Designing Social Translucence Over Social Networks

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

ABSTRACT
Social translucence is a landmark theory in social computing. Modeled on physical life, it guides designers toward elegant social technologies. However, we argue that it breaks down over modern social network sites because social networks resist its physical metaphors. In this paper, we build theory relating social translucence to social network structure. To explore this idea, we built a tool called Link Different. Link Different addresses a structural awareness problem by letting users know how many of their Twitter followers already a saw link via someone else they follow. During two months on the web, nearly 150K people used the site a total of 1.3M times. Its widespread, viral use suggests that people want social translucence, but network structure gets in the way. We conclude the paper by illustrating new design problems that lie at the intersection of social translucence and other unexplored network structures.

Author Keywords
social networks, social translucence, design, cmc, twitter

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

INTRODUCTION
Social translucence [22] is a landmark theory in social computing. In essence, it describes ways to build social technologies that support social life. Social translucence argues that we should make online social behavior *visible* to facilitate *awareness*, ultimately creating social spaces where we feel *accountable* to one another. Erickson & Kellogg model the theory on social processes we see in real, everyday life. For example, they share a story where authors gather in a room to arrange the chapters of an edited book ([22], pgs. 62–63). It works because the room affords visibility: if I try to move a chapter somewhere that upsets the book's delicate balance, you can intervene because you see me moving it. Typically, we have to carefully design social media to afford what this room affords effortlessly.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

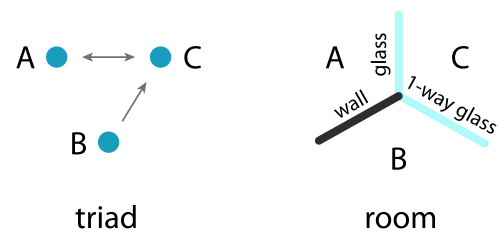


Figure 1. A hypothetical room mirroring a social media triad.

Early social media resembled real-life places like this room. Chat clients evoke conference rooms. Forums recall café bulletin boards. However, today many of the internet's most popular sites (e.g., Facebook, Twitter, Google+) don't look like rooms at all. Rather, they revolve around the concept of social networks. To illustrate, consider an ordinary group of three people on Twitter. We've depicted their relationships in Figure 1. A and C follow each other, and B follows C. Now, picture the room you would have to build to mirror Twitter's constraints. You need to keep A and B separate, so you might build a wall between them. B should see C, but not the other way around, so put one-way glass between those two. At first glance, it seems like you don't need anything between A and C, but you should keep them separate so A doesn't wander over, permitting B to see A.

This room is unlike anything you would find in everyday life. And it's just three people. What would it look like with four, five or six? It becomes hard to imagine such a room. The complexity grows so quickly because social networks disrupt sight lines in complex ways. Consequently, the metaphors around physical space so central to social translucence simply break down in the context of social networks.

This paper aims to bridge the gap, offering a new way to think about social translucence in the era of networks. We build a theory relating specific structures within complex social networks to the problems they cause for social translucence. Using it, we work through an awareness problem arising from one of these structures, the triad $A \leftarrow B \rightarrow C$. Here, A has difficulty sharing new information with B because she cannot know what B already heard from C. We designed a tool called Link Different to address this problem on Twitter. When someone wants to share a link with their followers, Link Different reports how many of them already saw the link via someone else they follow. Deployed openly on the web for two months, nearly 150K people used Link Different 1.3M+ times in total. We think these orders of magnitude speak volumes: people still want social translucence's cues, but network structure gets in the way.

We begin by reviewing social translucence and its descendent systems, as well foundational work on social networks. Next, we introduce Link Different and the sharing behavior of its users. Finally, we conclude the paper by extending the theoretical work behind Link Different, arguing that new social computing problems lie at the intersection of social translucence and various unexplored network structures.

LITERATURE REVIEW

First, we review the theory of social translucence. We also consider some of the many systems architected around it. Because our approach connects social translucence with social network structures, we conclude this section with a summary of fundamental social network research.

We focus on social translucence because it provides specific design guidance that we can overlay on social networks. However, theoretical alternatives exist. For example, the venerable CHI paper "Beyond Being There" [34] argues that we shouldn't try to replicate the affordances of real life in social technologies—a central tenet of social translucence. This tension deserves resolution, with new work perhaps arising by adopting an alternative theoretical lens.

Social Translucence

Comprising three major papers [21, 22, 23], the theory of social translucence aims to take what we do so naturally in real social life and map it onto online social media. (However, the term "social media" didn't exist at the time and doesn't appear in those papers.) Erickson & Kellogg write:

> The difficulty of digital communication and collaboration stands in stark contrast to our ability to communicate and collaborate with one another in the physical world . . . We have evolved an exquisite sensitivity to the actions and interactions of others. Whether it is wrapping up a talk when the audience starts fidgeting, or deciding to forego the grocery shopping because the parking lot is jammed, social information like this provides the basis for inferences, planning, and coordination of activity. [22] (p. 60)

(Hereafter, we refer to Erickson & Kellogg as "E&K.") This is the yardstick against which E&K argue we should measure our social technologies. Despite being published more than ten years ago, we can see technologies still catching up with the vision. For example, the recently launched startup Localmind lets users ask "how crowded the bar is, or how long the line to the club is" from afar [42]. Presumably, you might choose to stay home if the club line is too long, recalling E&K's jammed grocery store parking lot.

Unless explicitly designed into systems, most social information remains hidden inside databases. E&K propose three guiding principles for bringing it to the forefront, which we referred to at the beginning of the paper: *visibility*, *awareness* and *accountability*. Each gives rise to the next, and visibility is the most basic. Social translucence argues that we should make collective activity visible to everyone in a space. Of course, privacy is an issue, and this is where the word "translucence" takes on particular importance. Rather than build a totally transparent system (e.g., publish GPS coordinates), a system should abstract away from raw data without sacrificing too much useful social information (e.g., publish neighborhoods). In addition, constraints act on visibility. For example, our ability to see things decays as we move farther away from them. Social software can model this, for instance, by providing greater resolution to people nearby.

