

# A Tool for Retrieving Meaningful Privacy Information from Social Networks

Ricard L. FOGUÉS<sup>1</sup>, Jose M. SUCH<sup>1</sup>, Agustin ESPINOSA<sup>1</sup>, and  
Ana GARCIA-FORNES<sup>1</sup>

<sup>1</sup>*Departament de Sistemes Informàtics i Computació, Universitat Politècnica de  
València, {rilopez, jsuch, aespinos, agarcia}@dsic.upv.es*

## Abstract

The use of social networking services (SNSs) such as Facebook, Flickr, or MySpace has grown explosively in the last few years. People see these SNSs as useful tools to find friends and interact with them. SNSs allow their users to share photos, videos, and express their thoughts and feelings. Even though users enjoy the capabilities that these SNSs offer, they have become aware of privacy issues. The public image of a subject can be affected by photos or comments posted on a social network. Therefore it is important for SNS users to control what others can see in their profile. Recent studies demonstrate that users are demanding better mechanisms to protect their privacy. An appropriate approximation to solve this problem is a tool that automatically suggests a privacy policy for any item shared on a SNS. The first step for any mechanism to recommend and predict privacy policies is to retrieve meaningful privacy information from the SNS, such user communities and the relationships of them. Most SNSs rely on groups to help users specify their privacy policies. Therefore, a basic functionality of such a mechanism is to group the user's friends automatically. Although SNSs treat all of the friends of a user the same, without taking into account different degrees of the friendship, this is not a realistic approach. Hence, another factor to consider when defining a privacy policy is the type of relationship between the owner of the item being shared and its potential viewers. In this work, we present a tool called Best Friend Forever (BFF) that automatically classifies the friends of a user in communities and assigns a value to the strength of the relationship ties to each one. We also explain the characteristics of BFF and show the results of an experimental evaluation.

**Keywords:** Information retrieval, social network, social media, privacy, tie strength.

## 1 Introduction

Social networking services (SNSs) are currently the services that are most more demanded by users worldwide. Facebook (with more than 800 million active users[1]) and Flickr (with 51 million registered members[2]) are two of the most successful

SNSs. People register to these SNSs and share images, videos, and thoughts because they perceive a great payoff in terms of friendship, jobs, and other opportunities [6]. However, the huge number of items uploaded to these SNSs and the persistence of these items in the social networks have the potential to threaten the privacy of their users[13]. For example, employers are becoming accustomed to checking the profile of the candidates in popular SNSs. If the privacy of the profile of a candidate is not properly set, what an employer sees in that candidate's profile may affect the employer's decision. It might even be possible for a stalker to infer the address of a person by looking at that person's photos posted in a social network.

Recent studies show that SNS users' awareness of privacy issues has increased lately [3]. To cope with these privacy threats users tend to adjust and modify the default privacy settings set up by the SNS since they feel that these default settings are not enough to protect them. Nonetheless, the current privacy setting mechanisms offered by the SNS seem difficult or confusing for users [16, 22]. For example, Facebook offers five privacy levels for each element shared: "public", "friends", "only me", "personalized", or "groups". While these five levels may seem sufficient, they require a certain amount of work by the user before they can be applied. In other words, groups have to be built in advance by the user. If we consider that the average number of friends in Facebook is 130 [1], classifying all of them into groups can represent a serious challenge. Furthermore, once groups are defined, if the user decides to exclude specific users of a group from seeing the shared content, he/she has to specify the excluded persons one by one.

Another problem users find when defining privacy policies is that most of the SNSs base their privacy models entirely on groups. Every friend of a user is the same; close friends, family or mere acquaintances are not distinguished. As Wiese et al demonstrated in [24] the willingness of users to share in social networks is dependent on the closeness (tie strength) of relationships. The works of Gilbert et al. [10, 9] and Xiang et al.[25] showed that it is possible to predict the strength of the relationship ties with the information available at the SNSs. As these related works prove, in order to suggest good privacy policies for SNS users the strength of the relationship ties must be taken into account.

These complications and obstacles may lead users to have privacy policies that do not fit their preferences. Another effect of not using properly adapted privacy policies is that the users feel as though they have lost control of their information and how it is shared among the SNS. Users both desire and need more tools to allow them to regain control over their privacy. Thus, in the long term, our aim is to develop autonomous agents that would help users define their privacy policies by automatically recommending them privacy policies that are appropriate.

