Predicting Tie Strength with the Facebook API

Tasos Spiliotopoulos
Madeira-ITI, University of Madeira, Portugal / Harokopio University, Greece
tspiliot@gmail.com

Diogo Pereira
University of Madeira
Campus da Penteada, Funchal, Portugal
fabiodfpereira@gmail.com

Ian Oakley
School of Design and Human Engineering
UNIST
Ulsan, South Korea
ian.r.oakley@gmail.com

ABSTRACT
This paper presents a user study that employed a Facebook application to calculate the strength of Facebook users' friendships. Specifically, 18 variables were collected via the Facebook API for 1728 friendships and used to predict tie strength reported by 90 participants. The resulting model had an accuracy of 65.9% in differentiating between strong and weak ties, and 86.3% in differentiating between very strong and weaker ties. The tie-strength calculation was performed in real time by the application, conferring the key advantage that the result can be instantly used by the live application. We argue this functionality has the potential to enable many novel customization and recommendation scenarios. Furthermore, examining the effect of the use of different Facebook features and types of communication on the perceived tie strength gives a more comprehensive understanding of the concept of tie strength in social media and sheds light on people's use of social network sites.

Categories and Subject Descriptors
H.5.3. [**Information interfaces and presentation (e.g., HCI)**]: Group and Organization Interfaces - *web-based interaction, theory and models*.

General Terms
Human Factors; Theory.

Keywords
Social media; Facebook; Social Network Sites; computer-mediated communication; tie strength; API;

1. INTRODUCTION
Social Network Sites (SNSs) are a popular way of creating new connections, keeping in touch with old connections, and communicating with friends and acquaintances [2, 13]. Facebook, in particular, holds a prime position among SNSs with over one billion monthly active users [9]. It has attracted a considerable amount of research attention, much of which has explored the connection between Facebook and key social constructs such as social capital [4, 5], motivations for use [13, 21], information seeking [16], personality [19], information diffusion [1, 22], socioeconomic status [4], and user privacy [3]. However, most of this research occurs at the user level, focusing on the attributes of individuals. This paper argues that the *connections* among users are key building blocks of SNSs such as Facebook and this paper highlights the need for more research examining these links.

Facebook connections take the form of reciprocal friendship links between users that typically enable access to the majority of posted and stored content that each user maintains on the site. As is currently the case with many other SNSs, originally there was no differentiation between the different levels of friendship a user may have with individuals in their Facebook social network; all friendships were treated (and presented to the user) identically by the service. However, such a model seems counter intuitive – a romantic partner is likely a substantially closer friend that a sporting partner. To counter this imbalance, Facebook introduced the friend list feature, whereby users can assign group-based permissions to (mostly) manually specified groups of friends. Recent research, however, has uncovered practical issues with this feature: low adoption of the feature and low effort towards maintenance of the lists; inconsistencies in the mechanisms, strategies, and ultimately the groups that people create; and a general lack of efficiency of this one-time grouping approach for meeting users' need to share content in a privacy-preserving manner [15]. Addressing these problems, automated approaches that help users create and modify Facebook friend lists have been well received [15] and automated creation of friend lists based on tie strength has been found to be effective on Twitter [10]. In addition to assisting the automated grouping of friends, the calculation of tie strength provides valuable characterization of friendships that this paper argues can enable a larger range of adaptation and customization services.

Notable prior work that has examined tie strength in Facebook relationships has done so asynchronously [11, 18] - the data gathered have been manually collected by the researchers in a dedicated process and using non-standard tools and technologies. This paper presents a study that employs a standard Facebook application to capture data that can be used to calculate tie strength in real time. This approach has the key advantage that the result of the calculation of tie strength can be instantly incorporated into the application. This effectively moves tie strength calculation outside the lab environment, enabling naturalistic studies with real-time calculation of the term. It also has important design implications, as it opens the door for the development of live, adaptive applications that offer customized services, such as adaptive privacy controls, user and content recommendations, and UI personalization.

PCI '14, October 02 - 04 2014, Athens, Greece
Copyright 2014 ACM 978-1-4503-2897-5/14/10…$15.00.
http://dx.doi.org/10.1145/2645791.2645817.

2. TIE STRENGTH

The strength of a tie was introduced by Granovetter [12] as "a (probably) linear combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize a tie". More simply, tie strength refers to the bonding level or closeness between two people and a tie is typically classified as *strong* or *weak*. Strong ties are the people one usually trusts: family and close friends. On the other hand, weak ties are looser or shallower relationships, i.e. acquaintances.

