

Predicting Tie Strength With Social Media

Eric Gilbert and Karrie Karahalios

University of Illinois at Urbana-Champaign

[egilber2, kkarahal]@cs.uiuc.edu

ABSTRACT

Social media treats all users the same: trusted friend or total stranger, with little or nothing in between. In reality, relationships fall everywhere along this spectrum, a topic social science has investigated for decades under the theme of *tie strength*. Our work bridges this gap between theory and practice. In this paper, we present a predictive model that maps social media data to tie strength. The model builds on a dataset of over 2,000 social media ties and performs quite well, distinguishing between *strong* and *weak* ties with over 85% accuracy. We complement these quantitative findings with interviews that unpack the relationships we could not predict. The paper concludes by illustrating how modeling tie strength can improve social media design elements, including privacy controls, message routing, friend introductions and information prioritization.

Author Keywords

Social media, social networks, relationship modeling, ties, sns, tie strength

ACM Classification Keywords

H5.3. Group and Organization Interfaces; Asynchronous interaction; Web-based interaction.

INTRODUCTION

Relationships make social media *social*. Yet, different relationships play different roles. Consider the recent practice of substituting social media friends for traditional job references. As one hiring manager remarked, by using social media “you’ve opened up your rolodex for the whole world to see” [38]. To the dismay of applicants, employers sometimes cold call social media friends expecting a job reference “only to find that you were just drinking buddies.” Although clearly not the norm, the story illustrates a basic fact: not all relationships are created equal.

For decades, social science has made much the same case, documenting how different types of relationships impact individuals and organizations [16]. In this line of research, relationships are measured in the currency of *tie strength* [17]. Loose acquaintances, known as *weak ties*, can help a

friend generate creative ideas [4] or find a job [18]. They also expedite the transfer of knowledge across workgroups [20]. Trusted friends and family, called *strong ties*, can affect emotional health [36] and often join together to lead organizations through times of crisis [24]. Despite many compelling findings along this line of research, social media does not incorporate tie strength or its lessons. Instead, all users are the same: friend or stranger, with little or nothing in between. Most empirical work examining large-scale social phenomena follows suit. A link between actors either exists or not, with the relationship having few properties of its own [1, 2, 27].

This paper aims to bridge the gap, merging the theory behind tie strength with the data behind social media. We address one central question. With theory as a guide, can social media data predict tie strength? This is more than a methodological or theoretical point; a model of tie strength has the potential to significantly impact social media users. Consider automatically allowing the friends of strong ties to access your profile. Or, as one participant cleverly suggested, remaking Facebook’s Newsfeed to get rid of “people from high school I don’t give a crap about.” The model we present builds on a dataset of over 2,000 Facebook friendships, each assessed for tie strength and described by more than 70 numeric indicators. It performs with surprising accuracy, modeling tie strength to 10-point resolution and correctly classifying friends as *strong* or *weak* ties more than 85% of the time.

We begin by reviewing the principles behind tie strength, and then discuss its proposed dimensions. Using theory to guide the selection of predictive variables, we next present the construction of our tie strength model. It performs well, but not perfectly. To understand our model’s limitations, we also present the results of follow-up interviews about the friendships we had the most difficulty predicting. The paper concludes by applying our findings toward implications for theory and practice.

TIE STRENGTH

Mark Granovetter introduced the concept of *tie strength* in his landmark 1973 paper “The Strength of Weak Ties” [17]. In this section we review *tie strength* and the substantial line of research into its characteristics. We then discuss four researchers’ proposals for the dimensions of tie strength, laying a foundation for our treatment of it as a predictable quantity. The section concludes by introducing the research questions that guide the rest of this paper.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.

Copyright 2009 ACM 978-1-60558-246-7/09/04...\$5.00.

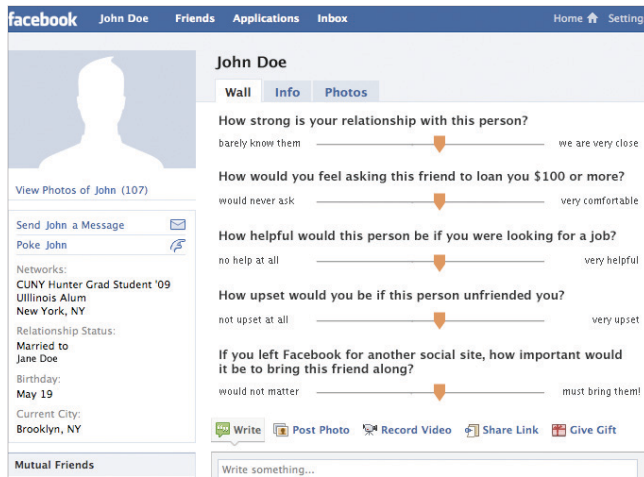


Figure 1. The questions used to assess tie strength, embedded into a friend's profile as participants experienced them. An automated script guided participants through a random subset of their Facebook friends. As participants answered each question by dragging a slider, the script collected data describing the friendship. The questions reflect a diversity of views on tie strength.

Definition and Impact

The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie. [17]

While Granovetter left the precise definition of tie strength to future work, he did characterize two types of ties, *strong* and *weak*. Strong ties are the people you really trust, people whose social circles tightly overlap with your own. Often, they are also the people most like you. The young, the highly educated and the metropolitan tend to have diverse networks of strong ties [31]. Weak ties, conversely, are merely acquaintances. Weak ties often provide access to novel information, information not circulating in the closely knit network of strong ties.

Many researchers have adopted tie strength as an analytic framework for studying individuals and organizations [16]. (Google Scholar, for instance, claims that over 7,000 papers cite "The Strength of Weak Ties" [15].) The social support offered by strong ties can actually improve mental health [36]. Banks that find the right mix of weak and strong ties to other firms tend to get better financial deals [39]. It has also been shown that weak ties, as opposed to strong ones, benefit job-seekers [18]. However, socioeconomic class reverses this effect: job-seekers from lower socioeconomic backgrounds often rely heavily on strong ties [16].