A first step in creating appropriate privacy policies is to gather information about how the user interacts with others in the social network. In other words, we need to know more details about the social network connections of the user. In this work, we introduce Best Friend Forever (BFF), which is a tool that automatically obtains friend groups and a value for the strength the relationship ties. Moreover, it provides support so that the user can refine the results. The tool has been implemented as a Facebook application and is publicly available at [gti-ia.dsic.upv.es/bff](http://gti-ia.dsic.upv.es/bff). The main objective of BFF is to offer users enough information to modify and adapt their privacy policies on Facebook. Our experimental results

suggest that our software accurately predicts user groups and tie strength values, and also requires little user intervention.

The rest of the paper is organized as follows. Section 2 introduces preliminary notions of tie strength and community finding algorithms in social networks. Section 3 presents an overview of BFF and its different elements. Section 4 reports the results of the experimental evaluation. Section 5 discusses some related works. Finally, Section 6 concludes the paper and outlines future research directions.

## 2 Background

One of the main features of our tool is that it is able to predict a value for the tie strength of each of the social connections of the user. Current SNSs have made little effort to differentiate users. Users are either friends or strangers, with nothing in between. This approach does not properly represent human relationships. As introduced in the paper of Granovetter[12], the concept of *tie strength* defines the relationship between two individuals. In his work, Granovetter describes two different types of ties: *strong* and *weak*. On the one hand, strong ties usually include relationships such as family and close friends. On the other hand, weak ties may refer to coworkers or less trusted friends. Granovetter defined four tie strength dimensions: duration, intimacy, intensity and reciprocal services. Later works proposed three additional dimensions[4, 23, 15]. These three additional dimensions are: (i) structural, which refers to factors like social circles, (ii) social distance, which refer to factors like political affiliation or education level, and (iii) emotional support, which embodies elements such as offering advice.

In Wiese’s work [24], the authors find a high correlation between the willingness to share information and tie strength. Their research proved that the strength of a tie is even more significant than grouping for predicting sharing. They suggest that a mixture of grouping and tie strength might provide richer sharing policies. Their conclusions support our thesis that tie strength is an important variable when deciding who is able to access a given information. Therefore, it is necessary to know the tie strength in order to suggest adequate privacy policies.

Automatic friend grouping is the other main feature of our software. Many SNSs offer grouping features. However, they are not automatic and the users need to take an effort and group all their contacts. If we consider that the average number of contacts in Facebook is 130, this can represent a time-consuming task. Friend groups become useful when defining privacy policies. It is easier for a human user to assign the same access privileges to a group of friends than specifying a privacy policy separately for each persons in the group.

In order to group persons we use *communities* [11]. Communities are usually defined as natural divisions of network nodes into densely connected subgroups. In our context, the nodes are the contacts or friends of a given participant, and the connections between the nodes are friend relationships. There are many community finding algorithms [8]. In this work we use the hierarchical diffusion algorithm proposed by Shen et al. in [20]. Section 3.2 introduces some details of this algorithm.

