Quantifying the Invisible Audience in Social Networks

Michael S. Bernstein[1,2], **Eytan Bakshy**[2], **Moira Burke**[2], **Brian Karrer**[2]

Stanford University HCI Group[1]
Computer Science Department
msb@cs.stanford.edu

Facebook Data Science[2]
Menlo Park, CA
{ebakshy, mburke, karrerb}@fb.com

ABSTRACT

When you share content in an online social network, who is listening? Users have scarce information about who actually sees their content, making their audience seem invisible and difficult to estimate. However, understanding this invisible audience can impact both science and design, since perceived audiences influence content production and self-presentation online. In this paper, we combine survey and large-scale log data to examine how well users' perceptions of their audience match their actual audience on Facebook. We find that social media users consistently underestimate their audience size for their posts, guessing that their audience is just 27% of its true size. Qualitative coding of survey responses reveals folk theories that attempt to reverse-engineer audience size using feedback and friend count, though none of these approaches are particularly accurate. We analyze audience logs for 222,000 Facebook users' posts over the course of one month and find that publicly visible signals — friend count, likes, and comments — vary widely and do not strongly indicate the audience of a single post. Despite the variation, users typically reach 61% of their friends each month. Together, our results begin to reveal the invisible undercurrents of audience attention and behavior in online social networks.

Author Keywords

Social networks; audience; information distribution

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Web-based interaction

General Terms

Design; Human Factors.

INTRODUCTION

Posting to a social network site is like speaking to an audience from behind a curtain. The audience remains invisible to the user: while the invitation list is known, the final attendance is not. Feedback such as comments and likes is the only glimpse that users get of their audience. That audience varies from day to day: friends may not log in to the site,

may not see the content, or may not reply. While established media producers can estimate their audience through surveys, television ratings and web analytics, social network sites typically do not share audience information. This design decision has privacy benefits such as plausible deniability, but it also means that users may not accurately estimate their invisible audience when they post content.

Correct or not, these audience estimates are central to media behavior: perceptions of our audience deeply impact what we say and how we say it. We act in ways that guide the impression our audience develops of us [17], and we manage the boundaries of when to engage with that audience [2]. Social media users create a mental model of their imagined audience, then use that model to guide their activities on the site [27, 38, 26]. However, with no way to know if that mental model is accurate, users might speak to a larger or smaller audience than they expect.

This paper investigates users' perceptions of their invisible audience, and the inherent uncertainty in audience size as a limit for users' estimation abilities. We survey active Facebook users and ask them to estimate their audience size, then compare their estimates to their actual audience size using server logs. We examine the folk theories that users have developed to guide these estimates, including approaches that reverse-engineer viewership from friend count and feedback.

We then quantify the uncertainty in audience size by investigating actual audience information for 220,000 Facebook users. We examine whether there are reasonable heuristics that users could adopt for estimating audience size for a specific post, for example friend count or feedback, or whether the variance is too high for users to use those signals reliably. We then test the same heuristics for estimating audience size over a one-month period.

While previous work has focused on highly visible audience signals such as retweets [5, 32], this work allows us to examine the invisible undercurrents of attention in social media use. By comparing these patterns to users' perceptions, we can then identify discrepancies between users' mental models and system behavior. Both the patterns and the discrepancies are core to social network behavior, but they are not well understood. Improving our understanding will allow us to design this medium in a way that encourages participation and supports informed decisions around privacy and publicity.

We begin by surveying related work in social media audiences, publicity, and predicting information diffusion. We then perform a survey of active Facebook users and compare their estimates of audience size to logged data. We then de-

scribe the inherent unpredictability of audience size on Facebook, both for single posts and over the course of a month.

RELATED WORK

Social media users develop expectations of their audience composition that impact their on-site activity. Designing social translucence into audience information thus becomes a core challenge for social media [16, 15]. In response, speakers tune their content to their intended audience [12]. On Facebook and in blogs, people think that peers and close online friends are the core audience for their posts, rather than weaker ties [24, 38]. Sharing volume and self-disclosure on Facebook are also correlated with audience size [10, 40]. However, as the audience grows, that audience may come from multiple spheres of the user's life. Users adjust their projected identity based on who might be listening [17, 27] or speak to the lowest common denominator so that all groups can enjoy it [20]. Social media users are thus quite cognizant of their audience when they author profiles [8, 14], and accurately convey their personality to audiences through those profiles [18].