Strong ties between employees from different organizational subunits can help an organization withstand a time of crisis [24]. Yet, strongly tied coworkers are also the ones likely to create crises by pushing for institutional change [23]. Employees who weakly tie themselves beyond organizational boundaries tend to receive better performance reviews and generate more creative ideas [4]. Weak ties also act as a conduit for useful information in computer-mediated communication [8]. However, weak ties often rely on a few commonly available media [22], whereas strong ties diversify, communicating through many channels [21].

How strong is your relationship with this person?

barely know them _____ we are very close

How would you feel asking this friend to loan you \$100 or more?

would never ask _____ very comfortable

How helpful would this person be if you were looking for a job?

no help at all _____ very helpful

How upset would you be if this person unfriended you?

not upset at all _____ very upset

If you left Facebook for another social site, how important would it be to bring this friend along?

would not matter _____ must bring them

The Dimensions of Tie Strength

At what point is a tie to be considered weak? This is not simply a question for the methodologically curious ... the theory makes a curvilinear prediction. How do we know where we are on this theoretical curve? Do all four indicators count equally toward tie strength? [23]

Granovetter proposed four tie strength dimensions: *amount of time*, *intimacy*, *intensity* and *reciprocal services*. Subsequent research has expanded the list. Ronald Burt proposed that structural factors shape tie strength, factors like network topology and informal social circles [5]. Wellman and Wortley argue that providing emotional support, such as offering advice on family problems, indicates a stronger tie [40]. Nan Lin, et al., show that social distance, embodied by factors such as socioeconomic status, education level, political affiliation, race and gender, influences tie strength [29].

In theory, tie strength has at least seven dimensions and many manifestations. In practice, relatively simple proxies have substituted for it: communication reciprocity [11], possessing at least one mutual friend [37], recency of communication [28] and interaction frequency [13, 17]. In a 1984 study, Peter Marsden used survey data from three metropolitan areas to precisely unpack the predictors of tie strength [33]. While quite useful, Marsden pointed out a key limitation of his work: the survey asked participants to recall only their three closest friends along with less than ten characteristics of the friendship.

The present research can be seen as updating Marsden's work for the era of social media. Our work differs primarily in setting and scale. By leveraging social media, participants no longer have to recall; we can take advantage of long friend lists and rich interaction histories. In this way, our work also overcomes the problem of retrospective informant accuracy [3, 30, 32]. In addition, a tie strength model built from social media has the potential to feed back into social media, in ways that benefit its users.

Research Questions

The work above leads us to introduce two research questions that guide the remainder of this paper:

R1: The existing literature suggests seven dimensions of tie strength: *Intensity, Intimacy, Duration, Reciprocal Services, Structural, Emotional Support* and *Social Distance*. As manifested in social media, can these dimensions predict tie strength? In what combination?

R2: What are the limitations of a tie strength model based solely on social media?

METHOD

To answer our research questions, we recruited 35 participants to rate the strength of their Facebook friendships. Our goal was to collect data about the friendships that could act, in some combination, as a predictor for tie strength. Working in our lab, we used the Firefox extension Greasemonkey [19] to guide participants through a randomly selected subset of their Facebook friends. (Randomly sampling participants' friends guards against those with large networks dominating the results.) The Greasemonkey script injected five tie strength questions into each friend's profile after the page loaded in the browser. Figure 1 shows how a profile appeared to a participant. Participants answered the questions for as many friends as possible during one 30-minute session. On average, participants rated 62.4 friends ($\sigma = 16.2$), resulting in a dataset of 2,184 rated Facebook friendships.

Social media experiments often employ completely automated data collection. We worked in the lab for two important reasons. First, we captured all data at the client side, after a page loaded at the user's request. This allowed us to stay within Facebook's Terms of Service. More importantly, however, we asked participants to give us sensitive information: their relationship strengths plus personal Facebook data. We collected data in the lab to guard our participants' privacy and to increase the accuracy of their responses.

Predictive Variables

While participants responded to the tie strength questions, our script automatically collected data about the participant, the friend and their interaction history. The tie strength literature reviewed in the previous section pointed to seven major dimensions of predictive variables. With these dimensions as a guide, we identified 74 Facebook variables as potential predictors of tie strength. Table 1 presents 32 of these variables along with their distributions. In choosing these predictive variables, we tried to take advantage of Facebook's breadth while simultaneously selecting variables that could carry over to other social media. Below, we clarify some variables listed in Table 1 and present those not included in the table. All predictive variables make an appearance either in the text or in Table 1.

Intensity Variables

Each Facebook user has a Wall, a public communication channel often only accessible to a user's friends. *Wall words exchanged* refers to the total number of words traded between the participant and the friend via Wall posting. *Inbox messages exchanged* counts the number of appearances by a friend in a participant's Facebook Inbox, a private commu-

Predictive Intensity Variables	Distribution	Max
Wall words exchanged		9549
Participant-initiated wall posts		55
Friend-initiated wall posts		47
Inbox messages exchanged		9
Inbox thread depth		31
Participant's status updates		80
Friend's status updates		200
Friend's photo comments		1352
Intimacy Variables		
Participant's number of friends		729
Friend's number of friends		2050
Days since last communication		1115
Wall intimacy words		148
Inbox intimacy words		137
Appearances together in photo		73
Participant's appearances in photo		897
Distance between hometowns (mi)		8182
Friend's relationship status	 6% engaged 30% single 30% in relationship	
Duration Variable		
Days since first communication		1328
Reciprocal Services Variables		
Links exchanged by wall post		688
Applications in common		18
Structural Variables		
Number of mutual friends		206
Groups in common		12
Norm. TF-IDF of <i>interests</i> and <i>about</i>		73
Emotional Support Variables		
Wall & inbox positive emotion words		197
Wall & inbox negative emotion words		51
Social Distance Variables		
Age difference (days)		5995
Number of occupations difference		8
Educational difference (degrees)		3
Overlapping words in <i>religion</i>		2
Political difference (scale)		4

Table 1. Thirty-two of over seventy variables used to predict tie strength, collected for each of the 2,184 friendships in our dataset. The distributions accompanying each variable begin at zero and end at the adjacent maximum. Most variables are not normally distributed. The *Predictive Variables* subsection expands on some of these variables and presents those not included in this table.

nication channel. *Inbox thread depth*, on the other hand, captures the number of individual Inbox messages sent between the pair. A helpful analogy for *Inbox thread depth* is the number of messages in a newsgroup thread.