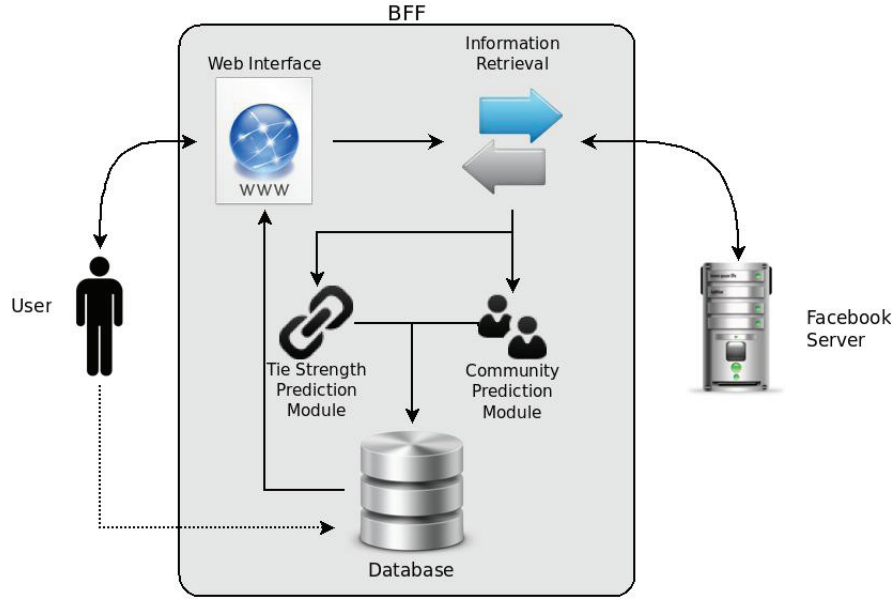


Figure 1: BFF Overview

### 3 Best Friend Forever







This section introduces our tool and gives a complete overview of it. BFF aims to retrieve information from the social network of a participant in order to help to automatically recommend privacy policies. Specifically, the data needed is tie strength and friend groups. BFF is written in PHP and Javascript and is publicly accessible. Due to our experimentation needs, BFF is currently working as a web page; however, in the future, we plan to distribute BFF as a software program that users can execute in their own computers or on a trusted web server in order to preserve their privacy.

BFF is composed of two modules: (i) community prediction, and (ii) tie strength prediction. The community prediction module is in charge of create chunks of users from the participant's contacts. The tie strength prediction module establishes a value of tie strength to each one of the participant's friends. In general, the input of BFF is the Facebook account of the participant, and the output is a set of user groups and a value of tie strength for each one of those users.

Figure 1 shows an overview of BFF and how it works. The interface between BFF and the user is a web page. BFF collects information from the user's Facebook account. Therefore, before users can use BFF, they have to login to Facebook and give permission to BFF to access their Facebook information. Once the permission is given, BFF requests information from the Facebook server. When all the necessary information has been collected, the information is passed to the community prediction module and to the tie strength prediction module. These modules predict a set of groups and tie strength values for the friends of the user. The predictions are shown to the user as a suggestion. The dotted line represents the

**Community5**

Community's new name

| Remove  | Name  | Tie Strength | New Tie Strength |
|---|---|--------------|------------------|
|  | Salvatore  | 3            | -- ▾             |
|  | Eddy       | 1            | -- ▾             |
|  | Vicente    | 1            | -- ▾             |

Add a friend to this community: -- ▾


 Apply Changes

Figure 2: Result Sample

possibility of the user to modify and adapt the suggestions created by the two modules. These modifications are stored in the database for future reference.

BFF aims to be a tool that can be used by real users of Facebook. Therefore, it has to work as fast as possible. BFF allows the users to configure the amount of time they want the whole process to take. The user can choose among three options *fast*, *normal*, and *thorough*. The amount of information collected depends on the configuration chosen by the user. The fast configuration only collects the user's most recent information on Facebook, normal configuration collects some old information, and thorough configuration collects every available information. Clearly, a faster process will be less accurate and more inclined to prediction errors than a thorough process where more information is collected.

Figure 2 shows the screen where the results of BFF are presented to the user. In this example, the figure only depicts one of the communities automatically created by BFF. Part of the name of the members of the community has been hidden to preserve their privacy. As shown in Figure 2, the members of the community are sorted by their tie strength value. The participant can change the name of the community, remove members from the community, add new ones, and change the tie strength value for any member in that community.

### 3.1 Tie Strength Prediction Module

As stated in the introduction, BFF predicts the tie strength of the relationships of the participant with each person that is socially connected to her. We model tie strength as a linear combination of predictive variables. During the creation of BFF the usability of our tool was a key factor. With this in mind, BFF has to be

capable of predicting the tie strength accurately in a reasonable amount of time, and every user should be able to get an accurate prediction. These two requisites (time cost and generalization) conditioned the selection of predictive variables. Next paragraph explains the selected predictive variables.

