Predicting Tie Strength in a New Medium

Eric Gilbert
School of Interactive Computing & GVU Center
Georgia Institute of Technology
gilbert@cc.gatech.edu

ABSTRACT
We have friends we consider very close and acquaintances we barely know. The social sciences use the term *tie strength* to denote this differential closeness with the people in our lives. In this paper, we explore how well a tie strength model developed for one social medium adapts to another. Specifically, we present a Twitter application called We Meddle which puts a Facebook tie strength model at the core of its design. We Meddle estimated tie strengths for more than 200,000 online relationships from people in 52 countries. We focus on the mapping of Facebook relational features to relational features in Twitter. By examining We Meddle's mistakes, we find that the Facebook tie strength model largely generalizes to Twitter. This is early evidence that important relational properties may manifest similarly across different social media, a finding that would allow new social media sites to build around relational findings from old ones.

Author Keywords
tie strength, social media, social networks, computer-mediated communication (CMC), collapsed contexts, streams

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

INTRODUCTION
While different on the surface, many social media share some fundamental similarities. They let people build relationships, primarily using language. We type messages to one another. Those messages happen at certain moments in time. Whether we articulate them publicly or not, we form our own micro social networks embedded within macro ones.

These building blocks suggest that we should be able to ask questions about relationships which transcend a particular medium. For example, we have colleagues with whom we correspond intensely, but not deeply; we have childhood friends we consider close, even when we fall out of touch. Do we leave clues in social media signaling our closeness with these people? How often we talk? Our places in a broader social network? The words and phrases we use?

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

The academic literature refers to this differential closeness by the name *tie strength*. We rely on our strong ties for emotional support, but our weak ties keep us in touch with current information [11]. People with many weak ties to other work groups often command higher salaries [2]. When our strong ties find happiness, we sometimes find it too [5]. In a 2009 paper, [7] reported a way to reconstruct tie strength using the data we leave behind in social media. Although it builds its model from Facebook data, [7] speculates that the model may adapt to relationships in other social media too.

Here, we explore that idea. We present *We Meddle*, a Twitter application built around [7]'s Facebook model. Simultaneously an experiment and a real-world system, We Meddle's central feature is that it infers tie strength between you and everyone you follow on Twitter. It generates exportable Twitter lists based on tie strength. The lists allow you to filter your Twitter stream to just a subset of people instead of everyone you follow, similar in spirit to "circles" in the recently released Google+ [9]. At the time of this writing, 2,114 people from 52 countries have used our site. After analyzing more than 200,000 We Meddle tie strength predictions, we arrive at early evidence that tie strength manifests similarly in Twitter and Facebook. To the best of our knowledge, this is the first paper to study in-depth how an aspect of relationships manifests *across* social media. Our method—a hybrid between building something people use voluntarily and studying social life online—has limitations, however. We also examine those limitations in this paper, hopefully directing future work toward ways of achieving greater precision.

We begin by reviewing work on tie strength, socially-rendered social media and the collapsed context problem. Next, we describe how we mapped [7]'s model from Facebook to Twitter. We then introduce We Meddle, which uses the model to generate Twitter lists. Adopting both quantitative and qualitative methods, we examine the model's generalizability and how people used We Meddle. The paper concludes by contextualizing our results within theory and design.

LITERATURE REVIEW
Tie strength refers to a general sense of closeness with another person. When that feeling is strong, we call it a *strong tie;* when it is weak, we call it a *weak tie.* Tie strength is also one of the most influential concepts in sociology. In the years since "The Strength of Weak Ties" introduced tie strength [11], the paper has attracted over 15,000 citations [10] from fields as diverse as Organizational Studies, Finance and Computer Science. Here, we focus on the most seminal and relevant papers from this literature.

We usually trust our strong ties, and their social circles tightly overlap with our own. Often, they also share our values, tastes and interests (i.e., homophily) [11, 20], yet this effect diminishes for people who live in cities [19]. Happiness even *flows* along strong ties in a network [5]. Weak ties, on the other hand, are our acquaintances. Most notably, they provide access to new information, information not flowing through our dense networks of strong ties. For instance, scientific discoveries seem to flow more efficiently through weak ties than through strong ones [12]. In a re-creation of a classic Milgram experiment [21], Lin et al. [18] asked participants to deliver a booklet to some unknown person in a distant place. They found that people who used more weak ties in their paths had greater success reaching the destination. People with weak ties outside their organizations often command higher salaries [2] and obtain better deals for their firms [29].

There are many reasons to care about tie strength, but historically simple heuristics have substituted for it. Communication reciprocity [6], one mutual friend [26], communication recency [18] and interaction frequency [8, 11] have all stood in for tie strength at one time or another. New research claims that we severely skew analyses when we use coarse heuristics like these [3], and leading scholars have called for more refined metrics [33]. Our data support these claims: we estimate that a heuristic like "call it a strong tie if they message each other at least N times" performs at 61% accuracy (at best) on a *strong vs. weak* classification task, with a baseline of 52%. Richer computational models, like the one presented in [7], perform with nearly 90% accuracy on the same task. In this paper, we reappropriate the model presented in [7], a computational technique for predicting tie strength from 74 Facebook interaction variables. Whereas other computational tie strength models have appeared between then and now [16, 34, 32], we chose to use [7]'s model because it builds in two things: simplicity as a path to generalizability and a large set of predictors. We put both properties to work when we map it from Facebook to Twitter.

With the present work, we ask if tie strength has a recognizable signature in social media, perhaps irrespective of a medium's implementation details. This idea, that some CMC phenomena persist across communities and media, has recently emerged in the literature [17]. We believe this is the first work to study a relational one.

Socially-rendered Social Media

Can social media interfaces reflect the social relationships they foster? Primarily, social media interfaces rely on *time* to organize themselves. Twitter, email, IM, IRC and many others all use time as their central design concept. We may call it "social media," but it's not a stretch to instead call it "temporal media."

In the short body of literature on socially-rendered social media, two research projects have laid the groundwork. SNARF [4, 22] and ContactMap [31] each explore using interaction histories to reconfigure interfaces. As these two projects represent the extent of the literature on socially-rendered interfaces, we will discuss them in some detail. (Currently, the Facebook News Feed prioritizes information by some metric. However, its details are proprietary.)