Intimacy Variables

To complement our aggregate measures, we used the Linguistic Inquiry and Word Count (LIWC) dictionary to perform content analysis [34]. Our hypothesis was that friends of different tie strengths would use different types of words when communicating. LIWC matches text against lists of word stems assembled into categories. *Wall intimacy words* refers to the number of Wall words matching at least one of eleven LIWC categories: Family, Friends, Home, Sexual, Swears, Work, Leisure, Money, Body, Religion and Health. Similarly, *Inbox intimacy words* refers to the number of Inbox words matching at least one of these categories. The Home category, for example, includes words like *backyard* and *roommate*, while the Work category includes *busy*, *classes* and *commute*. In total, the intimacy variables checked for matches against 1,635 word stems. Although not presented in Table 1, we also included each LIWC intimacy category as its own predictive variable.

Days since last communication measures the recency of written communication in some Facebook channel (Wall, Inbox, photo comments) from the day we collected data.

Duration Variable

We did not have access to the date when two people became friends. Instead, *Days since first communication* is a proxy for the length of the friendship. It measures time in the same way as *Days since last communication*.

Reciprocal Services Variables

Facebook friends have relatively few opportunities to exchange informational, social or economic goods. (These practices clearly differ by social media; consider a LinkedIn user who exploits his social capital by introducing business contacts to one another.) To capture *Reciprocal Services* on Facebook, *Links exchanged by wall post* measures the number of URLs passed between friends via the Wall, a common Facebook practice. Similarly, *Applications in common* refers to the number of Facebook applications a participant and friend share. Facebook applications usually provide a tightly scoped service (e.g., displaying a virtual bookshelf on a profile) and often spread between friends by word of mouth.

Structural Variables

Facebook allows users to join groups organized around specific topics and interests. *Groups in common* refers to the number of Facebook groups to which both the participant and the friend belong. *Normalized TF-IDF of interests and about* measures the similarity between the free text *interests* and *about* profile fields. It does so by computing the dot product between the TF-IDF vectors representing the text. TF-IDF is a standard information retrieval technique [10] that respects the baseline frequencies of different words in the English language. We also measured *Number of overlapping networks*, the number of Facebook networks to which both the participant and the friend belong. Facebook

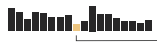



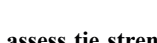
Dependent Variables	0–1 Scale	Median
How strong is your relationship?		0.411
How comfortable asking for loan?		0.076
How helpful if looking for job?		0.362
How upset if unfriended?		0.552
How important to bring friend?		0.324

Table 2. The five questions used to assess tie strength, accompanied by their distributions. The distributions present participant responses mapped onto a continuous 0–1 scale. Our model predicts these responses as a function of the variables presented in Table 1.

networks often map to universities, companies and geographic areas.

Emotional Support Variables

In a way similar to the content analysis variables described above, *Wall & inbox positive emotion words* is two variables referring to matches against the LIWC category Positive Emotion. The Positive Emotion category includes words like *birthday*, *congrats* and *sweetheart*. Similarly, *Wall & inbox negative emotion words* is two variables counting matches in the Negative Emotion category, including words like *dump*, *hate* and *useless*. We also recorded the number of gifts given between a participant and a friend. A Facebook gift is a small icon often given to a friend to show support. Gifts sometimes cost a small amount of money.

Social Distance Variables

We measured the difference in formal education between a participant and a friend in terms of academic degrees. It is computed by searching for the letters *BS*, *BA*, *MS*, *MA*, *JD*, *MD* and *PhD* in the *education* profile field. *Educational difference* measures the numeric difference between a participant and a friend along a scale: 0: None, 1: BS/BA, 2: MS/MA, 3: JD/MD/PhD.

1,261 people in our dataset completed the *politics* profile field. Of those, 79% reported their political affiliation as *very conservative*, *conservative*, *moderate*, *liberal* or *very liberal*. Assigning a scale in that order, *Political difference* measures the numeric difference between a participant and a friend. While the education and politics scales do not completely reflect the diversity of our sample, they do provide useful tools for assessing the importance of these variables for the majority of it.

Demographic and Usage Variables

Finally, in addition to the variables described above, we collected demographic and usage information on our participants and their friends: gender, number of applications installed, number of inbox messages, number of wall posts and number of photo comments.

Dependent Variables

Previous literature has proposed various manifestations of tie strength [17, 18, 21, 24]. To capture a diversity of views, we asked our participants to answer five tie strength questions. Participants moved a slider along a continuum to rate a friend. Figure 1 shows how those questions were embed-

ded into a friend's profile. Table 2 illustrates the responses. We chose a continuum instead of a discrete scale for three reasons. First, Mark Granovetter conjectured that tie strength may in fact be continuous [17]. The literature has not resolved the issue, let alone specified how many discrete tie strength levels exist. A continuum bypasses that problem. Second, a continuum lends itself to standard modeling techniques. Finally, applications can round a continuous model's predictions to discrete levels as appropriate.

Participants

Our 35 participants, primarily students and staff from the University of Illinois community, came from more than 15 different academic departments. The sample consisted of 23 women (66%) and 12 men (34%) ranging between 21 and 41 years old, with a mean and median of 26. The minimum number of Facebook friends was 25; the maximum was 729 (median of 153). In terms of age and number of friends, previous empirical work suggests that our participants fall within the mainstream of Facebook users [14, 35]. All participants used Facebook regularly and had been members for at least one year.