The variable *Days since last communication* measures the recency of the communication. *Days since first communication* is an approximation of the duration of the friendship. *Wall messages* counts the number of messages exchanged using the wall. *Photos together* counts the photos where both persons (participant and friend) are tagged. *Links shared* counts the number web page links traded between the friend and the participant. *Initiated wall posts* counts the number of publications posted by the friend on the participant's wall. "*Likes*" counts the number of likes given by the friend to the participant's publications. *Inbox messages exchanged* counts the number of private messages traded between both persons. *Number of friends* is the total number of friends of the friend. *Educational difference* measures the difference in a numeric scale: none = 0, high school = 1, university = 2, PhD = 4. Finally, *the mean strength of mutual friends* is also taken into account, and it captures the idea of how a relationship is modified by the tie strength of mutual friends.

The selected predictive variables are based on the variables proposed in [10]. In their work, the authors propose a set of 72 predictive variables. The authors did not consider the cost of collecting the variables and their generalization, they only considered the predictive capabilities of the variables. As stated before, two requisites for the predictive variables are their collecting cost and their generalization. With regard to time cost, BFF collects the information from Facebook; each time BFF needs to ask Facebook for an item, it has to send an HTTP request to Facebook. This operation may take a few seconds; therefore, collecting many variables from a very active participant account can take a long time. BFF had to restrict the number of predictive variables. In the matter of generalization, BFF tie strength prediction cannot depend on variables that require the participant to have specific characteristics. For example, language dependent variables are inappropriate as they would limit the different users that would be able to use BFF. The ten selected predictive variables for BFF satisfy both requisites, they can be collected fast and are valid for any user. Moreover, the selected predictive variables cover every tie strength dimension. Table 1 shows the tie strength dimensions and the predictive variables that belong to each dimension.

The equation below represents the tie strength  $s_i$  of the  $i^{th}$  friend.  $R_i$  stands for the vector of ten predictive variables of the  $i^{th}$  friend.  $\mu_M$  is the mean strength of mutual friends between the user and the  $i^{th}$  friend. Finally,  $\beta$  is the vector of weights applied to the predictive variables and  $\gamma$  is the weight applied to the mean strength of mutual friends. In order to set the weight of each variable we used the findings of [10] as we wanted to avoid the use of a model that completely lacked information on the relative importance of each variable to predict tie strength. As a future work, we plan to perform a fine tuning of the weights of the variables.

$$s_i = \beta R_i + \gamma \mu_M$$

$$M = \{s_j : j \text{ and } i \text{ are mutual friends}\}$$

| Dimension            | Variables   |
|----------------------|---|
| Intimacy             | Number of days since first communication.<br>Number of friends. |
| Intensity            | Wall messages. Initiated wall post. Inbox messages exchanged.   |
| Duration             | Days since first communication.                                 |
| Social distance      | Educational difference.   |
| Reciprocal services  | Links shared  |
| Emotional support    | Likes   |
| Structural dimension | Mean strength of mutual friends.                                |

Table 1: Predictive variables and tie strength dimensions

After collecting the predictive variables for the friends of the user, the variables are normalized. Then, the tie strength is calculated for each user. The results are normalized to a numeric scale 1-5, where 1 represents that both persons are very distant (mere acquaintance) and 5 that they are very close. The results are presented graphically, as shown in Figure 2, so that users are sorted by group and by tie strength. It is easier to figure out the value of the tie strength of a person by comparing that relationship to others. As in the grouping step (explained below), the participant can refine the results of the tie strength calculation.

### 3.2 Community Prediction Module

The community prediction module is based in the hierarchical diffusion algorithm proposed by Shen et al. in [20]. The algorithm is founded on the triadic closure principle, which suggests that, in a social network, there is an increased likelihood that two people will become friends if they have friends in common. The algorithm is divided into two steps: (i) the thresholding step, and (ii) the diffusion step. In the first step, the core members of the communities are chosen. These members are those users that are highly connected to others. Once the core members have been selected, the diffusion step performs a cascade joining, and new members are added to the communities formed by the core members. The diffusion step follows a direct-benefit model in networks, which states that people benefit from directly copying others' decisions. In their paper, the authors did not test their algorithm on a network of a SNS like Facebook. However, according to the results of our experimental evaluation, it performs very well in this environment. Moreover, this algorithm has great performance in terms of computational cost for the average size of Facebook communities.