SNARF, the "Social Network and Relationship Finder," is a social sorting prototype designed to solve the email overload problem. SNARF uses features from past email exchanges to visually depict which people are most important to its users. The features include things like "emails sent to each person from the user" and "replies to each person from the user." In total, SNARF extracts 11 features. You can then select any of them as the key upon which the system sorts your email. ContactMap, on the other hand, explores an approach more like our own. Its authors developed hypotheses about things likely to be "important" and perform logistic regressions to predict it from six features.

Our work draws inspiration from these systems and builds on them in main two ways. First and foremost, it works on the open internet. We have as a primary goal testing theory. By putting our work on the open web, we lift sampling frame problems, encounter unexpected contexts and collect enough data to test theory. Second, We Meddle is the first application we know of to put a *relational* model at the heart of its design, as opposed to something defined in a single medium. With a model calibrated against a large relational dataset, we may be able to improve on the relatively low satisfaction scores users gave automatic importance in ContactMap [31].

The system we present in this paper groups Twitter accounts together, something a handful of commercial systems also do. Features like TweetDeck's groups [28] and Facebook's Friend Lists provide ways to group your friends. We have little information on their usage. However, the blog TechCrunch quoted Mark Zuckerberg [27], the founder of Facebook, as saying "Guess what? Nobody wants to make Lists," probably referring to how much time it takes to create a Friend List. Users put friends in lists one by one. The system we present here focuses on subtracting mispredictions, rather than adding each friend individually.

The Collapsed Context Problem

We Meddle attacks the *collapsed context problem.* We often think about it with this analogy: imagine living your whole life at your own wedding. Everyone you know from various parts of your life is there: grandmothers, in-laws, coworkers, cousins, childhood friends, etc. Writing a status update on a social media site is like forgetting you left the microphone on. Everyone hears everything. Consuming content (e.g., reading Twitter or the Newsfeed) is very much like standing in the receiving line. Everyone you know passes by in random order. danah boyd has termed this the *collapsing of context* [1], concentrating mainly on how it affects self-presentation. Do you want to share your latest party pictures with everybody, including the people you didn't invite? But it has other consequences too. In social streams, people from every part of life collapse into one channel, in temporal order, with nothing distinguishing one from any other.

If you want to monopolize your followers' streams, write about what you're eating, seeing or doing every ten minutes. Because of collapsed context, your messages will crowd out others and get more attention. In the real world, and even with varied media [13], we can enforce boundaries: turn on the TV to hear about the movie star; use the phone to talk to your best friend. Today's social streams make this harder.

Facebook predictors	Twitter predictors	β
Days since last comm.	Days since last comm.	-0.587
Days since first comm.	Days since first comm.	0.581
Intimacy x Structural	Intimacy x Structural	0.308
Wall words exchanged	@-reply words exchanged	0.230
Mean *ts* of mutual friends	Mean *ts* of mutual friends	0.198
Educational difference	Follower difference	-0.123
Structural x Structural	Structural x Structural	0.150
Reciprocal x Reciprocal	Links x Links	-0.146
Initiated wall posts	Initiated @-replies	0.112
Inbox thread depth	Direct message headers	-0.105
Number of friends	Following count	-0.105
Soc. Distance x Structural	Soc. Distance x Structural	0.100
Number of applications	no analog	
Wall intimacy words	@-reply intimacy words	0.085

Table 1. The top predictors as measured by standardized betas for the *How strong?* model in [7] and their Twitter analogs. *Follower difference* (gray) substitutes for the original social distance predictor, *Education difference*. It captures a difference in fame on Twitter.

METHOD OVERVIEW & RESEARCH QUESTIONS

To explore the adaptability of predicting tie strength, we built a website called *We Meddle*, open to anyone on the internet who uses Twitter. Users sign in and have the tie strength model applied to their own Twitter accounts. We Meddle is a hybrid, simultaneously an experiment and a system we hoped people would find useful. Via We Meddle, this paper addresses the following research questions.

R1: Does the computational tie strength model in [7], trained on Facebook data, adapt to another social medium?

R2: Can feedback from users improve the model? If so, how does it change?

MAPPING FACEBOOK TO TWITTER

When a user first signs into We Meddle, an agent gathers relational histories from the Twitter API to compute tie strengths. However, first we had to map [7]'s model from Facebook to Twitter. Table 1 presents the mapping: we directly reappropriated the coefficients from [7] for use on Twitter. (The coefficients here differ in absolute terms from [7] because this model only uses the top predictors (not all 74). They are proportionally the same and move in the same directions.) This was one of [7]'s goals: identify predictors that draw on the breadth of Facebook, but have analogs in other media.

In building We Meddle's model, we tried to find the most natural analogs to [7]'s predictors, changing it as little as possible to study how well it adapts. All of We Meddle's predictors are also internally normalized as z-scores and logged—the approach taken by [7].

Predictors

We now visit each We Meddle predictor in turn, discussing what it measures and how it compares to the predictors from [7]'s original Facebook model. Below, + denotes a replication of [7]; ++ refers to a minor deviation from [7], usually an analogy between the two sites; +++ denotes a significant departure from [7]'s original Facebook model.

+ ***Days since last communication*** measures the number of days between the last time two people interacted, accounting for both @-replies and direct messages.

++ ***Days since first communication*** measures the same concept in the same way as the predictor above, but from the first known time two people interacted. In addition, it has access to the order in which a user followed someone, a feature [7] wanted but could not obtain. For example, if I followed you 3rd in my list of 200 followees, but have never communicated with you, this predictor chooses $3 \div 200$.

+ ***Intimacy x Structural*** measures the interaction between [7]'s *Intimacy* and *Structural* dimensions. LIWC intimacy words proxy for the *Intimacy* dimension and *Median strength of mutual friends* proxies for *Structural*.

++ ***@-reply words exchanged*** measures the raw number of words exchanged by the dyad in the form of @-replies. On Twitter, the @-reply is the natural analog of Facebook's wall post: a semi-public, directed message.

+ ***Mean tie strength of mutual friends*** measures the mean tie strength of mutual friends in the ego network. We follow the same iterative regression approach as [7].