Granovetter [12] first demonstrated that there is value in weak ties; because they are in contact with different social circles, they can be bearers of novel information and can be useful in tasks such as looking for a job. However, more recent studies on social media have provided evidence to the contrary. Communication with strong ties was found to be more predictive of finding employment within three months [6] and provided a subtle increase in useful and novel information [18]. Furthermore, communication with strong ties over social media has been generally associated with improvements in stress levels, social support, and bridging social capital [6]. Interestingly, communication on Facebook is associated with changes in reported relationship closeness, over and above effects attributable to face-to-face, phone, and email contact [7].

3. CALCULATING TIE STRENGTH

Although a number of studies have investigated the strength of Facebook friendships, none has done so in a way that calculates it in real time, allowing the same application that calculates tie strength to incorporate the results into its ongoing behavior. For instance, Gilbert and Karahalios [11] employed a browser script that crawled Facebook web pages to collect data in the form of 74 different variables that would be used to calculate tie strength. Similarly, Panovich and colleagues [18] used Facebook's "Download Your Data" feature to collect six variables while Burke and Kraut [6] examined Facebook user behavior logs. Finally, other studies have examined publically available Facebook data sets and collected data from 50 variables [14], three variables [23], and a single variable [24]. Interestingly, a study utilizing a Facebook real-time tie strength model on Twitter, not only found useful applications for the model, but also found the model to largely generalize to the new medium [10].

Calculating tie strength in real-time from an API requires a complex approach. Even in the context of an academic study, a compromise is needed between users' privacy and the usefulness of the information that an application asks to be disclosed. The current study addresses this issue by selecting a subset of the 74 variables from the previous research [11] based on two criteria; first, the participants would not be made uncomfortable by being requested to disclose too much potentially private information to a Facebook application and, second, the most useful variables would be collected, as identified in the literature (interestingly, the utility distribution of the predictive variables forms a power-law distribution) [11].

4. METHOD

Participants were recruited with a request to complete an online survey, primarily through posts in social network sites, but also a relevant online forum and an online study repository. The online survey was in the form of a Facebook application. The first page of the application included a comprehensive description of the study, clearly framed the experiment as an academic study and explained the data collection process. In addition to this, during the installation of an application the Facebook API ensures that the application displays all data-access permissions granted to it. Thus, we believe participants had a good understanding of the data captured by the study. Apart from the basic profile information, the application requested a single extended permission, "Access posts in your News Feed". Users can deny extended permissions when they install an application, but the study participants were instructed to accept this permission and, in fact, the application was designed so that it would not proceed unless they did so. There was no monetary compensation, but participants were shown a ranking of their Facebook friends at the end of the study.

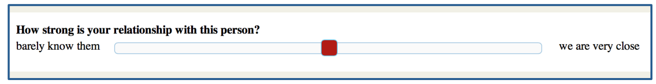

Figure 1. Screenshot of the horizontal slider used for friendship rating.

After completing demographic questions, each participant was presented with the details of a randomly selected friend and asked to rate their closeness to them based on a series of questions. This included one asking: "How strong is your relationship with this person?" Ratings spanned from "barely know them" to "we are very close" and were set by moving a horizontal slider (Figure 1). This was very similar to the method used by Gilbert and Karahalios [11], with the difference that in this study an application using the Facebook API was employed instead of a browser script. The slider had to be moved in order for the application to proceed to the next person as a means to ensure that the participant had rated the friend before moving on. In the meantime, the application gathered the values of 18 predictive variables (see Table 1) for this pair of Facebook users. After rating 20 friends, the participant was presented with summary

Table 1. The 18 variables gathered by the application for each Facebook friendship.

Data Collected	Mean	SD
Wall (timeline) posts exchanged	0.23	0.718
Comments exchanged on wall (timeline) posts	0.09	0.529
Comments on participant's photos [1]	0.05	0.675
Comments on photos where participant is tagged [1]	0.26	1.431
Likes on participant's wall (timeline) posts	0.36	3.718
Likes on photos where participant is tagged	0.23	1.363
Likes on participant's photos[1]	0.05	0.439
Number of mutual friends	34.79	43.38
Number of groups in common [2]	0.67	1.261
Mutual confirmed participation in events [3]	0.03	0.212
Family	0.03	0.217
Number of appearances together on photos	0.33	1.6
Number of wall (timeline) words exchanged	3.13	11.61
Days since first communication	844.1	831.4
Days since last communication	813.6	848.4
Difference in education level	0.59	0.657
Intimacy words exchanged in wall (timeline) posts	0.06	0.362
Participant-initiated wall (timeline) posts	0.22	0.68