When the participant uses our software, the community prediction module queries Facebook about the friends of the participant and the friends of those friends (mutual friends) in order to build the graph that will be the input of the algorithm. The community prediction module suggests the community division calculated by the algorithm. The participant can accept the groups proposed or modify them at will.



## 4 Experimental Evaluation

The goal of our experimental study is to evaluate the accuracy of our BFF tool in terms of community and tie strength prediction. Specifically, we want to answer the following questions:

- How effective is the community module in grouping the contacts of a user?
- How accurate are the predictions of the tie strength module?
- Do users perceive that BFF is a good tool in general? In other words, do they think that BFF is capable of inferring accurate information from their available data on Facebook?

To answer these questions, we performed an experimental evaluation with Facebook users. Our results indicate that BFF is an effective tool. Furthermore, users considered BFF to be a good tool and valued it positively. In what follows, we first introduce the experimental settings and then report our findings.

### 4.1 Participants

Our 17 participants were mostly students and members of the Polytechnic University of Valencia. The sample consisted of 4 women (23.5%) and 13 men (76.5%). The minimum number of Facebook friends was 58; the maximum was 529 (mean of 186.94). In total, we analyzed 3178 friend relationships. All of the participants used Facebook regularly.

### 4.2 Method

The participants in our experiment had to try BFF and evaluate its performance. BFF was created to ensure that its use would be easy for anyone. The participants only had to access to the web page of BFF, log in with their Facebook account, and start the application. During the experimental evaluation, the time configuration was deactivated since we wanted all of the participants to evaluate BFF with the same configuration settings. The forced configuration was “normal”, which on average takes 10 minutes to complete for a user with a number of friends of around 100.

After BFF completed its process, the participants were requested to correct any possible errors in tie strength prediction and in user grouping. Users could change the tie strength value of any contact, move users freely from one community to another, and create new communities. These possible corrections were stored in order to evaluate the performance of BFF.

Finally, the participants were requested to answer a short survey to find out their opinion about BFF. The survey was composed of the four following questions:

1. How well did BFF group your friends into communities?
2. How well did BFF predict the tie strength between you and your friends?
3. In general, how accurate do you think BFF is?



4. How accurate do you think BFF is considering it only accesses your information on Facebook? For example, if one of your friends on Facebook is your brother, but you have never interacted with him on Facebook, it is impossible for BFF to accurately predict the tie strength between you and your brother.

Each question was rated on a scale 1-5: 1 = very bad, and 5 = very good. The first and second question addressed specific parts of BFF (the grouping feature and the tie strength prediction respectively). The third and fourth questions were general questions. The intention of the fourth question was to clarify the limitations of BFF to the users. Currently, BFF is limited to the bounds of Facebook; therefore, it only considers the interactions and social connections that occur on Facebook. In future work, we expect to collect information from different sources than Facebook, so BFF will be able to avoid this limitation.

### 4.3 Results

With regard to tie strength prediction, the module performed very accurately. It achieved a Mean Absolute Error of 0.1155 on a discrete scale 1-5, where 1 is the weakest and 5 is the strongest. We chose to discretize<sup>1</sup> the tie strength in order to facilitate the understanding of the results to the users. Moreover, according to our findings, when the tie strength module predicted incorrectly, 51% of the time it overrated tie strength and 49% of the time it underrated the strength. This suggests that tie strength prediction is not biased.

The performance of the community prediction module was also very accurate; it achieved an accuracy of over 95%. The participants performed mainly two types of modifications on friend community predictions:

- The participant divided the largest community into several sub-communities. An interesting fact about this situation is that the moved contacts usually had a low tie strength (2.5 average). Thus, tie strength may affect the way a user groups his/her friends. An idea for future research is to determine how tie strength may play a role in the community prediction.
- The second more common modification was the user combining communities formed by only one or two members with low tie strength into a larger community. These new communities can be identified as communities of acquaintances. It seems that participants preferred to manage these contacts as a single group, even when they did not share anything but the fact that they had few friends in common and a low tie strength value.

Another important factor to analyze was the number of corrections that the participants needed to make to the suggestions. As stated previously, SNS users struggle to set up privacy settings. If the aim of BFF is to lighten the burden of this task, its suggestions cannot contain a huge number of errors that need correction.