+++ ***Follower difference*** measures the difference in follower counts between a We Meddle user and someone they follow. The original Facebook *Educational difference* predictor has no convenient Twitter analog. Instead of omitting it entirely, leaving our model without a representative from [7]'s *Social Distance* dimension, we substitute *Follower difference*. On Twitter, a big difference in follower counts (i.e., orders of magnitude since predictors are logged) is a fame differential. Rather than copy its coefficient directly from *Educational difference*, *Follower difference* derives its coefficient from the average of the four social distance predictors in [7]. This is the only predictor to change its coefficient substantially from the original model.

+ ***Structural x Structural*** measures the multiplicative effect of structure, where *Structural* is again proxied by *Median strength of mutual friends*.

+ ***Links x Links*** measures the multiplicative effect of link-sharing via @-replies. While [7] named its predictor *Reciprocal Services x Reciprocal Services*, [7] also used link-sharing as the underlying proxy variable.

++ ***Initiated @-replies*** measures the number of @-replies initiated by our user. This contrasts with *@-reply words exchanged*, which measures aggregate traffic. By contrast, this predictor is directional, measuring how often our user initiated contact.

++ ***Direct message headers*** measures the number of times a user corresponded with someone via direct messages. (No content was examined.) Direct messages are the Twitter-equivalent of the Facebook inbox: an internal messaging technology very much like email.

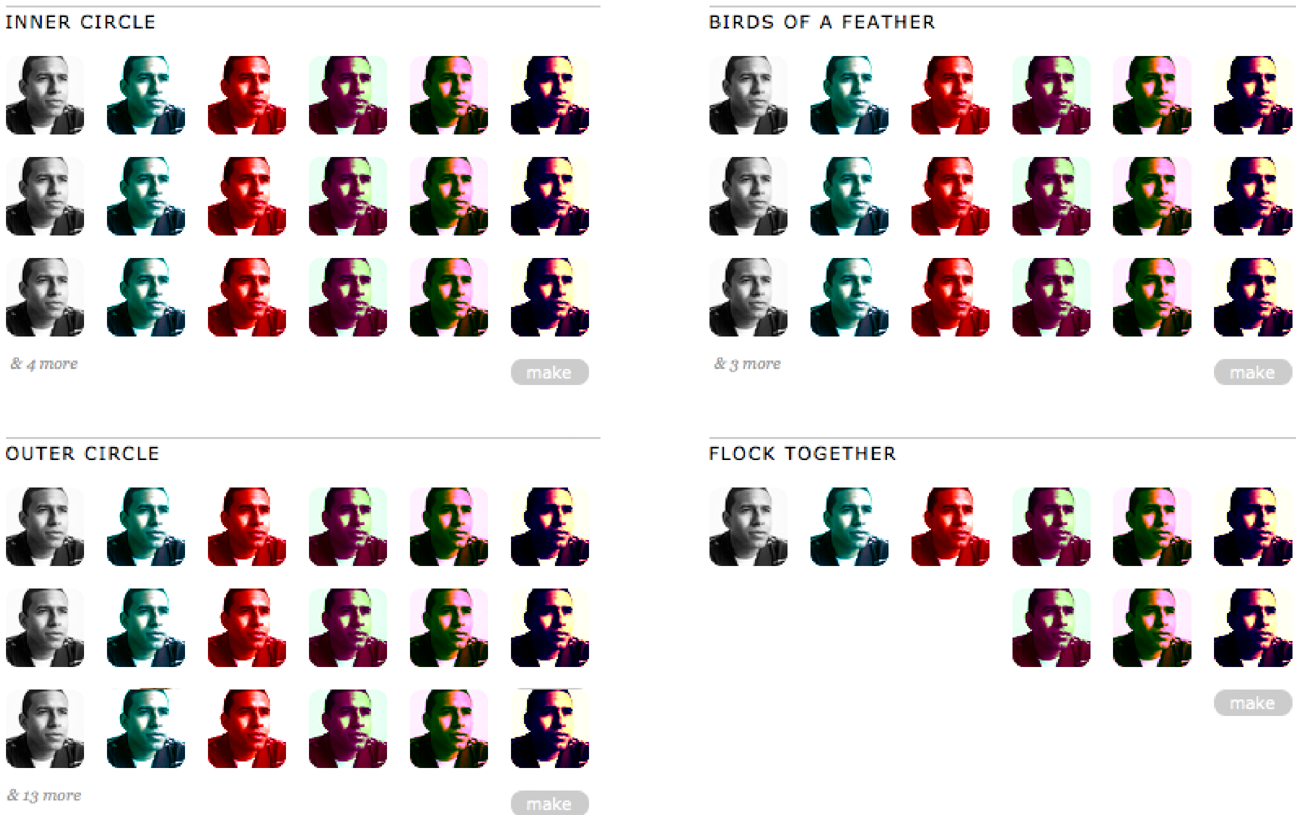

Figure 1. The We Meddle web interface, grouping a user's Twitter contacts via a computational model of tie strength. (This image has been modified to preserve privacy. Normally, different profile pictures appear in each spot.) On first log-in, We Meddle computes tie strengths for everyone the user follows. The "Inner Circle" and the "Outer Circle" correspond to strong ties and weak ties, respectively. We Meddle also computes lists corresponding to social communities in the underlying social network, labeled as "Birds of a Feather" and "Flock Together" here. Users can store these lists inside Twitter and use them in their client.

- ++ ***Following count*** measures how many people a user follows. [7] used *Number of friends* to discount the tie strength of any one friend as the number of friends grew. Since We Meddle computes tie strength over those you follow, this predictor is the natural analog.
- \+ ***Social Distance x Structural*** measures the interaction effect between the *Social Distance* dimension and *Structural*. As before, *Social Distance* is proxied by *Follower difference* and *Structural* is proxied by *Median strength of mutual friends*.
- ++ ***@-reply intimacy words*** measures the number of LIWC intimacy words among @-reply words.

Model Limitations

This is not a perfect mapping. In some cases, we took advantage of data not available to [7]; in others, we had to settle for imprecise analogs, such as fame difference substituting for education difference. Surely, we lose some precision. We aimed for the most natural analogs available, one of [7]'s goals. We hope that later work can resolve some of our inconsistencies. As this is the first in-depth study of how relational features manifest across media, we suggest that the mapping presented here provides first-order insight on an topic we know very little about today.