Limits: [1]last 200 photographs, [2]50 groups, [3]50 events

Table 2. Regression model of tie strength

Measure	β
Days since last communication	-.641**
Days since first communication	.550*
Family	.159***
Number of wall (timeline) words exchanged	.110**
Number of appearances together on photos	.093**
Comments on photos where participant is tagged	.072*
Number of groups in common	.068**
Number of mutual friends	-.051*
Likes on participant's wall (timeline) posts	.050*
Difference in education level	-.048*
Comments exchanged on wall (timeline) posts	.046
Likes on participant's photos	.027
Likes on photos where participant is tagged	-.023
Intimacy words exchanged in wall (timeline) posts	-.016
Mutual confirmed participation in events	-.010
Wall (timeline) posts exchanged	.010
Comments on participant's photos	.003
Intercept	.292***
R^2	.142
Adjusted R^2	.134

* p < .05, ** p < .01, *** p < .001, all beta coefficients are standardized.

results in the form of a comparison of the subjective ranking of their friends against the objective one performed by the algorithm, as well as some light-hearted commentary derived from the calculated tie strength. Participants were then given the option to rate more people but were also able to quit the application at any time.

In total, 90 participants (59% male) rated 1728 Facebook friendships. Participants with less than 20 and more than 1000 Facebook friends were excluded. The participants had a mean age of 26.9 years (SD = 8.7), and came from 11 countries with the vast majority (n=77, 85.6%) from Portugal and 4.4% (n=4) from the USA. They had a mean of 355 Facebook friends (SD = 218.9, range = 28 – 872) and reported using Facebook for an average of 13.4 (SD = 15.1) hours per week.

4.1 Gathered Data from the API

The application gathered 18 pieces of data for each rated pair (Table 1). Family relationships and the difference in education level were modeled as numerical values of 0, 1, 2, and 3, as in prior work [11]. The *days since last communication* measure refers to the most recent interaction between two users on Facebook from the day of data collection. The *days since first communication* would ideally be calculated based on the date that two users became friends. Since this information is only partially available on Facebook, it is based on the earliest interaction on record. The number of *intimacy words exchanged in wall (timeline) posts* was based on a relevant sentiment analysis dictionary [17], translated from English to Portuguese and used in both languages.

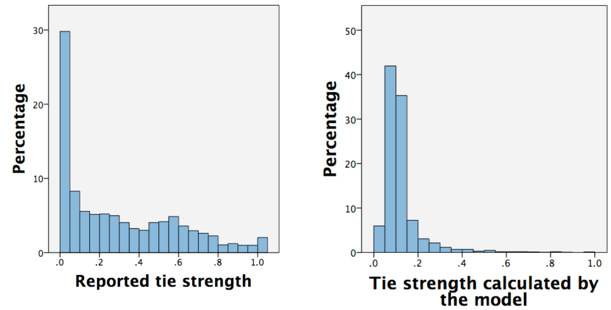

Figure 2. Distribution of reported tie strength of participants and tie strength calculated by the model

5. RESULTS AND DISCUSSION

A multiple regression was run with the data gathered from the API for each rated pair as predictors and the answer to the tie-strength question as the outcome variable. As expected, the two "days since" variables were highly correlated (r = .99, p < .01), as cases of 0 or 1 interaction result in the same "days since" number. Although high correlation among predictor variables is typically a problem in multiple regression, the fact that the two variables carry opposite effects in the regression support the theory from previous research [11] that these variables do not contain the same information about the dependent variable and therefore should both remain as predictors.

The two "days since" variables aside, the correlation matrix showed a number of strong relationships among the predictor variables, one of which exceeded the 0.8 benchmark that indicates potential multicollinearity: "wall (timeline) posts exchanged" and "participant-initiated wall (timeline) posts". Thus, the regression was run again excluding the second variable (with a total of 17 predictor variables). Examination of the Variance Inflation Factor (VIF) for each predictor variable found the highest value of VIF to be 2.667, well below the benchmark value of 10 that indicates multicollinearity. Therefore, we are confident that the regression carried out was free from multicollinearity concerns. Table 2 shows the result of the regression. The standardized beta coefficients show the effect (i.e. the relative weight) of each dependent variable on tie strength.

Overall, in the regression model for tie strength 10 of the 17 employed variables emerged as significant predictors. *Days since first* and *days since last communication* are signs of the duration and intensity of the relationship, respectively, and were found to have a very strong effect on tie strength, as indicated by the literature [11]. As expected, *family* was a significant positive predictor of tie strength, exhibiting the highest beta coefficient after the two *days since* variables.