Awareness arises from visibility. Having seen the social cues floating around me, I absorb them, and use them to decide what to do next. Furthermore, my social and cultural expectations come to bear on my newfound awareness. For example, suppose I see you methodically cleaning up files in our shared code repository. I become *aware* that soon you will reach my files, and social customs dictate that I release my lock on the files so you can do your work. Finally, out of this distributed social awareness comes a second-order effect: accountability. As E&K put it, "I know that you know that I know."

As we alluded earlier, social translucence has been very influential in social computing and CSCW, attracting over 1,000 citations since its publication a decade ago [30]. Across a wide variety of application areas, many systems have used it to guide design (e.g., [7, 9, 11, 24, 36, 46, 47]). Most work has focused on small groups. For example, [11], [24] and [36] use social translucence to abstract away from raw presence data in experimental IM clients. In [7], the theory provides a way to visualize abstract histories of presence among distributed workgroups. A notable exception to the small-group trend is WikiDashboard [46], which employed social translucence to make behind-the-scenes Wikipedia activity visible.

Particularly relevant to the present work, [9] built and field-tested a "directed content sharing" tool called FeedMe. While browsing RSS feeds in Google Reader, FeedMe recommends people who might also be interested in an article and lets you to share it with them via email. FeedMe uses social translucence to show you how many other FeedMe recommendations the target person has received. We take FeedMe as a point of departure. Where FeedMe allowed point-to-point sharing via email, we ramp the idea up to social media at scale on Twitter. To the best of our knowledge, the system we present in this paper is the first to connect social translucence and large-scale social networks.

Social Networks

Studied by sociologists for decades [38], social networks silently structure social life. The literature is simply too vast to cover here, but we attempt a very brief overview of fundamental work. Social networks start out simply: I become friends with you, you become friends with someone else, and so on. Friendships often form by way of homophily (i.e., shared interests and tastes) [43] and a process called preferential attachment [5], whereby people prefer to connect with those who are already well-connected. Over time, a complex, organic structure emerges. Via a network's macroscopic structural properties and dynamics, you can tell who is popular [6], who will have good ideas [17] and who in a company will get promoted [16].

It is important to note that social networks differ from *social network sites* [13]. Social networks are a concept. Social network sites (SNSs), on the other hand, revolve around social networks, using them as a central design element. SNSs

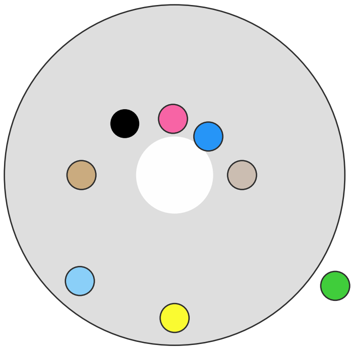


Figure 2. Babble's "social proxy," used to signal activity in a chat room. Avatars, represented by simple abstract circles, move to the center to signal interest in chatting. Reproduced from [21].

like Twitter, Facebook and Google+ are important: they can provide us with new social capital [20], answers to questions [44], and sometimes a sense of well-being [15].

We add to this growing literature by considering the design of SNSs as a function of embedded social network structures. We believe this is the first work to do this. In the next section, we consider how traditional social translucence cues interact with micro-structures called triads.

SOCIAL TRANSLUCENCE & SOCIAL NETWORKS

E&K often draw analogies between physical space and digital space. Consider, for example, the system discussed by E&K in *Social Translucence*, the chat program Babble. In Babble, we all enter the same digital space to chat. Your avatar moves to the center of the room to signal your interest in talking, and the geometry of Babble's circular design forces it into proximity with other avatars (see Figure 2). The parallels between physical space and online conversational space are direct and apparent. E&K explicitly connect digital social spaces with physical ones via an architecture metaphor:

> As designers of communication and collaboration systems, we find ourselves taking inspiration from work in the areas of architecture and urban design. This is not surprising, since, like architects and urban designers, we are concerned with creating contexts that support various forms of human-human interaction. [22] (p. 61)

As we discussed before, early social media resembled physical spaces. Chat rooms can be thought of as digital conference rooms—the insight made by Babble. You could argue that email distribution lists resemble neighborhood association meetings and therefore build affordances that mimic those meetings. In real life, as in these technological metaphors, social signals are tied to the space. For example, in real life, if I move, then everybody nearby notices me moving. If someone addresses our group, but I'm idly looking out the window, then everybody else can infer that I'm bored. The physics of everyday life affords these things.

However, modern social media—especially some of the most popular sites—often do not reflect everyday physical life. Rather, they rest on the primacy of social networks. Social networks do not have physical analogs. On Twitter, you follow diverse people who share things you find interesting. Those people follow who they want, and so on. Social signals—the

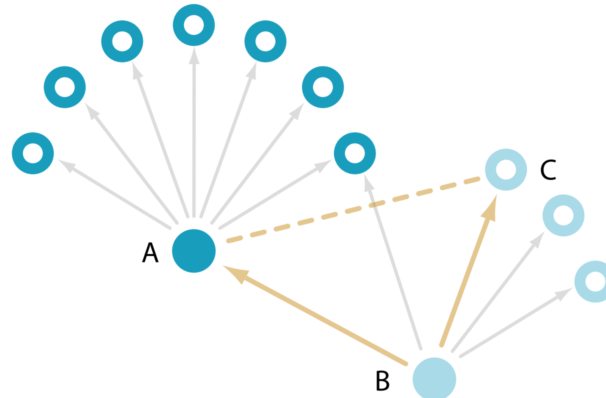


Figure 3. A fragment of a directed social network like the attention networks found in Twitter. Highlighted in gold is a triad that gives rise to a *triadic awareness problem*. B follows A and C, but A does not follow C. Consequently, A does not know what B has heard from C.

building blocks of visibility, awareness and accountability—are not tied to a specific digital space. Rather, social signals *flow through* networks. Whereas E&K envisioned digital "architecture" (i.e., rooms and spaces), today networks structure many aspects of online social life.