WE MEDDLE

We Meddle is a web application which uses this model to infer tie strengths between a Twitter user and the people they follow. We Meddle calculates tie strengths in the background, the first time a user logs in. It then generates lists of Twitter accounts, including lists of strong ties and weak ties. We chose to create lists simply because the opportunity presented itself. We moved into the space shortly after Twitter announced the Lists infrastructure, while it was new and there were few competitors. Figure 1 shows the interface (modified to preserve privacy)."Inner Circle" and "Outer Circle" are synonymous with strong and weak ties. We call someone a strong tie when they score one-half standard deviation above the mean, and vice versa for weak ties. This serves to reduce clutter in the interface, making We Meddle usable for users who follow many, many people.

A user can drop someone from a list by clicking on that person's profile picture. When she clicks, the profile picture goes to 25% opacity, holding its place to remind the user that she dropped the account. When she's happy with the list, she can create it, storing the list in Twitter. Storing it in Twitter means that she can access it from any Twitter client, using the list in ways the particular client affords. For instance, the Seesmic client [25] lets you view each list in its own column, meaning that you can slice the conventional Twitter stream into multiple views. A We Meddle user can go to Seesmic and see their Inner Circle flowing in a separate column beside the main stream, using the split view to make sure she does not miss any strong-tie tweets. A short demo video on the site shows users how this could look.

We Meddle takes its inspiration from real life social relationships. In real life, we do not pay everyone equal attention. Returning to the wedding analogy from earlier, imagine yourself standing in the receiving line while everyone in your life comes to talk to you, one by one, in random order. They can talk as long and as often as they want, effectively blocking the people you truly care about. Anyone who already finished can cut back into line anywhere they like. We permit something very similar from our stream clients. We Meddle tries to improve this by making guesses about a user's relationships with everyone they follow, and then lets users slice their stream by those guesses.

Users cannot add people to the lists We Meddle generates. Adding accounts to the We Meddle lists could change their meaning. If a user adds people, can we be sure that Inner Circle still corresponds strong ties? Perhaps the user started from We Meddle's suggestions but branched off to create a list with a different meaning. Limiting users to deletions is an important decision. It allows us to argue that the Inner Circle and Outer Circle retain their meanings. When a user removes an account from the Inner or Outer Circle, we learn where the model makes mistakes. These clicks are crucial data: during the natural process of using We Meddle, users leave a trail from which we study tie strength. (Note that the way ties are classified does not imply that everyone appears as either a strong tie or a weak tie. These two lists do not add up to everyone.)

Communities

When people discuss their social networks, they usually do so in two ways: tie strength and communities [14]. While this paper is about tie strength, We Meddle tries to support this practice. We Meddle uses a community detection algorithm to decompose a user's following network into as many as four social communities. These communities often correspond to groups we can easily name when we see them, like "College Friends," "Former Colleagues" and "Researchers." Users can store these lists in Twitter, too. We Meddle uses a freely available community detection algorithm, called the Markov Cluster Algorithm [30], to generate them. The algorithm does random walks of a network, noting that nodes within a community have more paths between one another than nodes in different communities. The algorithm then works to optimize modularity [23]. While not central to this paper, we built the communities feature to give users another reason to visit We Meddle.

Architecture

The We Meddle tie strength engine is written in Perl, appropriating the output of an R statistical model. When a user first signs in to We Meddle, the system builds a database of tie strengths for each account the user follows. The sign-in forks off hundreds and sometimes thousands of API requests against Twitter.

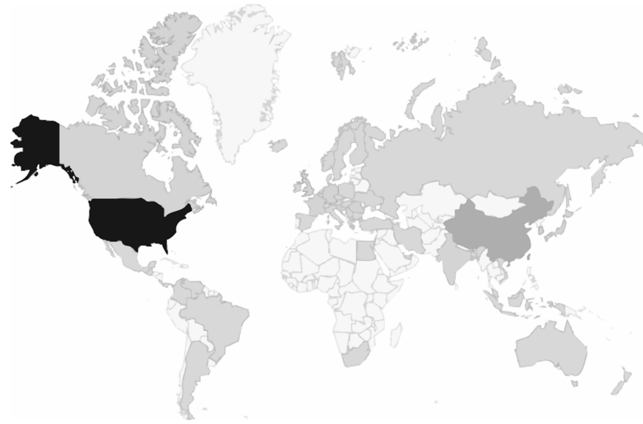

Figure 2. A heatmap of We Meddle's 2,114 users spread over their 52 countries. Most users come from the United States, but China also has a strong presence, accounting for about 10% of We Meddle's users.

As the API requests come back, they first filter through the non-structural parts of the tie strength model. When they all come back, the Perl-based model percolates the non-structural tie strengths through the user's network. People often formulate this as an eigenvector centrality problem, but in practice a simple percolation loop with a few iterations seems to always converge (and consumes far fewer resources). Near the end of the tie strength computation, the engine projects all tie strengths onto a $[0, 1]$ interval by mapping the mean to [7]'s mean and capping at the ends of the interval. The web interfaces seen by We Meddle users are written in PHP and Javascript, making use of the jQuery [15] toolkit for animations and asynchronous communication.

Deployment

In January 2010, we made We Meddle open to any Twitter user on the web. We Meddle was available at `http://wemeddle.com`. After announcing it on a few mailing lists and through our Twitter accounts, the site spread by word of mouth. We considered doing a lab study: recruit Twitter users from around campus to tell us whether We Meddle's guesses match how they feel about their networks. But, we decided on an open web deployment for two reasons. First, releasing We Meddle on the web means people can choose to use it. Voluntary usage is good in our view. Second, by lifting the college campus sampling frame, we have more confidence in the results that come from the present study. We Meddle has seen users from all over the world—something we never could have replicated in the lab. Figure 2 illustrates We Meddle's user distribution by country. At the time of this writing, 2,114 people from 52 countries have used We Meddle.

At the same time, We Meddle largely spread by word of mouth, starting with our personal networks. This is called snowball sampling, and some research traditions see it as problematic. This does introduce complications. Yet, this is primarily how social applications spread on the web. Putting it online also means it reached much further than we ever could come in the lab. Since We Meddle spread relatively wide, less than 5% of its users were our Twitter followers or their followers (i.e., 2 hops away). We argue that this reduces snowball sample validity threats.