The number of *wall (timeline) words exchanged* showed a significant positive effect on tie strength; writing text to each other in public may signify greater intimacy than the plain, lightweight communication achieved with likes. The *number of co-appearances in photographs* also emerged as a positive predictor, indicating that strong ties will typically also engage in offline relationships. Similarly, the *number of groups* that a participant had in common with a friend was a strong predictor of tie strength, hinting at the value of homophily; strong ties have similar interests and belong to the same groups. The number of a *friend's comments on photos that the participant was tagged in* was positively associated with tie strength, whereas no other types of comments were found to have a significant effect. This is likely a sign of intimacy, as photographs where the participant is present are conceivably more personal than other types of photos. The number of *likes on a participant's wall (timeline) posts* was also positively associated with tie strength.

Interestingly, the number of *mutual friends* was negatively associated with tie strength, something that at first seems counter-intuitive. This can be explained by the fact that a Facebook user may belong to several social contexts, some of which can involve large clusters of acquaintances (such as school, university, workplace). People in these large clusters have a large number of mutual Facebook friends because of the very fact that they belong to the same cluster. However, some traditionally very strong ties, such as family members, childhood friends, or generally old and close friends with friendships that span many years of time may not belong in such clusters. Finally, social distance, derived from the *difference in education* variable in our dataset, was also negatively associated with tie strength.

The unstandardised beta coefficients (not shown here but available from the authors on request) were used for the creation of a new linear model for calculating tie strength. Figure 2 shows the distribution of the tie strength of the relationships reported by the participants, as well as the distribution of the tie strengths calculated by our model. As in similar literature [11, 18], the range is normalized between 0 and 1 for each participant, where 0 is the weakest tie strength of a friend, and 1 is the strongest. In line with the findings of previous work [11] the model showed a bias towards underestimation of tie strength. More specifically, the mean of the reported tie strength was measured at 0.29 (median = 0.21) and the mean of the model's tie strength was 0.13 (median = 0.1). It is notable, however, that 19.7% of friendships rated by the participants were set to zero. This means that participants acknowledged that a substantial percentage of their Facebook relationships are essentially non-existent.

To demonstrate the practical applicability of this model, we reduced the tie-strength based relationships to two fundamental categories, weak ties and strong ties. In line with Gilbert and Karahalios' approach [11] we classified all friends above the model's mean value as strong ties and all those below as weak. Correct predictions are those where the participant's rating is correspondingly above or below the mean in the questionnaire dataset. Our tie strength model classified with 65.9% accuracy using this procedure, χ^2 (1, N = 3456) = 135.08, $p < 0.001$. However, given the large number of friendships per person that a Facebook application can have access to and the positive skew of the distribution of the reported and calculated tie strengths, it is also meaningful to examine the ability of the model to predict the few strongest ties. These would correspond to the close friends of the participant, those that the participant communicates more often and possibly more meaningfully with. Thus, we classified the strongest 10% of ties as very strong and ran the chi-square test again, showing that our model had an accuracy of 86.3% in differentiating between very strong and weaker ties, χ^2 (1, N = 3456) = 107.83, $p < 0.001$.

5.1 Implications for Design

While prior work has demonstrated that tie strength can be derived from social media, this paper demonstrates the feasibility of doing this live and using a standard API. This opens the door to a wide range of novel applications based on adaptable and customized services. For instance, social media systems could recommend information items and filter newsfeeds based on the tie strength of connections. Different types of information, such as questions or status updates, could be broadcast to different contacts for more efficient answers [18] or information diffusion [1, 24]. The default values of privacy controls, or recommended privacy settings for new connections can be set based on tie strength. Better recommendations for new connections could be made. For instance, if strong ties A-B and A-C exist, and if B and C are aware of one another, then it has been shown that a "psychological strain" may exist between B and C [12] and recommendations that these users become friends might best be avoided. Finally, even though social network analysis has proven useful in providing an understanding of social media [4, 21], it has also been suggested that appropriate use of network analysis depends on choosing the right network representation for the problem at hand [8]. This paper argues that studying alternative network representations, such as those whose links are weighted based on tie strength instead of binary friendships, has the potential to be the basis for substantial advances in understanding user behavior in social media.

There are also limitations to this work. The sample used in the study is relatively homogeneous, i.e. most participants are Portuguese. More diverse samples should be used in future studies in order to be more representative of the Facebook user population. There are also more conceptual issues. For example, although studies employing computationally collected usage metrics (like those captured in this paper) can provide many practical insights, arguably there is much about social media usage that falls outside the scope of such metrics. For example, recent work suggests that one reason why users choose not to post content is because they are in an offline social context, such as a meeting, where it would be inappropriate to do so [20]. Since such behavior would not be reflected in the usage data, it is possible that a more complete picture could be obtained with the combination of computational and qualitative data. To deal with this issue, future work on this topic should explore such mixed methods approaches.