Statistical Methods

We modeled tie strength as a linear combination of the predictive variables, plus terms for dimension interactions and network structure:

$$s_i = \alpha + \beta R_i + \gamma D_i + N(i) + \epsilon_i$$

$$N(i) = \lambda_0 \mu_M + \lambda_1 med_M + \sum_{k=2}^4 \sum_{s \in M} \lambda_k (s - \mu_M)^k$$

$$+ \lambda_5 min_M + \lambda_6 max_M$$

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

More complex models were explored, but a (mostly) linear model allows us to take advantage of the full dataset and explain the results once it is built. In the equations above, s_i represents the tie strength of the i^{th} friend. R_i stands for the vector of 67 individual predictive variables. ϵ_i is the error term. D_i represents the pairwise interactions between the dimensions presented in Table 1. Pairwise interactions are commonly included in predictive models [12]; in this case, including all pairwise interactions would force more variables than data points into the model. Instead, we nominated variables with the fewest missing values to represent each dimension. (Not every participant or friend contributes every variable.) D_i represents all pairwise interactions between the 13 variables with a 90% or greater completion rate. Choosing 90% as a threshold ensured that every dimension was represented. To the best of our knowledge, exploring the interactions between the dimensions of tie strength is a novel approach.

$N(i)$ encodes network structure. It captures the idea that a friendship's tie strength not only depends on its history, but also on the tie strengths of mutual friends. In other words, it models the idea that a friend who associates with your business acquaintances is different than one who knows your mother, brother and sister. Since every friend has a potentially unique set of mutual friends, the model uses seven descriptors of the tie strength distribution over mutual

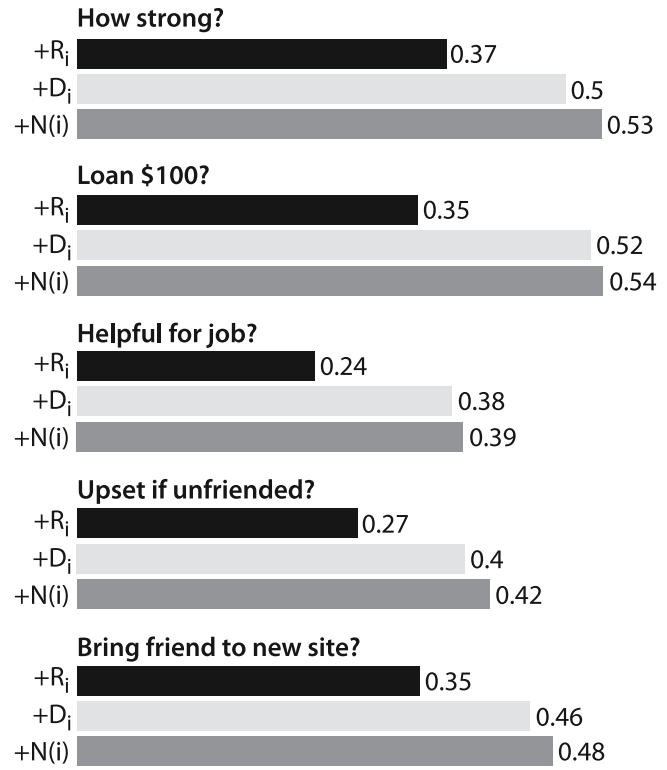


Figure 2. The model's Adjusted R^2 values for all five dependent variables, broken down by the model's three main terms. Modeling interactions between tie strength dimensions results in a substantial performance boost. The model performs best on *Loan \$100?* and *How strong?*, the most general question.

friends: mean, median, variance, skew, kurtosis, minimum and maximum. These terms belong to the *Structural* dimension. However, $N(i)$ introduces a dependency: every tie strength now depends on other tie strengths. How can we incorporate the tie strengths of mutual friends when it is tie strength we want to model in the first place? To solve this problem, we fit the equations above using an iterative variation of OLS regression. In each iteration, the tie strengths from the previous round are substituted to calculate $N(i)$, with all s_i initially set to zero. (Note that $N(i)$ is mostly linear in the predictive variables.) Using this procedure, all s_i converged in nine iterations (0.001 average relative change threshold). This approach parallels other "neighborhood effect" models [6].

We did not standardize, or "ipsatize" [9], the dependent variables. Because we employed network subsampling, we could not be sure participants saw the Facebook friend they would rate highest or lowest. Furthermore, not all real-life friends have Facebook accounts. It is reasonable to assume that some participants would reserve the ends of the spectra for people our experiment would never turn up. Finally, to account for the violations of normality exhibited by the distributions in Table 1, every variable is log-transformed.

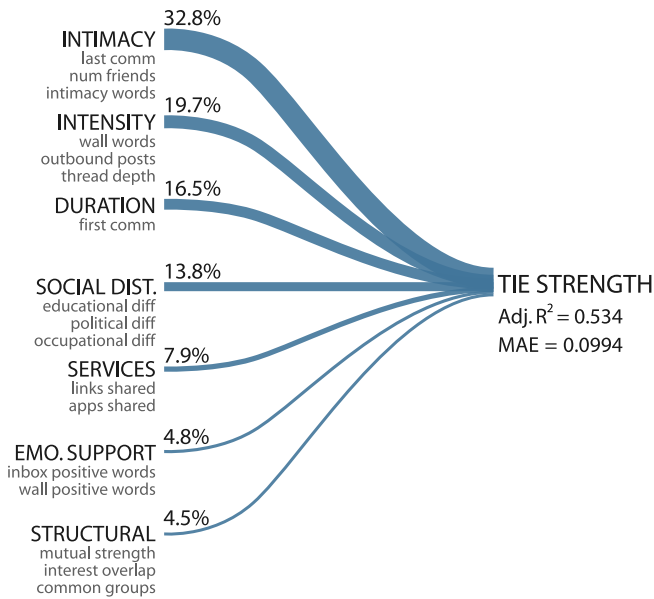


Figure 3. The predictive power of the seven tie strength dimensions, presented here as part of the *How strong?* model. A dimension's weight is computed by summing the absolute values of the coefficients belonging to it. The diagram also lists the top three predictive variables for each dimension. On average, the model predicts tie strength within one-tenth of its true value on a continuous 0–1 scale.