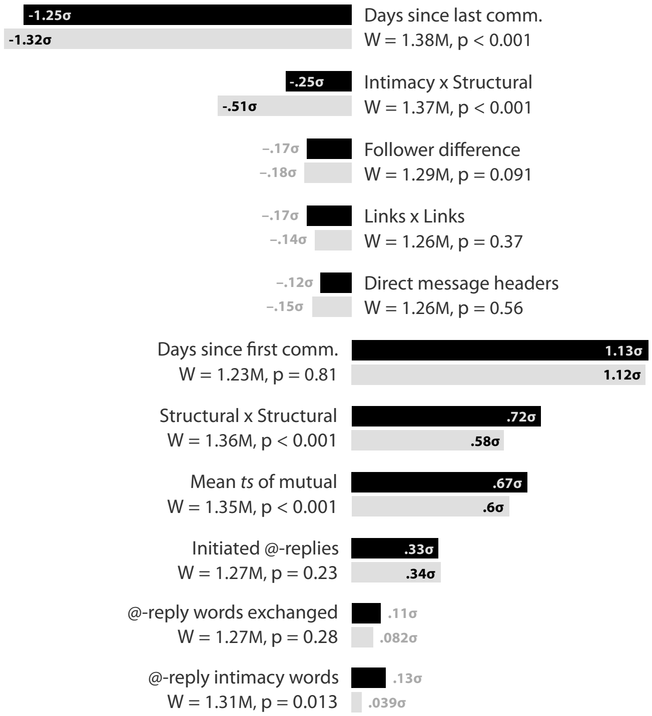

Figure 3. An analysis of We Meddle's *strong tie* mistakes in terms of the model's input predictors. Each pair of bars compares two groups, correct predictions (black) and mistakes (gray). At the end of each bar is the within-user, standardized median for the predictor. *W* refers to the Wilcoxon statistic. Note the recency effect indicated by *Days since last communication* and many strong effects by network-based predictors: *Intimacy x Structural, Mean strength of mutual friends,* and *Structural x Structural.* These highlight differences between Twitter and Facebook.

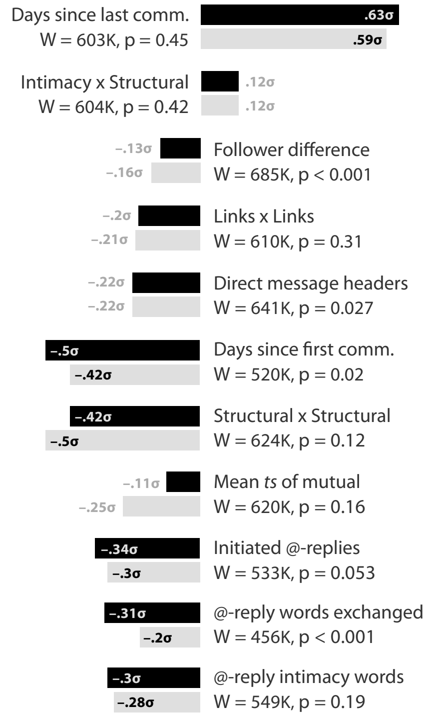

Figure 4. An analysis of We Meddle's *weak tie* mistakes in terms of the model's input predictors. Colors and statistics have the same meaning as in the figure to the left. We see very different reasons for weak tie mistakes than we do for strong tie mistakes. The network-based predictors vanish. Instead, we see strong effects by *Follower difference* and *@-reply words exchanged,* perhaps signaling the ease with which you can message a higher-status user on Twitter.

We Meddle's Limitations

We Meddle allows users to perform a first-pass sort of the people they follow on Twitter. One limitation is that the interface does not give users the ability to create lists with finely tuned meanings, such as the intersection of social circles. For example, perhaps someone would like a research circle only composed of people they feel close to (i.e. strong ties). While We Meddle takes a step toward enabling better consumption of social streams, we hope that future work can improve on some of these design limitations. This seems especially relevant and timely given the rise of Google+.

GENERALIZATION

When a user drops someone from the Inner Circle or Outer Circle, we can infer that the model made a mistake. Whether the model generalizes hinges on the question "How many mistakes did We Meddle make?" We argue that the tie strength model generalizes when we see a comparable error rate to the one seen in [7]. In other words, it generalizes to Twitter if it gets about 87–88% of its predictions right. We Meddle received 1,105 corrections from 236 different users. Most We Meddle users made no corrections. We could view this as a huge success: the majority of users experienced complete and utter success with We Meddle. Of course, this is a blindly optimistic interpretation. Some of them probably thought We Meddle got everything right. (Some qualitative data support this.) However, other people probably did not realize they could correct We Meddle (qualitative data also support this), or felt it wasn't worth the effort to make the clicks. Because We Meddle straddles the boundary between an experiment and a system people want to use, we have to correct for issues like these.

First, we consider only data from the 236 We Meddle users who made at least one correction and at least one list, putting aside users who did not. This method leaves out anyone who felt We Meddle's predictions closely matched reality and did not make corrections. (Several hundred people made lists without a correction.) But, we know these 236 users understood the process by which they could correct We Meddle and subsequently made a list. We cannot guarantee that these users corrected *all* of We Meddle's mistakes, but remember that they ultimately created a list for their own use within a Twitter client. They had every incentive we could provide—short of forcing them—to make the tie strength lists they wanted. We Meddle observed 27,529 relationships from these 236 users. Of these, users had a chance to correct accounts that We Meddle marked as a strong tie or a weak tie, a total of 14,075. So, users corrected 1,105 out of 14,075 relationships, or 7.85%.

However, as explained in the previous section, users had no way to tell We Meddle that it had forgotten someone: they could only drop people from lists. Since We Meddle is as likely to underestimate tie strength as it is to overestimate it (see Figure 4 in [7]), we double this percentage, obtaining an upper bound of 15.7%. While slightly higher, it closely resembles [7]'s error rate of 12–13%, especially considering that 15.7% is an upper bound. ([7]'s error rate was based on the difference between model predictions and participant-provided scores; We Meddle's user corrections are a direct analog.) Primarily from this number, but with support from qualitative data presented next, we find evidence that [7]'s tie strength model largely generalizes to a new social medium.