Our notion of the invisible audience is tied to the imagined audience in social media [27, 26]. The imagined audience usually references the types of groups in the audience — whether friends from work, college, or elsewhere are listening. In this paper, we focus not on the composition of the audience but on its size. Both elements play important roles in how we adapt our behaviors to the audience.

Cues can be helpful when estimating aspects of a social network, but there are few such cues available today. For example, to estimate the size of a social network, it can be helpful to base the estimate on the number of people spoken to recently [21] or to focus on specific subpopulations and relationships such as diabetics or coworkers [28]. However, it can be difficult to estimate how many people within the network will actually see or appreciate a piece of content. Public signals such as reshares [32] and unsubscriptions [22] give users feedback about the quality of their content, but users often consume content and make judgments without taking any publicly visible action [3, 13]. In addition, there are consistent patterns in online communities that might bias estimates: for example, the prevalence of lurkers who do not provide feedback or contribute [30] and a typical pattern of focusing interactions on a small number of people in the network [4].

Questions of audience in social media often reduce to questions of privacy. Users must balance an interest in sharing with a need to keep some parts of their life private [2, 31]. However, early studies on social network sites found no relationship between disclosure and privacy concerns [1, 34, 40]. Instead, people tended to want others to discover their profiles [40], and filled out the basic information in their profile relatively completely [23]. However, young adults are increasingly, and proactively, taking an active role in managing their privacy settings [35].

Design decisions with respect to audience visibility can influence users' interactions with their audience [9]. For example, Chat Circles allows members of a chat room to modify their audience by moving their avatar closer or further from other participants [39]. Visualizations can also support socially translucent interactions by making the members of the audience more salient [25, 11] or displaying whether an intended audience member has already seen similar content [7]. BlackBerry Messenger, Apple iMessage, and Facebook Messenger all provide indicators that the recipient has opened a message, and online dating sites OkCupid and Match.com reveal who has viewed your profile. Few studies have addressed users' reactions to this explicit audience indicator [33], though some users of the social networking site Friendster expressed concerns that they were uncomfortable with this level of social transparency and would consequently not view as many profiles [29].

Our work pushes the literature forward in important respects: we augment previous research with quantitative metrics, we focus on contexts where the intended audience is friends and not the entire Internet, and we empirically demonstrate that audience size is difficult to infer from feedback. We also contribute empirical evidence for the wide variance in audience size, something not previously possible for blogs or tweets, because we can track consumption across the entire medium.

METHOD AND DATA

To study audiences in social media, we use a combination of survey and Facebook log data. Most Facebook content is consumed through the News Feed (or *feed*), a ranked list of friends' recent posts and actions. When a user shares new content — such as a status update, photo, link, or check-in — Facebook distributes it to their friends' feeds. The feed algorithmically ranks content from potentially hundreds of friends based on a number of optimization criteria, including the estimated likelihood that the viewer will interact with the content. Because of this and differences in friends' login frequencies, not all friends will see a user's activity.