RESULTS

Because each participant rated more than one friend, observations within a participant were not independent. This is a common obstacle for ego-centric designs. To roughly adjust for it, all of the results presented here cut the degrees of freedom in half, a technique borrowed from the social networks literature [33].

On the first tie strength question, *How strong is your relationship with this person?*, the model fits the data very well: $Adj. R^2 = 0.534$, $p < 0.001$. It achieves a Mean Absolute Error of 0.0994 on a continuous 0–1 scale, where 0 is weakest and 1 is strongest. In other words, on average the model predicts tie strength within one-tenth of its true value. This error interval tightens near the ends of the continuum because predictions are capped between 0 and 1. In addition, we found strong evidence of four dimension interactions ($p < 0.001$): *Intimacy* \times *Structural*, $F_{1,971} = 12.37$; *Social Distance* \times *Structural*, $F_{1,971} = 34$; *Reciprocal Services* \times *Reciprocal Services*, $F_{1,971} = 14.4$; *Structural* \times *Structural*, $F_{1,971} = 12.41$. As we demonstrate shortly, the *Structural* dimension plays a minor role as a linear factor. However, it has an important modulating role via these interactions. One way to read this result is that individual relationships matter, but they get filtered through a friend's clique before impacting tie strength.

Figure 2 summarizes the model's performance on all five tie strength questions, broken down by the model's three main terms. Modeling dimension interactions boosts performance significantly, with smaller gains associated with modeling network structure. The model fits the second tie strength question as well as the first: *How would you feel*

Top 15 Predictive Variables	β	F	p-value
Days since last communication	-0.76	453	< 0.001
Days since first communication	0.755	7.55	< 0.001
Intimacy \times Structural	0.4	12.37	< 0.001
Wall words exchanged	0.299	11.51	< 0.001
Mean strength of mutual friends	0.257	188.2	< 0.001
Educational difference	-0.22	29.72	< 0.001
Structural \times Structural	0.195	12.41	< 0.001
Reciprocal Serv. \times Reciprocal Serv.	-0.19	14.4	< 0.001
Participant-initiated wall posts	0.146	119.7	< 0.001
Inbox thread depth	-0.14	1.09	0.29
Participant's number of friends	-0.14	30.34	< 0.001
Inbox positive emotion words	0.135	3.64	0.05
Social Distance \times Structural	0.13	34	< 0.001
Participant's number of apps	-0.12	2.32	0.12
Wall intimacy words	0.111	18.15	< 0.001

Table 3. The fifteen predictive variables with highest standardized beta coefficients. The two *Days since* variables have large coefficients because of the difference between never communicating and communicating once. The utility distribution of the predictive variables forms a power-law distribution: with only these fifteen variables, the model has over half of the information it needs to predict tie strength.

asking this friend to loan you \$100 or more? However, it does not fit the last three questions as well. The lower performance on these questions may have resulted from participant fatigue. We considered randomizing the questions for each friend to account for ordering effects like fatigue, but we feared that randomizing would confuse and frustrate our participants, contributing to lower accuracy across the board. Therefore, we chose to prioritize the first question, the most general of the five. With the exception of *How helpful would this person be if you were looking for a job?*, all dependent variable intercorrelations were above 0.5 (Table 4).

Figure 3 visualizes the predictive power of the seven tie strength dimensions as part of the *How strong?* model. The figure also includes each dimension's top three contributing variables. The weight of a dimension is calculated by summing the coefficients of the variables belonging to it. Although not uniformly distributed, no one dimension has a monopoly on tie strength.

Table 3 presents the standardized beta coefficients of the top fifteen predictive variables. The F statistics signify a variable's importance in the presence of the other variables. The two *Days since* variables have such high coefficients due to friends that never communicated via Facebook. Those observations were assigned outlying values: zero in one case and twice the maximum in the other. In other words, the simple act of communicating once leads to a

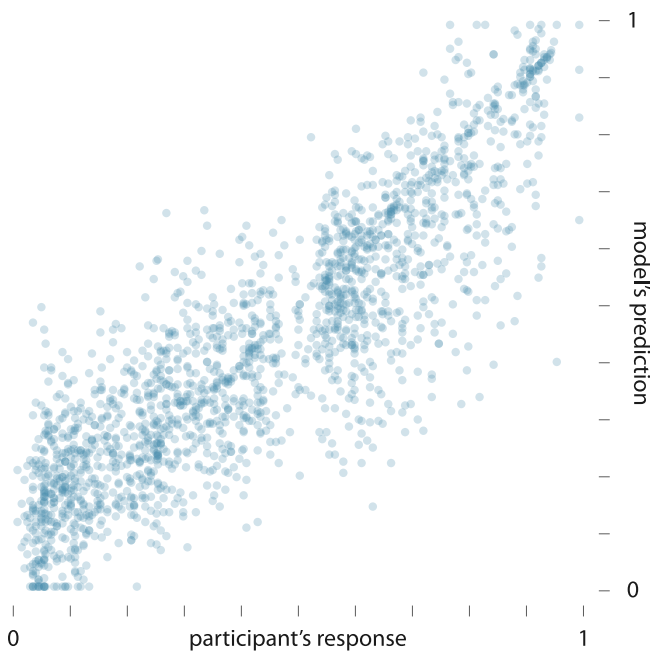


Figure 4. The model's performance across all ties in our dataset. There is a strong correlation, yet the model shows a slight bias toward underestimation, represented as the larger cloud in the bottom-right of the figure. The gap in the center results from participants' inclination to move the slider from its starting point, if only slightly.

very large movement in tie strength. *Educational difference* plays a large role in determining tie strength, but that may reflect the university community from which we sampled participants. Curiously, *Inbox thread depth* has a negative effect on tie strength; the more messages friends exchange on a single topic, the lower their tie strength. It is important to note that Table 3 orders the variables by their weights, or β coefficients, not their p -values. The p -value for *Inbox thread depth* does not express confidence in its coefficient; it expresses confidence in its utility relative to other variables. (The coefficient confidence is greater than 99.9%.) For example, *Inbox thread depth* is highly correlated with *Inbox intimacy words*, resulting in a lower F statistic.