Mistakes in Terms of Predictors

Next, we analyze how We Meddle's mistakes express themselves in terms of the model's input predictors. This can tell us in what subtle ways the Twitter tie strength model differs from the Facebook model, and where we should look to improve it. Figures 3 and 4 summarize the results. For example, Figure 3 shows that true strong ties have a lower *Days since last communication* predictor than mistakes, Wilcoxon $W = 1.38\text{M}$, $p < 0.001$. (Use of Wilcoxon tests reflects the non-normality in the data.) We within-user standardize every predictor. That is, we compare numbers standardized against all relationships corresponding to a user. As the same model generated both the correct predictions and the mistakes, differences stem from differences between correct predictions and mistakes, not artifacts of the model.

What jumps out most is the contrast between strong and weak tie mistakes, particularly the role of the network in strong tie mistakes. Three network predictors have large effects: *Intimacy x Structural* (0.263 standard deviations), *Structural x Structural* (0.134 standard deviations) and *Mean strength of mutual friends* (0.073 standard deviations). Interestingly, something outside the model itself suggests why. The strong tie mistakes disproportionately belong to big clusters, evidenced by their membership in the lists generated by community detection on the underlying social network, $\chi^2 = 37.43$, $p < 0.001$. Recall that the model blends tie strength recursively through the network: it is a function of the tie strengths of mutual friends. Due to the summary statistics over the cluster, this means that many relationships benefit from a single strong tie in a big cluster. These network predictors make a strong case for a more refined view of the network in the tie strength model, something we discuss in more detail later.

Weak tie mistakes, on the other hand, express themselves most definitively in *@-reply words exchanged,* with correct predictions lower than mistakes. Perhaps this stems from how easily you can message someone you do not know on Twitter. Whereas on Facebook someone must confirm my friend request before we can exchange messages, on Twitter I can send messages to President Obama if I like.

HOW USERS EXPERIENCED WE MEDDLE

The quantitative data address the generalizability question, but they do not tell us how users felt toward We Meddle. To understand it, we conducted interviews with our users. We picked We Meddle users at random from the logs and @-mentioned them on Twitter. After they replied, we conducted eight semi-structured interviews in whatever medium they preferred (e.g., phone, IM or email). (This is a typical response rate and sample size for follow-up interviews.) The participants ranged widely in backgrounds and in how they used Twitter, from young coders to small business owners

Figure 5. A comment on We Meddle via Twitter. The user writes, "This thing is miraculous. I really want to know how it works. It even classifies my GF's account into Inner Circle, but I never send tweets to her account, and it also classifies accounts I don't care about into outer circle. The four categories are very accurate, almost 100%."

who primarily used Twitter for promotional reasons. We Meddle also received many hundreds of comments via Twitter, and we present a selection of them at the end of this section. The point of the interviews was to elicit feedback from people who did not say anything on the web, hopefully removing the inherent self-selection bias that comes from speaking up publicly. The interviews took about 30 minutes. Email interviews consisted of questions similar to IM and phone interviews, but we often had follow-up conversations to clarify points. Typically, participants logged into We Meddle to look at the lists while we talked.

When asked about the composition of the tie strength lists, participants reflected on the lists' accuracy and the tie strength concept in their lives.

> [Did the lists reflect your real social life?] Um, I was pretty amazed to tell you the truth. Really amazed cause, um, the one I had an extreme [sic] hard time with trying to figure out was the Outer list. And this was probably the same with most people cause they're not people you communicate with much. So, I only remember one person I actually recognized on the Outer list. But the other three [Inner Circle, two communities] were pretty close to right on.

> It's actually kind of fun to look at the Inner Circle and say, "Wow, look at that person. I haven't talked with him in a long time, but they totally fit there." Yeah, there's some of those in here.

> It's kind of astounding and scary how good it is. [The people it chose for your Inner Circle, did they fit?] I would say the Inner Circle is about 70% accurate, maybe 80%. And it does actually a fairly—well, it's such an interesting question, right? So, it includes my wife, which is *good.* And people like [name] who runs [my online community], which is *good.* And my baby blogs, but doesn't tweet. But mostly what it is, is the Inner Circle is a sort of a blend of my immediate personal universe and probably my two most important social universes.

The last statement captures the way scholars often talk about tie strength: it skims off people from different social circles. In some cases, users expressed surprise when We Meddle correctly identified certain people in their lives.

> A few of the . . . well I remember a few of the people in the Inner Circle are actually relatives, and that was pretty cool. I didn't expect that.

> It's interesting that it actually placed my girlfriend four rows down versus at the very top, where I would expect her to be. [But she's in the list?] She's in the list, yeah absolutely, most of the lists actually. I hope she doesn't see the four rows down part. [participant laughs]

However, we did hear about problems. For instance, two users discussed relational contexts which fell outside We Meddle's built-in assumptions.

> Some people on Twitter just say stupid things. Or, they might say something that's inaccurate. So there's a few people I see who are not in my Inner Circle, you know, my group . . . people I actually hang out with. But we have had disagreements on Twitter. We argued.

> The Inner Circle is actually not super accurate. Yeah, the Inner Circle is basically all the people I used to work with . . . But I talk with them, sort of irregularly, now that I'm not at [company] anymore. [How close were you when you worked there?] Yeah, I was close to them while I was there, so it's fun to see them here from an old job. Maybe there should be a category for them: like people you used to be close to, but you know, aren't anymore.

Relationships can be intensely *negative.* (In fact, we received a handful of mildly annoyed emails from We Meddle users wondering why their ex-partner appeared in the Inner Circle.) The tie strength model does not understand these relationships. The former colleague story shows how a *biographical break* can influence someone's viewpoint. This participant felt close to his co-workers one month, and then did not the next. [7] saw similar problems, and future work could do better by trying to resolve them. However, a perhaps strange way to see these problems is as a sign of generalization: the model makes the same kind of mistakes as [7].

Reactions via Twitter

We Meddle received hundreds of unprompted comments via Twitter. In addition to Figure 5, we include a selection of these tweets here.

> wemeddle.com is a very cool idea for making twitter lists. It was good enough to re-create lists I made myself! Worth checking out.

> Fascinating clustering of those you follow by strong and weak ties.

> Oh geez, this application is a great filter for your Twitter contacts. [translated from Romanian]

> trying we meddle. [. . . six minutes later . . .] didn't like we meddle.

> Liking what the folks at We Meddle are doing with the @TwitterAPI!

> Easiest Twitter list maker from @wemeddle.

> @[previous tweet] Just did and guess what, you are in my inner circle.