Of course, they also permit a scale of social media unthinkably large for 2000, when social translucence was first introduced. Twitter has at least 38 million active users; imagine them all in the same chat room. (Where "active" is defined, for the sake of argument, as following at least 16 people [18].) Networks let us carve up our attention, looking at a select group of people rather than at everyone. This is probably a prerequisite for scale.

Yet while the underlying architecture may have changed, we are still social creatures attuned to social cues. We want social information so we can act appropriately. We want to know who's interested in what we say. We want to know whether our audience is bored and looking out the window or hanging on our every word. We want rich social cues that also respect privacy. We think the spirit of *Social Translucence* remains the same. Only the landscape has changed.

A Triadic Awareness Problem

What does this mean for design? As we discussed earlier, we believe social translucence breaks down because of fundamental structural differences between physical space and social networks. Therefore, we adopt a structural perspective in this paper. Our major theoretical claim is that different structures embedded within complex social networks give rise to different social translucence design problems.

We will work an example end-to-end in this paper. However, we revisit this broader claim—that network structures unlock design problems—in our *Discussion* section as it suggests new open problems for social computing. Now, we introduce a particular *triadic awareness problem*. A triad is a small fragment of a social network: three people and the ties that bind them. They can exhibit various shapes. Let us consider the directed attention networks like the ones that underly Twitter and Google+. Given three people—call them A, B and C—each relationship can take on one of four

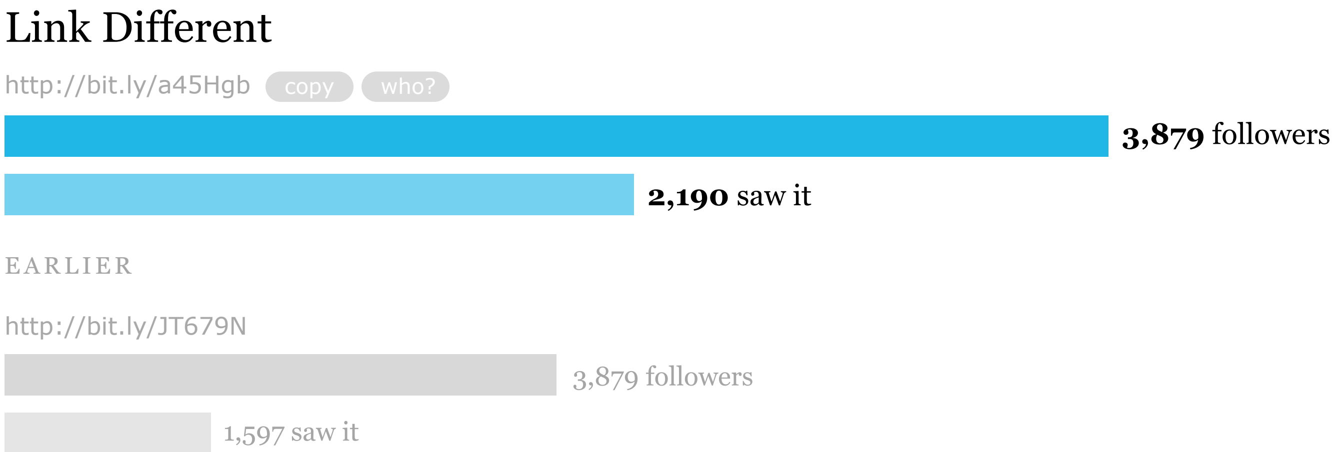


Figure 4. The Link Different web interface. When a user visits a webpage, she can use the Link Different bookmarklet to compute how many of her Twitter followers have already seen it from someone else they follow. Link Different displays the information as simple bar charts and includes some recent history. Users can also copy the link and view a random sample of followers who already saw it.

states: absent, left-directional (Twitter's "following"), right-directional (Twitter's "followed by") or bidirectional. $64=4^3$ possible triads emerge from the multiplicative combination of pairwise ties. For example, a triad might be fully closed, $A \leftrightarrow B \leftrightarrow C \leftrightarrow A$ or some less-connected variant, like $A \leftrightarrow B \leftarrow C$. (See [19] for background on triads and [28] for their recent use in social media research.)

A triadic awareness problem arises from a certain triad, the one shown in gold in Figure 3: $A \leftarrow B \rightarrow C$. (The arrows indicate attention flowing in that direction.) Set in the context of Twitter, this triad means that B follows both A and C, but A does not follow C (or vice versa). However, because A is not listening to C, this presents a problem. If A wants to share novel information with B, then she has to estimate whether B already heard it from C.

In practice, however, this problem is orders of magnitude harder than what we just described. A must estimate whether B already heard her information from C *or from anybody else* B *follows*. Furthermore, A has followers other than B and must estimate that same thing for each of them. Let us assume that the average active Twitter user follows 100 people and has 100 followers—probably a conservative estimate [3]. Then A has to estimate novelty for 100 followers, each of whom follows 100 people. Therefore, a typical scenario involves roughly 100^2=10,000 triadic awareness problems. All this complexity arises from a simple question: "What have your followers already heard?"

Contrast what we just described with a space like a chat room. There, A simultaneously listens to B and C. In fact, we might consider a chat room a special theoretical case: the fully closed triad $A \leftrightarrow B \leftrightarrow C \leftrightarrow A$. In a chat room,

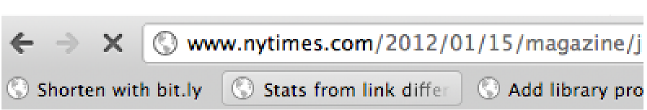


Figure 5. Link Different bookmarklet installed in Chrome. While visiting a webpage, users click it to run Link Different on the page's URL.

everyone can talk and listen to everybody else. (They may not always talk to and listen to everyone, but it is always possible.) Social networks purposely dissect this idea. A and C *choose* not to pay attention to each other; they don't want social translucence's visibility. Like in Figure 1 at the beginning of the paper, networks disrupt visibility (and hence, awareness and accountability) between people.