---

<sup>1</sup>The discretization process might have caused a higher prediction error. For example, a user with a tie strength of 3.6 and another with a strength of 4.4 will be both assigned a strength of 4 during the discretization process. As future work, we plan to study the effect of discretization in the prediction error, so that we could achieve a trade-off between the understandability of the results and the error introduced because of the discretization.

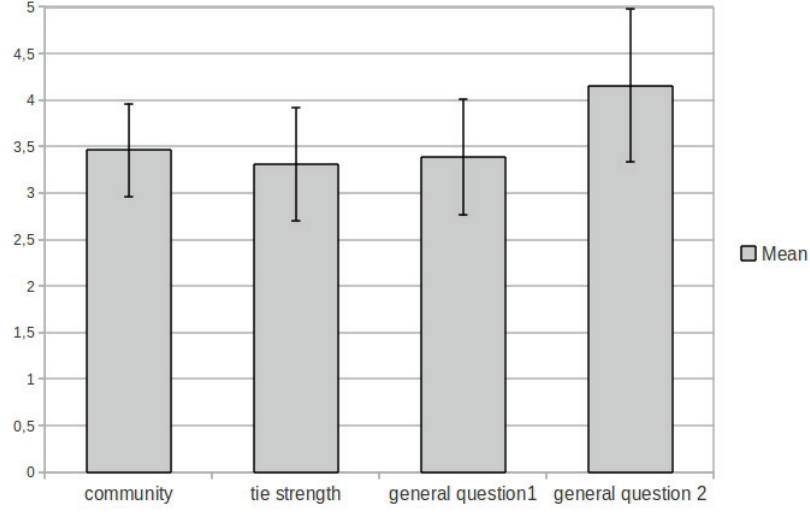


Figure 3: Mean and standard deviation for the survey questions

The mean number of corrections made was 19.3 per participant. Specifically, the participants made an average of 13.12 tie strength corrections and an average of 6.2 community corrections. Considering that the average number of friends of our participants was 186.94, having to perform only 19.3 changes could speed up the process of organizing friends before setting privacy policies.

The participants also rated the performance of BFF by answering a short survey. The results show that the participants rated BFF performance positively. The participants perceived a slightly better accuracy in community prediction than in tie strength. This shows that tie strength prediction is a more complex task due to the high number of variables that the model considers. Another result to note is that the participants rated the second general question (question 4) higher than the first general one (question 3). When answering the first general question, the participants did not consider the limitations of BFF. Therefore, even when almost every friend was rated correctly, they detected mistakes. Due to the brief explanation in the second question about how BFF works, the participants realized that BFF is limited by the bounds of Facebook, and, for example, that it cannot predict the tie strength of a relationship that mainly occurs outside Facebook. When the participants became aware of the limitations of BFF, they took into account how they interacted with others on Facebook in order to make their judgments. This explains the better rating for the second general question.

## 5 Related Work

Recent works have proposed models to predict tie strength. Gilbert et al.[10] proposed a model, based on Granovetter's work, that predicted tie strength among the users of Facebook. The authors identified a set of 74 predictive variables that can be found on Facebook. They achieved an accuracy of 84%. Another work that

predicts tie strength of social links is [14]. Like in the work of Gilbert, the authors define a set of 50 predictive variables. In this work the authors aim to discriminate strong links from weak links. However, they do not consider a scale in the strength of the link, they are either strong or weak. These two works use a supervised learning model that needs human intervention to work properly. Aiming at the same objective, Xiang et al.[25] proposed a model to infer relationship strength based on profile similarity and interaction activity, with the goal of automatically distinguishing strong relationships from weak ones. It is worth noting that this model relies on an unsupervised learning method, but it lacks a empirical evaluation with real users. All three works show that it is possible to infer tie strength from the available personal data in a SNS. These three works differ from ours in that they aim to create models to predict tie strength from the information available on a SNS. However, they do not offer tools that social network users can use to help them to form friend groups and set privacy policies. Moreover, they only consider the predictive capabilities of the variables chosen for their models, but they do not take into account factors like the computational cost of collecting these variables, which is an important factor when creating a usable tool.