> If you are too lazy to take the time to make lists, we now have a new choice. We Meddle automatically analyzes your friends to generate several lists, which you can then make. [translated from Japanese]

> We Meddle is really good. It automatically divides my friends into groups, so that occasional tweets from my important friends won't be buried in an ocean of other tweets. [translated from Chinese]

These tweets speak to the value users found in We Meddle. Each tweet originally included a link to direct people to the site, which we edited out in the interest of brevity. We see this willingness on the part of our users to spread the site as evidence of its value to them. We Meddle did not have any prompt anywhere asking users to share the site with others.

DISCUSSION

From our data, we arrive at early evidence that [7]'s predictive tie strength model largely generalizes to a new social medium. The error structure, based on the intersection of quantitative and qualitative data, resembles what we see in [7].

If the model aligned precisely with social dynamics in Twitter, Figures 3 and 4 would show no effects for any predictors. However, the error analysis reveals wrinkles in the tie strength model. Remember, though, that these mistakes represent at most 15.7% of the data; their effect on an updated model would be subtle. Some predictors from Figures 3 and 4 probably reflect Twitter itself. For instance, the part played by *Days since last communication* in strong tie mistakes suggests that Twitter is more recency-driven than Facebook: tie strength seems to decay faster there. *@-reply words exchanged* in strong tie mistakes may suggest that people need to use more evocative language to maintain relationships in Twitter's lean medium. The role of *@-reply words exchanged* in weak tie mistakes probably reflects non-reciprocal ties in Twitter. [7]'s model learned a social structure in which a handshake precedes any communication. In Facebook, we have to accept friend requests; Twitter usually has no barrier.

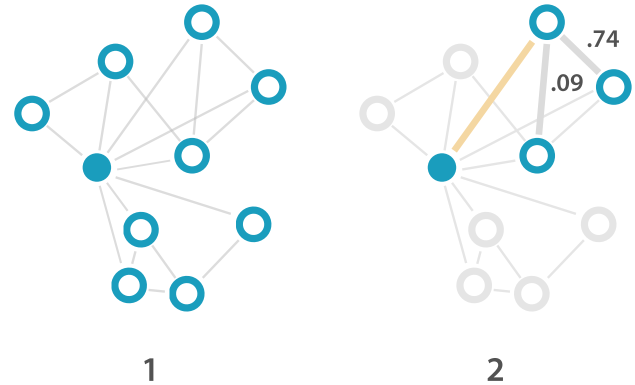

Figure 6. A way to update the network component of the tie strength model. 1) A user's ego network, including links between mutual friends. 2) We want to estimate the tie strength of the link in gold. Instead of weighting the two mutual tie strengths equally, we weight them by low-fi tie strengths on the cross-cutting links, 0.09 and 0.74 in this example. We think it could reduce noise associated with big clusters; however, the extra crawl costs time.

We also find that intensely negative relationships often confound the model. [7] indicts "asymmetric friendships," friendships with big power differentials. How can we account for these relationships? This remains an unanswered and attractive target for future research. We tentatively propose a direction: measure politeness. We could measure it textually (e.g., hedges, deference, formal greetings, etc.) or via inter-message response time. For example, we know that in corporate contexts, upper-level management often lets messages sit for long periods of time before responding [24]. Similar temporal dynamics may happen in everyday social media. Politeness may signify asymmetry; lack of politeness coupled with *Intensity* may signify negative relationships.

A New View of the Network

The network predictors truly stand out in Figure 3. In [7]'s model, every mutual friend contributes equally to the network part of the tie strength model. But, why should my wife's sister contribute to my best friend's tie strength, if they only felt obligated to become friends in the first place? Instead, we propose a new, weighted network model, one where each mutual tie has its own tie strength. Practically speaking, this is easier said than done. Not only would we have to estimate every ego-centric tie (as We Meddle does), but also every cross-cutting tie between alters. In other words, not only would the model have to estimate tie strength for you and all our mutual friends, but also *between* you and all our mutual friends. To operate in web-response time, this is probably intractable; plus, from the outside we simply cannot see everything every dyad does.

Instead, consider low-fi tie strengths on the ties between alters. Draw a few roughly orthogonal tie strength predictors: we could pick *Days since last communication* and *@-reply intimacy words*. The model could use tie strengths made only from these two predictors to weight contributions to the tie strength of interest. Figure 6 illustrates the idea. This would of course require deeper knowledge of the communication record (e.g., being inside Twitter) or extra crawls at prediction time. But, as we see in Figures 3 and 4, it's the easiest way to bring down the error rate. With it, perhaps 90-95% prediction accuracy is within reach.

Theoretical Implications

This paper looks at *one* new medium. We see this as a substantial step for the study of social media and CMC: we believe this is the first work to quantitatively study how an important property of relationships manifests in two social media. However, many others remain. It is an open question how or even if tie strength can be reconstructed in all of them. Does it work in email? Does it work in IM? Perhaps it will need modifications. What modifications? Future work may find traction with the approach presented here: offer [7]'s model in the new medium; let your users tune it via feedback.

Our findings suggest that some important properties of online relationships resist sites' implementation details. This is important not only because we have so many social media, but because they change so often. Facebook has changed, perhaps substantially, since [7]'s study in 2008. Among many changes, Facebook has added comments directly on statuses, photos and videos. In 2008, everything happened threadlessly on the Wall. Maybe practices have changed in response. But, how much? [7]'s model worked for We Meddle, built a year and a half later. We see this as a deep question for future study: Do some fundamental properties of online relationships manifest the same way, regardless of design?

We have a vast literature on how tie strength modulates all kinds of social phenomena, from financial trading to the spread of values to the cohesion of groups. Consequently, we see opportunity to study other things by predicting tie strength. For example, what mix of ties keeps users on sites the longest? Mostly strong? A core of strong, but otherwise mostly weak? Do certain mixes of ties promote prosocial behavior online? Answers would have profound consequences for how we see online communities and how owners operate them. Or, instead of asking questions *about* social media, we might ask questions that simply use social media as a setting. For example, we could re-ask Fowler & Christakis's question [5]: Does tie strength modulate happiness in online networks?