LINK DIFFERENT

The theory we have developed suggested a need for a tool to solve the triadic awareness problem on social network sites. This is why we developed Link Different. Link Different is a web-based application that lets Twitter users know how many of their followers already saw a link from someone else they follow. We built Link Different to enhance the common practice of link-sharing on Twitter. Recent estimates suggest that at least 25% of all tweets contain a link to somewhere else on the web [45]. We focused on links, rather than more general concepts like "newsworthy topics" (e.g., how Google News groups stories) for tractability reasons which we explain shortly. Link Different is designed to support a scenario like the following:

> Sue takes a break from the report she's writing to browse the *New York Times*. She reads a fascinating article about the Chinese economy's rising service sector. Having built a Twitter following based in part on her economics expertise, she goes to share the story on Twitter. But then Sue notices its publication date: *yesterday*. News flies fast on Twitter, and she sees the distinct possibility that her followers already saw this article from someone else. She brings the URL to Link Different. Link Different provides her with a simple bar chart showing the proportion of her followers who already saw the *New York Times* piece from somebody else they follow. Sue chooses to share it after seeing that only 22% of her followers saw the story already.

People interact with Link Different via a bookmarklet they install in their browser. Much like bit.ly's bookmarklet [12], users install a small snippet of Javascript in their browser's bookmarks bar (see Figure 5). While visiting some page on the web and considering sharing it on Twitter, users can click the bookmarklet to have Link Different operate on the their

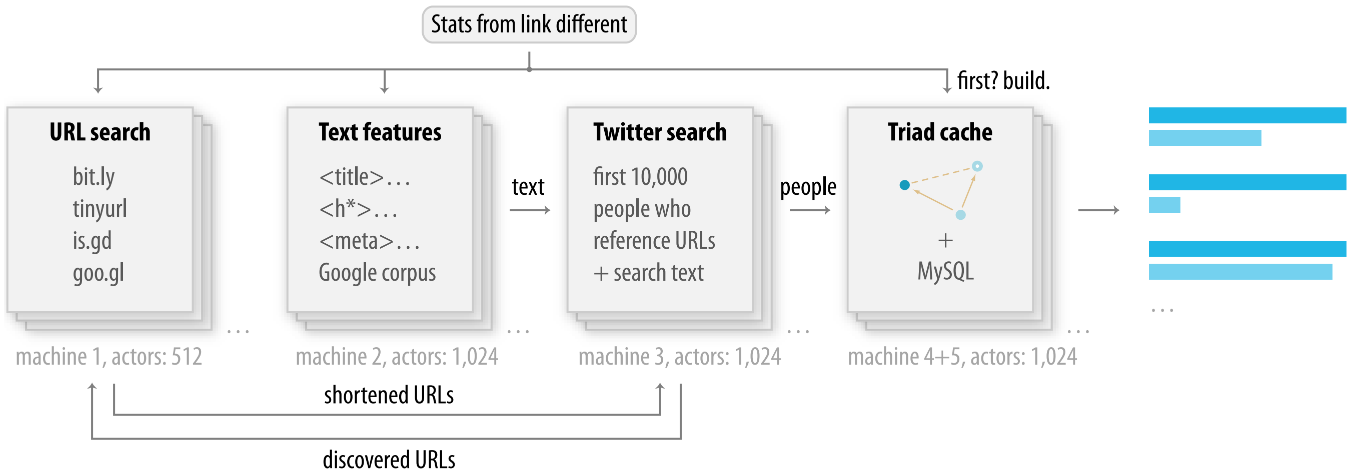


Figure 6. Link Different's server-side architecture. The application is distributed across five load-balanced machines, each running its own pool of parallel Scala actors. URLs and textual features get passed to a Twitter search to find references to the target webpage. The first 10,000 people to reference the target page are compared to the triad cache, which we build when someone uses Link Different for the first time.

page's URL. To get access to the bookmarklet, people first have to authenticate Link Different's Twitter app—something they do once at sign-up. (We explain the reason behind authentication in the *Implementation* section.)

When a user clicks the Link Different bookmarklet, a new tab opens, bringing them to a page like the one in Figure 4. We modeled this interaction on bit.ly's behavior when shortening URLs. After Link Different finishes its crawl of the user's Twitter network, Figure 4's simple bar charts appear. The top one shows the total number of followers, while the one below it shows how many followers already saw it from someone else they follow. In Figure 4, slightly more than half of the user's followers have seen it already. Link Different also shows a recent history in gray. We pursued a very simple design, hiding the complexity of the distributed Twitter crawl in the server-side software.

Link Different automatically shortens the link for users, using bit.ly's API, and it also provides a copy button to let users quickly put the shortened link on their clipboard. After our release, we received many tens of requests from users asking for a feature to let them know *which* followers had seen the link already. The logic behind many of these requests seemed to be that different followers are differentially important, and our users might be willing to bother some as long as a few others benefitted. In response, we added the "who?" button. It reveals a random sample of up to 50 followers who already saw the link, sorted by their number of followers—a proxy for their status on Twitter.

Implementation

Link Different's user interface is simple HTML, CSS and jQuery [39] served by PHP. Link Different's server-side architecture, on the other hand, is a parallelized distributed system. Figure 6 presents an overview. To ensure acceptable response times with potentially thousands of concurrent users, Link Different is distributed among five high-end machines (e.g., Quad-core i7, 16GB+ RAM, etc.) that interact with each other via message passing. The software is written in Scala and uses its actor and message-passing implementations [32]. That is, the components broken down in gray boxes in Figure 6 operate in parallel with one another, responding to requests from other components when they have available time and resources. The Link Different architecture makes thousands of HTTP requests every time a user clicks the Link Different bookmarklet, yet its distributed design means that most people receive responses within 10 seconds. (This is admittedly long by web standards, but we considered it acceptable for a research prototype.) Also, Twitter granted us a whitelisted IP address, providing Link Different with 20,000 API requests per hour, per user—the reason we require users to authenticate our Twitter app.