The other main feature of BFF is that it suggests friend groups to the participant user. The main idea is that with the grouping and tie strength information the user has enough elements to create appropriate privacy policies. The work of Fang and coworkers [7] proposes a tool that suggests privacy policies for certain elements of a Facebook user profile. This work bases the privacy suggestions in grouping user's contacts in contexts. Every contact in the same context is granted the same access permissions. The authors present a tool called Privacy Wizard that helps user to set the privacy policies to protect user's traits, like birth date, address, and telephone number. However, this work does not consider tie strength, and as the authors proved in [24], it is a key variable to consider when determining the disclosure degree of the elements being shared in a social network.

Other works present mechanisms that can partially infer users' social network and its characteristics from sources of information different than SNSs. In [5] the authors propose a method that extracts a social network for a user given her mailbox and the information available on Internet. A similar approach is presented in [18]. In this work the authors present POLYPHONET. From a given set of persons, the authors find the social connections among them by querying to Google. The authors estimate the strength of the relationship between two persons by co-occurrences of their two names. These two works differ from ours in that they do not rely in a SNS to extract social information from users. However, this approach also has limitations. Relying on information sources that do not necessarily contain social relevant information may lead to errors. For example, two persons may appear in several web pages together but do not have any social link. In order to avoid this problem, both works ([18, 5]) require a predefined set of persons that will form the social community. In contrast, relying only on Facebook data guarantees that the social links will actually exist, but may also lead to errors. Even when the connection truly exists, the interactions between two persons may occur outside Facebook. Therefore, the strength of such link will be incorrectly predicted by our software. In the future, we plan to expand the search of variables for defining the groups and the tie strength with information

that can be found outside the social network, like the information available in the participant's mailbox or in the personal web page of a user of the social network.

The work of Murukannaiah and Singh [19] presents Platys Social. The authors developed a software that runs on a mobile device. This software learns a user's social circles and the priority of the user's social connections from daily interactions. The software infers the interactions from information that is available on mobile devices, such as wi-fi networks, bluetooth connections, phone calls, and text messages. The work of Murukannaiah and Singh presents a new approach for extracting social information from the real world, and not only from Internet. Their work and ours could be merged so that tie strength could be computed taking into account day by day encounter frequency and the information stored on a SNS like Facebook.

## 6 Conclusions and Future Work

In this paper, we have presented a new tool for social network mining. This tool is our first attempt to build a software that can help users to better understand their social relationships on a SNS like Facebook. Currently, BFF is focused on community and tie strength prediction. However, in the long term, we plan to expand it with new functionalities and features. The modular architecture of BFF allows us to develop new modules that can be easily added to BFF. These new modules will rely on the capability of BFF to properly predict tie strength and user communities. In order to be confident in the current capabilities of BFF, we evaluated it using real-world data from real users of Facebook. BFF achieved a Mean Absolute Error of 0.1155 for predicting tie strength and an accuracy of 95% in friend grouping. Furthermore, on average, participants only needed to perform 19.3 corrections to BFF suggestions, taking into account that the average number of friends of the participants was 186.94, BFF can positively accelerate the process of organizing friends. Finally, users considered that BFF was good at predicting tie strength and groups, and they considered it to be a good tool overall.

Many research paths open from here. The first one, and the motivation of this work, is to use the extracted information to predict privacy policies. Users limit what they share and with whom depending on the type of the relationship. Therefore, a tool that correctly infers the types of relationships may be able to predict suitable privacy policies. Furthermore, the ability of BFF to create groups of users also matches the functionality of many SNSs that offer the possibility for users to group their friends. Using these two features, we can create a new functionality for privacy policy recommendation. Users perceive the utility of SNSs by sharing photographs, videos, and other items with their contacts. However, privacy issues can stop users from fully enjoying the functionalities of a SNS. By automating the process of privacy policy definition and how the information is disclosed on a SNS, we can reduce the burden that these systems impose on users, thus increasing their utility.

Another path for further research is to add the possibility for BFF to predict tie strength and friend communities using not only the information available at Facebook, but also using other environments for searching. As the works [5, 18, 19]

prove, social information can be extracted from several different environments. The information available at users' mailbox, personal web pages, Internet search engines could be collected by BFF. The development of a module that could be deployed on a mobile device would allow BFF to also consider daily user interactions. The addition of new sources of information will change how tie strength and grouping are predicted. For the tie strength model, new variables will have to be considered, so the weight of the variables may have to change. With regard to the community finding algorithm, connections outside Facebook may increase the weight of some edges, thus changing the selection of core members during the threshold step. Besides, it will be necessary to take into account how the addition of new variables can affect the efficiency of the tie strength and community predictions.