Design Implications

In this paper, we take a step toward showing that computing tie strength helps users deal with the consumption side of the collapsed context problem. However, if we controlled all of Twitter, we would not have picked the consumption side of the collapsed context problem. We would have picked the production side: imagine sending messages only to strong tie colleagues (e.g., the intersection of the Inner Circle and a community list) without having to work at generating those lists yourself. We hope that our We Meddle case study encourages designers to experiment with tie strength. Whereas [7] briefly sketches many ways tie strength could inform design, here we conclude by drawing a longer scenario showcasing tie strength.

Imagine a woman interacting with her friends and family via a social network site. She posts photos, talks about her job, her family life and how night classes are coming along. And she uses it to keep up with everybody's busy lives. Now let's imagine that when she next logs in, she's been on vacation for two weeks with limited internet access. She wants to catch up. What should the system show her? All 5,000 things that happened?

In addition to topics she cares about, we might want to show her what happened to the most important people. Maybe her best friend changed jobs, or her sister took some great photos of her vacation. Maybe she missed a few strong tie birthdays and can sneak in belated congratulations. Perhaps a weak tie posted an inspiring web video that got everybody talking. We think understanding tie strength is the first step to building systems that can do these things.

LIMITATIONS

We used an imperfect mapping to build We Meddle, a system that could not support deep fine-tuning of social circles. We see this as a first step in an area we know very little about today. At the same time, we hope future work can improve upon the precision and design flexibility of what we present here. Systems that flexibly understand social life stand a chance to improve a wide range of social media.

CONCLUSIONS

In this paper, we find early evidence that [7]'s predictive tie strength model generalizes to a new social medium, one in which it did not train. This was a critical step: a model that works only in Facebook has little value outside that site. Perhaps most importantly, our findings suggest that a core property of online relationships may manifest similarly across social media. We see this as an important new step for social media and CMC theory.

Via We Meddle, we also show how computing tie strength can help users cope with the collapsed context problem. We find this encouraging and hope to see practitioners exploring new interfaces which incorporate tie strength. Perhaps we do not have to think linearly and temporally: we could bubble old but important messages to the top of a linear interface using tie strength. By intersecting tie strength and community detection, perhaps we might make a dent in social media's privacy problem.

We believe this is the first work to study in-depth how something fundamental about online relationships manifests *across* social media. But, we also think this work raises as many questions as it answers. Does tie strength continue to manifest this way in other social media, like email and IM? Do other core aspects of relationships manifest in their own ways across media?

ACKNOWLEDGEMENTS

We would like to thank the Social Spaces group at Illinois, as well as the comp.social lab at Georgia Tech. Amy Bruckman, Sarita Yardi and Ethan Zuckerman contributed valuable feedback on We Meddle's design. Google helped support this work. We also thank our users for giving We Meddle a try, and for chatting with us when we asked.

REFERENCES

1. d. boyd. Faceted id/entity: Managing representation in a digital world. Master's thesis, MIT, 2002.
2. R. S. Burt. Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399, 2004.
3. M. De Choudhury, W. Mason, J. Hofman, and D. Watts. Inferring relevant social networks from interpersonal communication. In *Proc. WWW*, pages 301–310, 2010.
4. D. Fisher, B. Hogan, A. J. Brush, M. Smith, and A. Jacobs. Using social sorting to enhance email management. In *Human-Computer Interaction Consortium (HCIC'06), Fraser, Colorodo*, 2006.
5. J. H. Fowler and N. A. Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *BMJ*, 337:a2338+, 2008.
6. N. E. Friedkin. A test of structural features of granovetter's strength of weak ties theory. *Social Networks*, 2:411–422, 1980.
7. E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proc. CHI*, pages 211–220, 2009.
8. E. Gilbert, K. Karahalios, and C. Sandvig. The network in the garden: an empirical analysis of social media in rural life. In *Proc. CHI*, pages 1603–1612, 2008.
9. Google Plus. http://plus.google.com.
10. Google Scholar. http://scholar.google.com/scholar?q=strength+weak+ties.
11. M. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.
12. M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233, 1983.
13. C. Haythornthwaite and B. Wellman. Work, friendship, and media use for information exchange in a networked organization. *J. Am. Soc. Inf. Sci.*, 49(12):1101–1114, October 1998.
14. B. J. Hogan. *Networking in Everyday Life*. PhD thesis, University of Toronto, 2009.
15. jQuery. http://jquery.com.
16. I. Kahanda and J. Neville. Using transactional information to predict link strength in online social networks. In *Proc. ICWSM*, 2009.
17. A. Kittur and R. Kraut. Beyond wikipedia: coordination and conflict in online production groups. In *Proc. CSCW*, pages 215–224, 2010.
18. N. Lin, P. Dayton, and P. Greenwald. Analyzing the instrumental use of relations in the context of social structure. *Sociological Methods & Research*, 7(2):149, 1978.
19. P. V. Marsden. Core discussion networks of americans. *American Sociological Review*, 52(1):122–131, 1987.
20. M. Mcpherson, L. S. Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.
21. S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.
22. C. Neustaedter, A. Brush, M. Smith, and D. Fisher. The social network and relationship finder: Social sorting for email triage. In *Proc. CEAS*, 2005.
23. M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
24. D. Owens, M. Neale, and R. Sutton. Technologies of status management: status dynamics in e-mail communications. *Research on Managing Groups and Teams*, 3:205–230, 2000.
25. Seesmic. http://seesmic.com/app.
26. X. Shi, L. A. Adamic, and M. J. Strauss. Networks of strong ties. *Physica A: Statistical Mechanics and its Applications*, 378(1):33–47, May 2007.
27. TechCrunch. http://tcrn.ch/cuzaqp.
28. TweetDeck. http://www.tweetdeck.com.
29. B. Uzzi. Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing. *American Sociological Review*, 64(4):481–505, 1999.
30. S. van Dongen. *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, 2000.
31. S. Whittaker, Q. Jones, B. Nardi, M. Creech, L. Terveen, E. Isaacs, and J. Hainsworth. Contactmap: Organizing communication in a social desktop. *ACM TOCHI*, 11(4):445–471, 2004.
32. A. Wu, J. M. DiMicco, and D. R. Millen. Detecting professional versus personal closeness using an enterprise social network site. In *Proc. CHI*, pages 1955–1964, 2010.
33. S. Wuchty. What is a social tie? *Proceedings of the National Academy of Sciences*, 106(36):15099–15100, September 2009.
34. R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proc. WWW*, pages 981–990, 2010.