Figure 4 compares the model's prediction to participant responses across the entire dataset. The figure illustrates a strong correlation and another view on the MAE presented above. We discuss the practical significance of the findings illustrated in Figure 4, along with the discretization of tie strength, in the next section.

Error Analysis Interviews

The model performs well, but not perfectly. To understand its limitations, we conducted ten follow-up interviews about the friendships we had the most difficulty predicting. After identifying the friends with the highest residuals, we asked participants to tell us about this particular friendship, including anything that makes it special. For instance, one participant described a "friend" he barely knew:

I don't know why he friended me. But I'm easy on Facebook, because I feel like I'm somehow building (at least a miniscule amount of) social capital, even when I don't know the person.

Correlations	Strong	Loan	Job	Un	Bring
Strong	1	0.69	0.45	0.75	0.7
Loan	0.69	1	0.4	0.55	0.55
Job	0.45	0.4	1	0.5	0.46
Unfriend	0.75	0.55	0.5	1	0.74
Bring	0.7	0.55	0.46	0.74	1

Table 4. The intercorrelations of the five dependent variables. With the exception of *Job-Strong*, *Job-Loan* and *Bring-Job*, the dependent variables are well-correlated with one another.

We went to the same high school and have a few dozen common friends. We've never interacted with each other on Facebook aside from the friending.

rating: 0; prediction: 0.44

Notice how the participant recalls that "he friended me." Although these friends had communicated via Facebook only twice (the participant mistakenly recalled "never"), the friend's clique confused the model. The friend came from a group of relatively strong friends. As we mentioned earlier, the model filters individual relationships through cliques, leading to the high residual. Perhaps having deeper network knowledge could help, such as how the mutual friends see this friend. But this is beyond our ego-centric design.

Asymmetric Friendships

Two participants rated a friend highly because of how the friendship compared to others like it. In one case, a participant described a close bond with a professor:

This is a professor from one of the classes I TA-ed. We have a very good relationship, because in the past we have worked out a lot of difficult class problems. The professor still remembers my name, which for some of my "friends" on Facebook may not be true. But not only that, she also knows how things are going at school, and when we meet in a hallway we usually stop for a little chat, rather than exchanging casual "Hi! Hello!" conversation.

rating: 0.85; prediction: 0.41

Educational difference and the directionality of the wall posts pushed this prediction toward *weak tie*. Many people would not remark that a close friend "remembers my name." However, in the context of this participant's "networking" friends, the professor breaks the mold.

Participants' responses often revealed the complexity of real-life relationships, both online and offline. One participant grounded her rating not in the present, but in the hope of reigniting a friendship:

Ah yes. This friend is an old ex. We haven't really spoken to each other in about 6 years, but we ended up friending each other on Facebook when I first joined. But he's still important to me. We were best friends for seven years before we dated. So I rated it where I did (I was actually even thinking of rating it higher) because I am optimistically hoping we'll recover some of our "best friend"-ness after a while. Hasn't happened yet, though.

rating: 0.6; prediction: 0.11

Confounding the Medium

As might be expected, Facebook friends do not only stick to Facebook. One participant described a close friendship with a diverse digital trail:

This friend is very special. He and I attended the same high school, we interacted a lot over 3 years and we are very very close. We trust each other. My friend and I are still interacting in ways other than Facebook such as IM, emails, phones. Unfortunately, that friend and I rarely interact through Facebook so I guess your predictor doesn't have enough information to be accurate.

rating: 0.96; prediction: 0.47

However, even friends that stick to Facebook sometimes do so in unexpected ways:

We were neighbors for a few years. I babysat her child multiple times. She comes over for parties. I'm pissed off at her right now, but it's still 0.8. ;) Her little son, now 3, also has an account on Facebook. We usually communicate with each other on Facebook via her son's account. This is our "1 mutual friend."

rating: 0.8; prediction: 0.28

This playful use of Facebook clearly confused our model. With the exception of the *Social Distance* dimension, all indicators pointed to a weak tie. In fact, it is hard to imagine a system that could ever (or should ever) pick up on scenarios like this one.

DISCUSSION

Our results show that social media can predict tie strength. The *How strong?* model predicts tie strength within one-tenth of its true value on a continuous 0–1 scale, a resolution probably acceptable for most applications. In other words, discretizing our continuum onto a 10-point Likert scale, the *How strong?* model would usually miss by at most one point. The *Intimacy* dimension makes the greatest contribution to tie strength, accounting for 32.8% of the model's predictive capacity. This parallels Marsden's finding that emotional closeness best reflects tie strength [33]. However, the *Intensity* dimension also contributes substantially to the model, contrasting with Marsden's finding that *Intensity* has significant drawbacks as a predictor. One way to explain this discrepancy is that the sheer number of people available through social media strengthens *Intensity* as a predictor. In other words, when you choose to interact with someone over and over despite hundreds of people from which to choose, it significantly informs tie strength. The number of variables representing each dimension also plays a role in its overall impact. For example, *Emotional Support* might impact tie strength more if more variables represented it. (*Emotional Support* is particularly hard to quantify.) However, more variables does not always equal greater impact. As *Duration* illustrates, a single variable can account for a large part of the model's predictive capacity.

Some applications will not need 10-point resolution; the coarse categories of *strong* and *weak* may suffice. In "The Strength of Weak Ties," Granovetter himself performs his analytic work with only these approximate distinctions. One way to accomplish this is to use the model's mean, classifying all friends above it as *strong* and all below it as *weak*. Correct predictions are those where the participant's rating

is correspondingly above or below the mean in the participant dataset. The *How strong?* model classifies with 87.2% accuracy using this procedure, significantly outperforming the baseline, $\chi^2(1, N = 4368) = 700.9, p < 0.001$. (Note that this situation does not require more sophisticated evaluation techniques, like cross-validation, because the model is highly constrained and the threshold is not learned.)