A historical quirk in Twitter's design substantially complicates the Link Different implementation. Since Twitter limits tweets to 140 characters (an artifact of Twitter's early text message focus), people have an incentive to shorten their URLs to save characters in their tweets. A number of services sprung up to serve this need, such as bit.ly, the most popular URL shortener [1]. Moreover, many of these shortening services provide their users with personalized click-through statistics. They do this by creating custom short-URLs for every person (e.g., http://bit.ly/AB4fgT for you and http://bit.ly/6kyk90 for me), each resolving to the same long-URL via HTTP redirects. This means that for a popular article on the web, say a *New York Times* article, thousands of short-URLs often exist. Link Different needs to track all of these. While bit.ly and tinyurl APIs permit discovering all short-URLs (and luckily account for nearly 90% of short-URLs [1]), other services do not.

Therefore, in addition to pushing the Link Different user's long-URL through various shorteners, we also perform Twitter searches on key textual information from the webpage. Our goal is to find references to the target article not discovered via the URL search. We extract text within the `title`, `header` (i.e, `<h1>`, `<h2>`, `<h3>` and `<h4>`) and `meta` tags. Every possible tri-gram phrase is used as input to its own Twitter search. In addition, we use the Google 1T corpus [14] to compute normalized TF-IDF statistics [25] for each uni-, bi- and tri-gram from the page's non-HTML text. The Google 1T corpus has baseline frequencies for phrases on the web, and computing these statistics uncovers a page's most descriptive phrases. The phrases scoring in the top decile become inputs for their own Twitter searches. For all these searches (a typical page creates around 500), any tweet containing a URL is queued for inspection by the "URL search" process. If they resolve to Link Different's target, then the person tweeting the URL is a candidate. We consider up to 10,000 Twitter users as candidates in our server's final stage. (An alternative to our textual approach, one taken by the recently acquired BackType [4], is to catalog and index every URL on the Twitter firehose. While it would greatly simplify Link Different, we did not have resources to build something at this scale.)

The "URL search" and "Text features" searches ultimately generate candidate Twitter users who could be parts of triads similar to the example in Figure 3. To find the subset that belong to these triads, we cache all people who could be Cs in a Figure 3 triad. The first time someone clicks the Link Different bookmarklet, we build a triad cache in parallel with the ongoing search described above. A distributed crawl is initiated, where we build an index mapping every C to a list of Bs that follow C. We store this cache in a MySQL database. Here, we traded space for time, as the average Link Different user requires roughly 1K of cache space. However, as a result, for every candidate Twitter user returned via searches, we can use our index to do a constant-time lookup, retrieving every one of our user's followers who follows the candidate. After removing duplicates by adding them to a set data structure, we arrive at the number of followers who already saw this link from someone else they follow.

Limitations

Link Different's user interface is very simple. It does not let you explore *who* mentioned a link to your followers, nor *when* they saw it. It cannot tell you whether your followers actually clicked on the link they saw from someone else, only that it appeared in their Twitter streams. People would almost certainly welcome this deep information. Furthermore, it would be interesting to extend Link Different to operate on stories rather than links. For example, when a major event happens (e.g., the Super Bowl), various outlets cover it. We can envision users welcoming the opportunity to explore concepts rather than simple URLs. Link Different's text searches might be the beginning of an approach.

Moreover, Link Different makes probabilistic tradeoffs in its server implementation. There is no guarantee that the "saw it" number is precise. For performance reasons, we traded off precision for response time. Deeper indexing strategies might resolve these inconsistencies.

WEB FIELD STUDY

We chose to evaluate Link Different as an open field study on the web. Other approaches also make sense, such as an in-depth, but small-N field study like the one in [9]. Yet, we were primarily interested in establishing a new theoretical claim: that this triadic awareness problem exists and people struggle with it. Rather than showing that a technology solves a problem we know to exist, in this paper we want to study whether the problem exists at all. For this reason, we chose to open Link Different up to any Twitter user on the web.

After finalizing the Link Different build, we "launched" it by sending short email messages describing the project to a handful of well-known technology blogs. We were fortunate enough to have one of them cover Link Different [41]. Often, one of the biggest hurdles for a release—especially for an academic lab—is getting the initial word out, and this helped us overcome it. Simultaneously, the blog coverage meant that we did not have to rely on our existing online networks, such as Twitter, Facebook and internal email lists. For example, this sidesteps validity threats due to snowball samples originating with a technology's creators [10]. After the initial coverage, Link Different primarily spread virally through Twitter. During its two months online, 144,232 unique people used Link Different 1,343,322 times in total. The blog coverage drove an initial burst of visitors to the site, but via HTTP referrer headers [29] we found that 84% of Link Different's usage came from our users discussing Link Different on Twitter.

Sharing Behavior

After releasing Link Different, we found that we were generating very interesting data about conformity and novelty effects on social media sharing. Although not our intention, you can view the simple Link Different interface as a prompt to this question: "Now that you know what proportion of your followers have already seen your link, do you still want to share it?" We realized that viewing sharing behavior as a function of the proportion of followers who have already seen the link may yield new insight into conformity and novelty in social media.

Because Javascript cannot copy text to the clipboard across browsers, Link Different uses a snippet of Flash code to let users copy the bit.ly link. Rather than load this code upfront, we use jQuery to load it dynamically into the page when the user clicks the copy button. Via this mechanism, our web server logs recorded when a user decided to copy a link and when they didn't. Presumably, when they use the copy button, they intend to share it on Twitter.

To verify this *copy-as-proxy* idea, we randomly sampled 5,000 instances where a user copied a link and 5,000 where they did not (out of the 1,343,322 total events). Our goal was to estimate how often a copy click leads to sharing and how often a non-click entails not sharing. If consistent, we can use the copy click as a proxy for sharing without having to crawl every tweet ever made by everyone who used Link Different—something we simply did not have the resources to accomplish. Over 90% of copy clicks lead to sharing and over 90% of non-clicks lead to not sharing. Therefore, for the remainder of this discussion, we will use copy-click data as a proxy for sharing data.