Apart from being of crucial importance for developing autonomous agents that recommend privacy policies to users, the information that our tool provides can also be the basis (or at least it can play a very important role) to solve many other problems. For instance, the tie strength among agents is used to obtain the optimal social trust path in complex social networks [17]. Moreover, agents could judge the outcome of a negotiation as being distributively fair based on the tie strength between them [21].

## 7 Acknowledgements

This work has been partially supported by CONSOLIDER-INGENIO 2010 under grant CSD2007-00022, and TIN 2008-04446 and PROMETEO/2008/051 projects. Ricard L. Fogués is working with a FPI grant from Programa de Ayudas de Investigación y Desarrollo (PAID) de la Universitat Politècnica de València.

## References

- [1] Facebook website. Facebook Statistics <http://www.facebook.com>.
- [2] Yahoo advertising solutions. <http://advertising.yahoo.com/article/flickr.html>.
- [3] D. Boyd and E. Hargittai. Facebook privacy settings: Who cares? *First Monday*, 15(8), 2010.
- [4] R. Burt. *Structural holes: The social structure of competition*. Harvard Univ Pr, 1995.
- [5] A. Culotta, R. Bekkerman, and A. McCallum. Extracting social networks and contact information from email and the web. 2004.
- [6] N. Ellison, C. Steinfield, and C. Lampe. The benefits of facebook friends: social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.
- [7] L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, pages 351–360. ACM, 2010.
- [8] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.
- [9] E. Gilbert. Predicting tie strength in a new medium. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, CSCW '12*, pages 1047–1056, New York, NY, USA, 2012. ACM.

- [10] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 211–220. ACM, 2009.
- [11] M. Girvan and M. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821, 2002.
- [12] M. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.
- [13] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80. ACM, 2005.
- [14] I. Kahanda and J. Neville. Using transactional information to predict link strength in online social networks. In *Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM)*, 2009.
- [15] N. Lin, W. Ensel, and J. Vaughn. Social resources and strength of ties: Structural factors in occupational status attainment. *American sociological review*, pages 393–405, 1981.
- [16] H. Lipford, A. Besmer, and J. Watson. Understanding privacy settings in facebook with an audience view. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, pages 1–8. USENIX Association Berkeley, CA, USA, 2008.
- [17] G. Liu, Y. Wang, and M. Orgun. Optimal social trust path selection in complex social networks. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence, AAAI*, pages 1391–1398, 2010.
- [18] Y. Matsuo, J. Mori, M. Hamasaki, T. Nishimura, H. Takeda, K. Hasida, and M. Ishizuka. Polyphonet: An advanced social network extraction system from the web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(4):262 – 278, 2007. World Wide Web Conference 2006 Semantic Web Track.
- [19] P. Murukannaiah and M. Singh. Platys social: Relating shared places and private social circles. *Internet Computing, IEEE*, (99):1–1, 2011.
- [20] K. Shen, L. Song, X. Yang, and W. Zhang. A hierarchical diffusion algorithm for community detection in social networks. In *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2010 International Conference on*, pages 276–283. IEEE, 2010.
- [21] C. Sierra and J. Debenham. The LOGIC negotiation model. In *AAMAS '07: Proceedings of the 6th international joint conference on Autonomous agents and multi-agent systems*, pages 1–8. ACM, 2007.
- [22] K. Strater and H. Lipford. Strategies and struggles with privacy in an online social networking community. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1*, pages 111–119. British Computer Society, 2008.
- [23] B. Wellman and S. Wortley. Different strokes from different folks: Community ties and social support. *American journal of Sociology*, pages 558–588, 1990.
- [24] J. Wiese, P. Kelley, L. Cranor, L. Dabbish, J. Hong, and J. Zimmerman. Are you close with me? are you nearby? investigating social groups, closeness, and willingness to share. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 197–206. ACM, 2011.
- [25] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.