Some predictive variables surprised us. For instance, *Inbox thread depth* negatively (and strongly) affects tie strength. This finding also clashes with existing work. In [41], Whittaker, et al., report that familiarity between Usenet posters increases thread depth. One way to resolve this disparity is to note that there may be a fundamental difference between the completely private threads found on Facebook (essentially a variant of email) and Usenet's completely public ones. Common ground theory [7] would suggest that strong ties can communicate very efficiently because of their shared understanding, perhaps manifesting as shorter Inbox threads. *Educational difference* also strongly predicts tie strength, with tie strength diminishing as the difference grows. This may have resulted from the university community to which our participants belonged. On the other hand, the result may have something to do with Facebook itself, a community that spread via universities. Some variables we suspected to impact tie strength did not. *Number of overlapping networks* and *Age difference*, while intuitively good predictors, made little appreciable difference to tie strength. ($\beta = 0.027, F_{1,971} = 3.08, p = 0.079$ and $\beta = -0.0034, F_{1,971} = 10.50, p = 0.0012$, respectively.)

The error analysis interviews illustrate the inherent complexity of some relationships. They also point the way toward future research. A model may never, and perhaps should never, predict some relationships. Wanting to reconnect with an ex-boyfriend comes to mind. Relationships like these have powerful emotions and histories at play. However, it may be possible to make better predictions about relationships like the professor-student one, a strong relationship relative to others like it. Incorporating organizational hierarchy may also improve a system's ability to reason about relationships like these. Merging deeper network knowledge with data about who extended the friend request also looks promising, as evidenced by the "he friended me" interview.

Practical Implications

We foresee many opportunities to apply tie strength modeling in social media. Consider privacy controls that understand tie strength. When users make privacy choices, a system could make educated guesses about which friends fall into trusted and untrusted categories. This might also depend on media type, with more sensitive media like photos requiring higher tie strengths. The approach would not help users set privacy levels for brand new friends, ones with whom there is no interaction history. Yet, it has two main advantages over the current state of the art: it adapts with time, and it establishes smart defaults for users setting access levels for hundreds of friends.

Or, imagine a system that only wants to update friends with novel information. Broadcasting to weak ties could solve

this problem. Consider a politician or company that wants to broadcast a message through the network such that it only passes through trusted friends. Because strongly tied friends often reconcile their interests [17], a politician might look for new supporters among the strong ties of an existing one. Limiting the message's audience in this way may increase the success rate relative to the effort expended.

Social media has recently started suggesting new friends to users. However, sometimes we choose not to friend someone with good reason. For instance, a strong tie of a strong tie is not necessarily a friend at all: consider the beloved cousin of a best friend. Granovetter writes, "if strong ties A-B and A-C exist, and if B and C are aware of one another, anything short of a positive tie would introduce a 'psychological strain' into the situation" [17]. A system that understands tie strength might avoid "strain" by steering clear of these delicate situations. In fact, weak ties of existing friends may make better friend candidates, as it is less likely that users have already declined to friend them. More broadly, systems that understand tie strength might apply it to make better friend introductions, although deeper study would need to uncover how best to use it in this context.

Recent work suggests that the average number of social media friends continues to grow, currently above 300 [25]. With users keeping so many friends, social media has started to consolidate friend activity into a single stream. Facebook calls this the Newsfeed. However, the multiplicative nature of the types of friends crossed with the types of updates, e.g., photos, status, new friends, comments, etc., presents a difficult design problem. A system that prioritizes via tie strength, or allows users to tune parameters that incorporate tie strength, might provide more useful, timely and enjoyable activity streams.

Theoretical Implications

There is still more variance to understand. Certainly, more predictive variables could help, such as "behind-the-scenes" data like who friended who. However, throwing more data at the problem might not solve it; perhaps social media needs novel indicators. This raises new questions for theory. When modeling tie strength exclusively from social media, do we necessarily miss important predictors? What is the upper limit of tie strength predictability?

We believe our work makes three important contributions to existing theory. First, we defined the importance of the dimensions of tie strength as manifested in social media. This is novel especially in light of the fact that these weights do not always align with prior work. Second, we showed that tie strength can be modeled as a continuous value. Third, our findings reveal how the *Structural* dimension modulates other dimensions by filtering individual relationships through cliques. Previously, it was not well-understood how or if tie strength dimensions interacted.

Finally, we see a home for our results in social network analysis. Most work to date has assumed a present link or an absent link, omitting properties of the link itself. Introducing a complete tie strength model into social network

analyses, perhaps even joining a social media model with real-world data, may enable novel conclusions about whole systems [26].

Limitations

We purposely worked from theory to extend this research beyond just Facebook. The specific predictive variable coefficients may not move beyond Facebook, but the dimension weights may. That being said, this work looks only at one social media site, at one time, using data available through the browser. We look forward to work evaluating the utility of "behind-the-scenes" data and to work contrasting these findings with other social media.

CONCLUSION

In this paper, we have revealed a specific mechanism by which tie strength manifests itself in social media. Many paths open from here. Social media designers may find traction fusing a tie strength model with a range of social media design elements, including privacy controls and information prioritization. Our follow-up interviews suggest profitable lines of future work. We hope that researchers in this field will find important new theoretical questions in this work, as well as opportunities to use tie strength to make new conclusions about large-scale social phenomena.

We believe this work addresses fundamental challenges for understanding users of socio-technical systems. How do users relate to one another in these spaces? Do the data left behind tell a consistent story, a story from which we can infer something meaningful? We think this work takes a significant step toward definitively answering these questions.

ACKNOWLEDGMENTS

This work was funded in part by NSF Grant 0643502 and by Microsoft. We thank Keith Hampton, Caroline Haythornthwaite, Cliff Lampe, Scott Golder and the entire Social Spaces group for early feedback on this work. We also thank Carolyn Gilbert, who in no way helped this paper get written, but is wonderful so she gets mentioned anyway.