From our 1,343,322 events, we computed the proportion of links shared as a function of a moving 2% "already seen" window. Figure 7 illustrates the results. For example, the leftmost bar in Figure 7 says that people shared links 42% of the time when only 0-2% of their followers had seen the link from

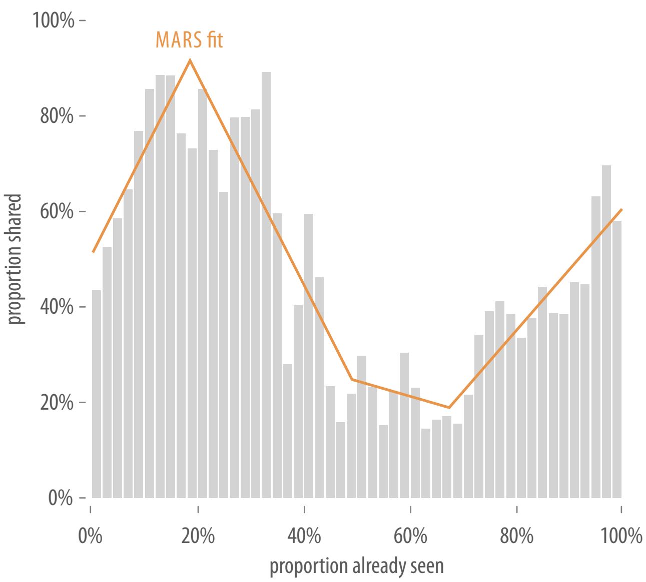


Figure 7. The proportion of links shared (*y*-axis) as a function of the proportion depicted in the Link Different interface (*x*-axis). Given a stimulus of how many followers already saw a link, the graph shows the likelihood of sharing it. The orange line plots a Multiple Adaptive Regression Splines (MARS) fit, uncovering four different regions.

someone else. We chose a 2%-wide window because it was the smallest integer to produce 99% confidence intervals of $\pm1\%$ using the Agresti-Coull confidence interval technique for proportions [2]. Other window sizes could presumably change Figure 7's shape slightly.

It would be nice to abstract away from some of the noise in Figure 7. Therefore, we used R's MDA library [33] to compute a Multiple Adaptive Regression Splines (MARS) fit of our data. In essence, the technique iteratively converges on a piecewise linear fit, uncovering regions with fundamentally different shapes. It learns these regions from data. The orange line in Figure 7 shows the MARS fit. It uncovers four segments, each modeled as a line with a different slope. From left to right, they have slopes (i.e., β-weights), 3.86, -4.46, -2.23 and 1.9. This model fits the data reasonably well, with an Adj. $R^2 = 0.445$. While some regions in Figure 7 are distinctly curvilinear, the MARS decomposition provides an abstract layer over which we can reason and potentially build theory—something we attempt in the *Discussion* section.

Limitations

The analysis performed above does not tell us about how often people share based on what they estimate their followers have already seen. Presumably, this calculation is part of everyday sharing behavior on social media sites. It would be interesting and insightful to look into how sharing varies as a function of *perceived* novelty.

Word-of-mouth Twitter referrals drove visitors to Link Different. While blog coverage started it, this is a snowball sample nonetheless. Snowball samples arise when one participant refers another, who refers another, etc. A concern is that the participant pool is biased—or, in this context, that it follows the path of least resistance. Researchers often use snowball sampling to recruit difficult-to-reach populations [27], and you could argue that this is one. In any case, we make no claims that we have a random, representative sample, even of social media users. We revisit this issue shortly in the *Discussion* section.

DISCUSSION

We worked from theory to uncover a problem we suspected to exist on sites backed by social networks. We built Link Different to examine this idea, a strategy we might call a "technology probe" [37] at internet scale. We were honestly surprised by the response from Twitter. Within hours, our servers hit their memory limits and we rushed to deploy an optimized architecture (the one depicted in Figure 6). While we firmly believe social media research does not have to spread virally to show value, we should also embrace it when it happens—something other authors have also argued recently [8]. That is, word-of-mouth adoption is sufficient (and welcome), but not necessary. In this case, we think the widespread, rapid and voluntary adoption of Link Different speaks to the scale and reality of this triadic awareness problem.

That said, we were lucky. The press could have just as easily ignored us. A release strategy based on an authoritative blog linking to your project is risky at best. Like other authors, we have struggled with getting the word out in other projects. We can offer no conclusive strategies or advice. Simply put, this time it worked. Furthermore, viral adoption has its downsides. For example, we cannot claim that we recruited even a random *Twitter* sample, let alone a random sample. We attracted those people willing to experiment with a new tool. This is what we meant when we said "the path of least resistance" earlier. Like a river, Link Different probably found its way via referrals to the people who were most inclined to use it. While 100K+ users is quite high for an academic project, you could reframe the number the following way. Very conservatively, Twitter has 38 million active users. Do only 0.4% of them (i.e., 150K/38M) face the triadic awareness problem? Also, probably not. We think it is wise to not test the total numbers of users statistic too rigorously. However, we do claim that the widespread and voluntary response suggests a deep problem lurking under the network.

Users' sharing behavior surprised us. When we built Link Different, we hypothesized that sharing was strictly inversely proportional to novelty. We would have expected a graph that started high on the left and went primarily down from there. Rather, we find a bimodal distribution, with curious points at the leftmost and rightmost ends. The likelihood of sharing starts at around 40% when almost no one has already seen your link, steadily rising to over 80% when 16-18% of your followers haven't seen it. One way to interpret this is that perhaps we do not want to share things that *no one* seems to be interested in, and we wait until a bigger (but still small) fraction of people see it to finally share the link. From there, the data does what we would expect, moving steadily down through the next two MARS segments. Then it goes back up again. This truly surprised us. At around 72%, sharing behavior starts to track steadily up from its low points in the 40-60% range, finally peaking when almost everyone has seen it (the high bars in the 94-100% range). Perhaps an interpretation is that sharing with people who already know about it is valued because it serves to signal that you too are "in the know."

These data provide a starting point for asking questions about whether sharing behavior depends on the genre of the link shared. Also, does it vary as a function of follower counts or time on the site? These are all interesting questions, and we look forward to follow-up research.