6. CONCLUSIONS

This paper presented a user study that employed a Facebook application for the calculation of the strength of Facebook user friendships in real time. This approach enables a closer scrutiny of the factors that affect tie strength in social media. Furthermore, this kind of application effectively moves tie strength calculation outside the lab environment, enabling naturalistic studies with real-time calculation of the term. We believe that this work opens the door for the development of live, adaptive applications that offer customized services and we expect to see future research that builds on the results of this work.

7. ACKNOWLEDGEMENTS

The work reported in this paper is supported by FCT research grant SFRH/BD/65908/2009.

8. REFERENCES

[1] Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. The role of social networks in information diffusion. In *Proc. WWW 2012*, ACM (2012).

[2] boyd, D. and Ellison, N.B. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication 13*, 1 (2007), 210–230.

[3] boyd, D. Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence. *Convergence 14*, 1 (2008).

[4] Brooks, B., Welser, H.T., Hogan, B., and Titsworth, S. Socioeconomic Status Updates: Family SES and emergent social capital in college student Facebook networks. *Information, Communication & Society 14*, 4 (2011), 529–549.

[5] Burke, M., Kraut, R., and Marlow, C. Social capital on Facebook: Differentiating uses and users. In *Proc. CHI 2011*, ACM (2011), 571–580.

[6] Burke, M. and Kraut, R. Using Facebook after Losing a Job: Differential Benefits of Strong and Weak Ties. In *Proc. CSCW 2013*, ACM (2013).

[7] Burke, M. and Kraut, R. Growing Closer on Facebook: Changes in Tie Strength Through Social Network Site Use. In *Proc. CHI 2014*, ACM (2014).

[8] Butts, C.T. Revisiting the foundations of network analysis. *Science (New York, N.Y.) 325*, 5939 (2009), 414–6.

[9] Company Info - Facebook Newsroom. http://newsroom.fb.com/company-info/. Accessed 15 May 2014.

[10] Gilbert, E. Predicting tie strength in a new medium. In *Proc. CSCW 2012*, ACM (2012).

[11] Gilbert, E. and Karahalios, K. Predicting tie strength with social media. In *Proc. CHI 2009*, ACM (2009).

[12] Granovetter, M.S. The Strength of Weak Ties. *The Journal of applied psychology 78*, 6 (1973), 1360–1380.

[13] Joinson, A. “Looking at”, “looking up” or “keeping up with” people?: Motives and use of Facebook. In *Proc*. *CHI 2008*, ACM (2008), 1027–1036.

[14] Kahanda, I. and Neville, J. Using transactional information to predict link strength in online social networks. In *Proc. ICWSM 2009*, AAAI (2009), 74–81.

[15] Kelley, P.G., Brewer, R., Mayer, Y., Cranor, L.F., Sadeh, N. An Investigation into Facebook Friend Grouping. In *Proc. Interact 2011*, IFIP (2011), 216–233.

[16] Lampe, C., Vitak, J., Gray, R., and Ellison, N. Perceptions of Facebook’s value as an information source. In *Proc. CHI 2012*, ACM (2012).

[17] Nielsen, F. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Arxiv preprint arXiv:1103.2903*, (2011).

[18] Panovich, K., Miller, R., and Karger, D. Tie strength in question & answer on social network sites. In *Proc. CSCW 2012*, ACM (2012).

[19] Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., and Crowcroft, J. The personality of popular Facebook users. In *Proc. CSCW 2012*, ACM (2012).

[20] Sleeper, M., Balebako, R., Das, S., McConahy, A.L., Wiese, J., and Cranor, L.F. The post that wasn’t: exploring self-censorship on facebook. In *Proc. CSCW 2013*, ACM (2013).

[21] Spiliotopoulos, T., and Oakley, I. Understanding Motivations for Facebook Use: Usage Metrics, Network Structure, and Privacy. In *Proc. CHI 2013,* ACM (2013).

[22] Ugander, J., Backstrom, L., Marlow, C., and Kleinberg, J. Structural diversity in social contagion. In *PNAS 109*, 16 (2012), 5962–6.

[23] Xiang, R., Neville, J., and Rogati, M. Modeling relationship strength in online social networks. In *Proc. WWW 2010*, ACM (2010).

[24] Zhao, J., Wu, J., and Feng, X. Information propagation in online social networks : a tie-strength perspective. *Knowledge and Information Systems*, (2011).