers in C' (the new set of classifiers produced by starting from the reduced friends set F'), and specifically those classifiers in C' with Id's belonging to F', will be identical with the training set of the original classifiers in the full problem (i.e., those arising for $F' = F$, the full friend set).*

$$P'tr_i = Ptr_i, \forall I \in C' \cap F' \tag{18.39}$$

Proof. First notice that:

$$P'tr_i \leq Ptr_i \tag{18.40}$$

This is true since for any p in $P'tr_i$, it holds that:

$$I \in tags(p) \forall p \in P'tr \Rightarrow I \in tags(p) \forall p \in Ptr \Rightarrow p \in P'tr_i \tag{18.41}$$

Now assume that $P'tr_i < Ptr_i$, then there exists a photo p such that

$$p \in Ptr_i \tag{18.42}$$

$$p \notin P'tr_i \tag{18.43}$$

From (18.43) we get that

$$p \notin P'tr \text{ since I} \in tags(p) \tag{18.44}$$

From (18.42) we get that

$$p \in Ptr \tag{18.45}$$

Therefore from the definition of $P'tr$, we get that

$$tags(p) \cap F' = 0 \tag{18.46}$$

However, it holds that

$$I \in tags(p) \tag{18.47}$$

$$I \in F' \cap C' \tag{18.48}$$

therefore:

$$tags(p) \cap F' \geq \{I\} \neq 0 \tag{18.49}$$

The initial assumption has led us to a contradiction, and the training sets for the classifiers that also belong to the new F' friend subset are identically the same. In the special case, in which, $|tags(p)| = 1$, for every $p \in P$, (photos with one face), which is also the vast majority of our own photos, it is easy to prove the equality of the training sets for any classifier in the entire set C'. In brief, (18.46) in the above Lemma $\Rightarrow I \notin F'$, while $C' \cap F' \leq F', \Rightarrow I \in F'$, which is also a contradiction. Therefore, the entire set of classifiers C' will be trained with the same set of training images.

References

1. Mavridis N, Datta C, Emami S, BenAbdelkader C, Tanoto A, Rabie T (2009) FaceBots: social robots utilizing facebook. In: Proceedings of the 4th ACM/IEEE HRI '09. ACM, New York, NY
2. Michelson J, Ortiz J (2006) Auto-tagging the Facebook. http://www.stanford.edu/class/cs229/proj2006/MichelsonOrtiz AutoTaggingTheFacebook.pdf
3. Torralba A (2003) Contextual priming for object detection. Int J Comput Vision 53(2)(Jul 2003): 169–191
4. Hoiem D, Efros AA, Hebert M (2006) Putting objects in perspective. In: Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition, Vol 2 (17–22 June 2006). CVPR. IEEE Computer Society, Washington, DC, pp 2137–2144
5. Stone Z, Zickler T, Darrell T (2008) Autotagging facebook: social network context improves photo annotation. IEEE computer society conference on computer vision and pattern recognition workshops, Anchorage, Alaska, USA, June 2008
6. Mavridis N, Kazmi W, Toulis P, Ben-AbdelKader C (2009) On the synergies between online social networking. Face recognition, and interactive robotics, CaSoN 2009
7. Facebook online social networking website. http://www.facebook.com
8. Facebook statistics. http://www.facebook.com/press/info.php?statistics
9. Samar V (1999) A single sign-on using cookies for Web applications. In: Proceedings of the 8th workshop on enabling technologies on infrastructure for collaborative enterprises table of contents, pp 158–163, ISBN:0-7695-0365-9
10. Nelson ML, Maly K (2001) Buckets: Smart objects for digital libraries. Commun. ACM 44, 5 (May 2001) 60–62
11. Facebook API. http://wiki.developers.facebook.com/index.php/Mainpage
12. Kanda T, Ishiguro H (2006) An approach for a social robot to understand human relationships: friendship estimation through interaction with robots. Interact Stud 7:369
13. Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
14. Viola P, Jones M (2004) Robust real-time object detection. Int J Comput Vision 57(2):137–154
15. Nefian AV, Hayes MH (1998) Hidden markov models for face detection and recognition, in International Conference on Image Processing, Chicago, Illinois, USA