Social Translucence Over Social Networks

We built Link Different around a theoretical idea for designing social network sites: small structures embedded within social networks unlock social translucence design problems. Specifically, we employed a method where we started with a triad and looked for where social translucence breaks down over it. With Link Different, we looked at one triad and built a system to solve a an awareness problem that arises from it. The response from Twitter suggests the scope of the problem.

We believe this extension of social translucence has depth. If you count the members of a triad as interchangeable (i.e., the *A*, *B*, *C* labels have no real meaning), then directed social networks have 16 non-isomorphic triads ([26], pgs. 141–142). (In network/graph theory, "isomorphism" means you can turn one graph into another by relabeling the nodes/vertices.) This gives us 16 structures over which we can examine where social translucence breaks down. Intersecting any one of these structures with a social translucence design trait (i.e., visibility, awareness and accountability) could yield new insight into problems in today's social media. In other words, this means 48=16·3 new design spaces.

In Table 1, we have mapped out six problems we feel likely to exist over certain triads. This is the start of work in this space, and much research needs to be done cataloging and assessing triadic social translucence problems. We offer Table 1 as a sample of what this approach might produce. For example, in Table 1's first example design problem, we highlight tensions that arise from Granovetter's "forbidden triad" [31]. In his words, "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." (You have to allow for mapping his labels to ours and letting his "strong tie" be our "bi-directional.") So how does *B* carry on a simultaneous conversation with *A* and *C* without drawing attention to the forbidden triad? In a simultaneous conversation between the three, *B* would reference both *A* and *C*, but *C* would notice that *A*'s replies never mention him. While this triad occurs relatively infrequently in natural social networks [35], the sheer scale of modern social media sites suggests millions of these triads exist on the internet today. This also highlights an interesting point: design problems are conditioned upon the baseline probabilities for how often particular triads occur in social networks.

In Table 1's fourth example, we see a problem Twitter addressed a few years after it launched. In Twitter's early days online, *A* could see *B*'s conversation with *C*, even though *A* doesn't follow *C*. But since *A* doesn't follow *C*, she never saw *C*'s replies. Twitter solved this with software changing how it routes messages: *A* now only sees *B*'s conversation with *C* if she also follows *C*, cutting *A* out of the conversation entirely.

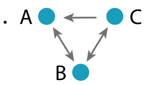
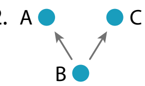
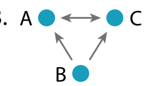
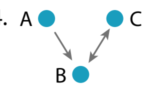
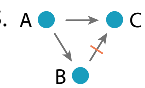
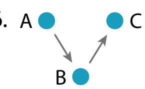
Triad	Trait	Example design problem
1. 	Visibility	When chatting with A and C, how does B not highlight the "forbidden triad?"
2. 	Visibility	B hears something from A relevant to C. How does B bring it to C's attention?
3. 	Awareness	Are A and C always aware that B hears everything they say to one another?
4. 	Awareness	A can hear what B says to C, but not what C says back. * addressed by Twitter.
5. 	Awareness	A followed C because B did. Now B severed the tie. Does A still want to follow?
6. 	Accountability	B can take credit for what C says, since A only hears C through B.

Table 1. Six triads and potential social translucence design problems that arise from them. We argue that design problems arise from specific structures, although more work needs to be done establishing prevalence and importance.

Other solutions exist, but we highlight this case to show a real-world occurrence of a triadic design problem. You can also consider the structural dynamics in triads and its role in design (i.e., how ties change over time). We take a first step in this direction with Table 1's fifth design problem.

What about bigger structures? In this paper, we have exclusively considered triads, leaving out higher-order structures connecting four, five or even six people. Higher-order structures may yield their own insights, but perhaps at the cost of greater conceptual complexity. The complexity of such structures rises roughly as the square of the number of people in them, so we must carefully choose what structures designers should consider. In other words, the bigger the structure, the harder it is to reason about. In any case, we need more work cataloging where social translucence breaks down over specific structures within networks.

Social translucence offers fundamental theoretical guidance for social media designers: *who-sees-what*, *who-knows-what* and *who-knows-that-I-know*. We argue that it is out of these building blocks that bigger theories about designing social media emerge [40]. We need to know as much as we can about the lower levels of design (i.e., their structural properties) to support higher-level problems, such as starting a new community. We think Link Different highlights the significance of structural design problems underlying today's social media.

Limitations & Future Work

We hope to see new work pick up where we left off. Link Different's user interface could benefit from extensions permitting deeper inspection of the social connections behind sharing. The server implementation could be extended to handle more complex stories, not just URLs. While we find the study of sharing behavior in this paper fascinating, it is just a start. It presents as many questions as answers.

We hope future work finds traction with the new theory in this paper. That said, what we have presented here is just the beginning. We need more work to uncover which problems occur in the real world and which matter most. We envision both theoretical work that extends this research, as well as more practically oriented systems work establishing the parameters of particular problems.

CONCLUSION

We believe this work addresses a fundamental issue for modern social media: providing social cues over social networks. The theory we would normally turn to—social translucence—breaks down because social networks resist social translucence's physical metaphors. In this paper, we built theory relating social translucence to social network structure, updating social translucence for the era of social networks. Based on the widespread and viral response we got when we applied it, we think this approach has the potential to uncover other problems lurking in social media today.

ACKNOWLEDGEMENTS

We would like to thank the comp.social lab and Amy Bruckman for reviewing early drafts of this work. Leslie Gilbert came up with the hypothetical example used in the introduction. Both Yahoo! and Google helped support the work.