REFERENCES

1. Adamic, L. A. and Adar, E. 2003. Friends and neighbors on the Web. *Social Networks*, 25(3), 211–230.
2. Albert, R. and Barabási, A. L. 2002. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74(1).
3. Bernard, H. R., Killworth, P., et al. 1984. The Problem of Informant Accuracy: The Validity of Retrospective Data. *Annual Review of Anthropology*, 13, 495–517.
4. Burt, R. S. 2004. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2), 349–399.
5. Burt, R. *Structural Holes: The Social Structure of Competition*. Harvard University Press, 1995.
6. Chopra, S., Thampy, T., et al. Discovering the Hidden Structure of House Prices with a Non-parametric Latent Manifold Model. *Proc. KDD*, 2007. 173–182.
7. Clark, H. H. *Arenas of Language Use*. University Of Chicago Press, 1993.

8. Constant, D., Sproull, L., et al. 1996. The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. *Organization Science*, 7(2), 119–135.
9. Cunningham, W. H., Cunningham, I. C., et al. 1977. The Ipsative Process to Reduce Response Set Bias. *Public Opinion Quarterly*, 41(3), 379–384.
10. Frakes, W. B. and Baeza-Yates, R. *Information Retrieval: Data Structures and Algorithms*. Prentice Hall, 1992.
11. Friedkin, N. E. 1980. A Test of Structural Features of Granovetter's Strength of Weak Ties Theory. *Social Networks*, 2, 411–422.
12. Gergle, D., Kraut, R. E., et al. The Impact of Delayed Visual Feedback on Collaborative Performance. *Proc. CHI*, 2006. 1303–1312.
13. Gilbert, E., Karahalios, K., et al. The Network in the Garden: An Empirical Analysis of Social Media in Rural Life. *Proc. CHI*, 2008. 1603–1612.
14. Golder, S., Wilkinson, D. M., et al. Rhythms of Social Interaction: Messaging Within a Massive Online Network. *Proc. 3rd International Conference on Communities and Technologies (CT2007)*, 2007.
15. Google Scholar. <http://scholar.google.com/scholar?q=weak+ties>. Accessed September 16, 2008.
16. Granovetter, M. 1983. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, 1, 201–233.
17. Granovetter, M. S. 1973. The Strength of Weak Ties. *The American Journal of Sociology*, 78(6), 1360–1380.
18. Granovetter, M. *Getting a Job: A Study of Contacts and Careers*. University Of Chicago Press, 1974.
19. Greasemonkey. <http://www.greasespot.net>. Accessed September 16, 2008.
20. Hansen, M. T. 1999. The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge Across Organization Subunits. *Administrative Science Quarterly*, 44(1), 82–111.
21. Haythornthwaite, C. 2002. Strong, Weak, and Latent Ties and the Impact of New Media.. *Information Society*, 18(5), 385–401.
22. Haythornthwaite, C. and Wellman, B. 1998. Work, Friendship, and Media Use for Information Exchange in a Networked Organization. *J. Am. Soc. Inf. Sci.*, 49(12), 1101–1114.
23. Krackhardt, D. The Strength of Strong Ties: The Importance of Philos in Organizations. In N. Nohria and R. Eccles (Ed.), *Networks and Organizations: Structure, Form and Action* (216–239). Boston, MA: Harvard Business School Press.
24. Krackhardt, D. and Stern, R. N. 1988. Informal Networks and Organizational Crises: An Experimental Simulation. *Social Psychology Quarterly*, 51(2), 123–140.
25. Lampe, C., Ellison, N., et al. Changes in Participation and Perception of Facebook. *Proc. CSCW*, 2008. 721–730.
26. Laumann, E. O., Gagnon, J. H., et al. 1989. Monitoring the AIDS Epidemic in the United States: A Network Approach. *Science*, 244(4909), 1186–1189.
27. Liben-Nowell, D. and Kleinberg, J. The Link Prediction Problem for Social Networks. *Proc. of the 12th international conference on Information and knowledge management*, 2003. 556–559.
28. Lin, N., Dayton, P. W., et al. 1978. Analyzing the Instrumental Use of Relations in the Context of Social Structure. *Sociological Methods Research*, 7(2), 149–166.
29. Lin, N., Ensel, W. M., et al. 1981. Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment. *American Sociological Review*, 46(4), 393–405.
30. Marin, A. 1981. Are Respondents More Likely To List Alters with Certain Characteristics?: Implications for Name Generator Data. *Social Networks*, 26(4), 289–307.
31. Marsden, P. V. 1981. Core Discussion Networks of Americans. *American Sociological Review*, 52(1), 122–131.
32. Marsden, P. V. 1990. Network Data and Measurement. *Annual Review of Sociology*, 16(1), 435–463.
33. Marsden, P. V. and Campbell, K. E. 1990. Measuring Tie Strength. *Social Forces*, 63(2), 482–501.
34. Pennebaker, J. W. and Francis, M. E. *Linguistic Inquiry and Word Count*. Lawrence Erlbaum, 1999.
35. Rapleaf. RapLeaf Study of Social Network Users vs. Age. http://business.rapleaf.com/company_press_2008_06_18.html. Accessed September 16, 2008.
36. Schaefer, C., Coyne, J. C., et al. 1990. The Health-related Functions of Social Support. *Journal of Behavioral Medicine*, 4(4), 381–406.
37. Shi, X., Adamic, L. A., et al. 2007. Networks of Strong Ties. *Physica A: Statistical Mechanics and its Applications*, 378(1), 33–47.
38. Tahmincioglu, E. Facebook Friends As Job References? <http://www.msnbc.msn.com/id/26223330>. Accessed September 16, 2008.
39. Uzzi, B. 1999. Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing. *American Sociological Review*, 64(4), 481–505.
40. Wellman, B. and Wortley, S. 1990. Different Strokes from Different Folks: Community Ties and Social Support. *The American Journal of Sociology*, 96(3), 558–588.
41. Whittaker, S., Terveen, L., et al. The Dynamics of Mass Interaction. *Proc. CSCW*, 1998. 257–264.