REFERENCES

1. D. Aamoth. "When It Comes To URL Shorteners, bit.ly Is Now The Biggest." *TechCrunch.* http://tcrn.ch/aqGvUh. Retrieved September 23, 2011.
2. A. Agresti and B. Coull. Approximate Is Better Than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, 52:119–126, 1998.
3. C. Arthur. "Average Twitter User has 126 Followers, and Only 20% Go via Website." *The Guardian.* http://bit.ly/3AiLy. Retrieved September 23, 2011.
4. BackType. http://backtype.com. Retrieved September 23, 2011.
5. A. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.
6. P. S. Bearman, J. Moody, and K. Stovel. Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American Journal of Sociology*, 110(1):44–91, July 2004.
7. J. B. Begole, J. C. Tang, R. B. Smith, and N. Yankelovich. Work Rhythms: Analyzing Visualizations of Awareness Histories of Distributed Groups. In *Proc. CSCW*, pages 334–343, 2002.
8. M. S. Bernstein, M. S. Ackerman, E. H. Chi, and R. C. Miller. The Trouble with Social Computing Systems Research. In *Proc. EA CHI*, pages 389–398, 2011.
9. M. S. Bernstein, A. Marcus, D. R. Karger, and R. C. Miller. Enhancing directed content sharing on the web. In *Proc. CHI*, pages 971–980, 2010.
10. P. Biernacki and D. Waldorf. Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research*, 10(2):141–163, 1981.
11. J. Birnholtz, L. Reynolds, E. Luxenberg, C. Gutwin, and M. Mustafa. Awareness Beyond the Desktop: Exploring Attention and Distraction with a Projected Peripheral-vision Display. In *Proc. GI*, pages 55–62, 2010.
12. Bit.ly. Bit.ly tools. https://bitly.com/pages/tools. Retrieved September 23, 2011.
13. d. boyd and N. Ellison. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2008.
14. T. Brants and A. Franz. Web 1T 5-gram Version 1. *Linguistic Data Consortium, Philadelphia*, 2006.
15. M. Burke, C. Marlow, and T. Lento. Social Network Activity and Social Well-being. In *Proc. CHI*, pages 1909–1912, 2010.
16. R. Burt. *Structural Holes: The Social Structure of Competition*. Harvard University Press, August 1995.
17. R. S. Burt. Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399, September 2004.
18. N. Carlson. "How Many Users Does Twitter Really Have?" *Business Insider.* Retrieved September 23, 2011.
19. D. Cartwright and F. Harary. Structural Balance: A Generalization of Heider's Theory. *Psychological Review*, 63(5):277, 1956.
20. N. Ellison, C. Steinfield, and C. Lampe. The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.
21. T. Erickson, C. Halverson, W. Kellogg, M. Laff, and T. Wolf. Social Translucence: Designing Social Infrastructures that Make Collective Activity Visible. *Communications of the ACM*, 45(4):44, 2002.
22. T. Erickson and W. Kellogg. Social Translucence: An Approach to Designing Systems that Support Social Processes. *TOCHI*, 7(1):59–83, 2000.
23. T. Erickson, D. Smith, W. Kellogg, M. Laff, J. Richards, and E. Bradner. Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of "Babble". In *Proc. CHI*, pages 72–79, 1999.
24. J. Fogarty, J. Lai, and J. Christensen. Presence Versus Availability: The Design and Evaluation of a Context-Aware Communication Client. *International Journal of Human-Computer Studies*, 61(3):299–317, 2004.
25. W. B. Frakes and R. Baeza-Yates. *Information Retrieval: Data Structures and Algorithms*. Prentice Hall, 1992.
26. O. Frank. Triad Count Statistics. *Discrete Mathematics*, 72(1-3):141–149, 1988.
27. O. Frank and T. Snijders. Estimating the Size of Hidden Populations Using Snowball Sampling. *Journal of Official Statistics*, 10(1):53–67, 1994.

28. S. Golder and S. Yardi. Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality. In *Proc. IEEE SocialCom*, 2010.
29. Google Analytics. http://google.com/analytics. Retrieved September 23, 2011.
30. Google Scholar. http://scholar.google.com/scholar?q=social+translucence. Retrieved September 23, 2011.
31. M. S. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.
32. P. Haller and M. Odersky. Scala Actors: Unifying Thread-based and Event-based Programming. *Theoretical Computer Science*, 410(2-3):202–220, 2009.
33. T. Hastie and R. Tibshirani. http://cran.r-project.org/web/packages/mda. Retrieved September 23, 2011.
34. J. Hollan and S. Stornetta. Beyond Being There. In *Proc. CHI*, pages 119–125, 1992.
35. P. Holland and S. Leinhardt. A Method for Detecting Structure in Sociometric Data. *American Journal of Sociology*, 76(3):492–513, 1970.
36. G. Hsieh, K. P. Tang, W. Y. Low, and J. I. Hong. Field Deployment of IMBuddy: A Study of Privacy Control and Feedback Mechanisms for Contextual IM. In *Proc. UbiComp*, pages 91–108, 2007.
37. H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderbäck. Technology Probes: Inspiring Design for and with Families. In *Proc. CHI*, pages 17–24, 2003.
38. M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.
39. jQuery. http://jquery.com. Retrieved September 23, 2011.
40. R. E. Kraut and P. Resnick. *Evidence-Based Social Design: Mining the Social Sciences to Build Online Communities*. MIT Press, 2011 (forthcoming).
41. F. Lardinois. "Don't Bore Your Twitter Followers: Link Different" *ReadWriteWeb.* http://rww.to/qQ2IBr. Retrieved January 13, 2012.
42. Localmind. http://foursquare.com/app/localmind. Retrieved September 23, 2011.
43. 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.
44. S. Paul, L. Hong, and E. Chi. Is Twitter a Good Place for Asking Questions? A Characterization Study. In *Proc. ICWSM*, 2011.
45. L. Rao. "Twitter Seeing 90 Million Tweets Per Day, 25 Percent Contain Links." *TechCrunch.* http://tcrn.ch/by1ZrB. Retrieved September 23, 2011.
46. B. Suh, E. Chi, A. Kittur, and B. Pendleton. Lifting the Veil: Improving Accountability and Social Transparency in Wikipedia with Wikidashboard. In *Proc. CHI*, pages 1037–1040, 2008.
47. M. Svensson, K. Höök, J. Laaksolahti, and A. Waern. Social Navigation of Food Recipes. In *Proc. CHI*, pages 341–348, 2001.