Audience Logs

We logged audience information for all *posts* (status updates and link shares) over the span of June 2012 from a random sample of approximately 220,000 US Facebook users who share with friends-only privacy. We also logged cumulative audience size over the course of the entire month. To determine audience size, we used client-side instrumentation that logs an event when a feed item remains in the active viewport for at least 900 milliseconds. Our measure of audience size thus ensures that the post appeared to the user for a nontrivial length of time, long enough to filter out quick scrolling and other false positives. However, being in the audience does not guarantee that a user actually attends to a post: according to eyetracking studies, users remember 69% of posts that they see [13]. So, while there may be some margin of difference between audience size and *engaged* audience size, we believe that this margin is relatively small. Furthermore, we note that univariate correlations are unaffected by linear transformations, so the correlations between estimated and actual audience are the same regardless of whether a correction is applied to the data.
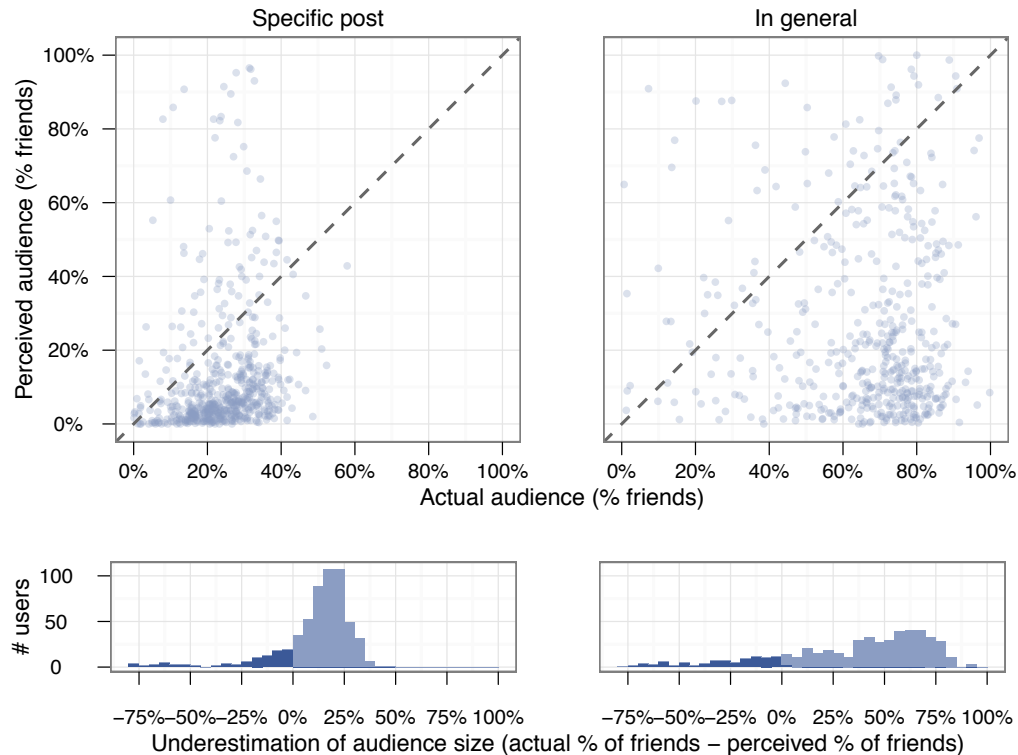
Figure 1: Comparisons of participants' estimated and actual audience sizes, as percentages of their friend count. Most participants underestimated their audience size. The top row displays each estimate vs. its true value; the bottom row displays this data as error magnitudes. Left column: The specific survey, where participants were shown one of their posts, comparing estimated audience to the true audience for that post. Right column: The general survey, where participants estimated their audience size "in general," comparing estimated audience to the number of friends who saw a post from that user during the month.

Our logging resulted in roughly 150 million distinct (source, viewer) pairs from roughly 30 million distinct viewers. All data was analyzed in aggregate to maintain user privacy. The number of *distinct friends providing feedback*, given in terms of likes and comments, was also logged for each post.

In our analysis, we did robustness checks by subsetting our data, e.g., active vs. less active users, and users with different network sizes. We saw identical patterns each time, so we report the results for our full dataset.

Survey

To compare actual audience to perceived audience, we surveyed active Facebook users about their perceived audience. Participants might estimate their audiences differently when they anchor on a specific instance than when they consider their general audience, so we prepared two independent surveys: *general* and *specific* audience estimation. In the general survey, participants answered the question, *"How many people do you think usually see the content you share on Facebook?"* In the specific survey, participants clicked on a page that redirected them to their most recent post, provided that post was at least 48 hours old. Specific survey participants then answered the question, *"How many people do you think saw it?"* In both surveys, participants then shared how they came up with that number. Finally, participants shared their desired audience size on a 5-point scale ranging from "Far

fewer people" to "Far more people".

We advertised the survey to English-speaking users in our random sample who had been on Facebook for at least 90 days, had logged into Facebook in the past 30 days, and who had shared at least one piece of content (e.g., status update, photo, or link) in the last 90 days. The general audience survey had 542 respondents, and the specific audience survey had 589. Sixty-one percent of respondents were female, with a mean age of 32.8 ($sd = 14.7$), and a median friend count of 335 ($mean = 457, sd = 465$).

PERCEIVED AUDIENCE SIZE

In this section, we investigate how users' perceptions of their audience map onto reality. This investigation has three main components. First, we quantify how accurate users are at estimating the audience size for their posts. Second, we perform a content analysis on users' self-reported folk theories for audience estimation. Third, we explore users' satisfaction with their audience size.

Audience size: perception vs. reality

We compared participants' estimated audience sizes to the actual audience size for their posts. Participants underestimated their audiences. Figure 1 plots actual audience size against perceived audience size for the two surveys. Accurate guesses lie along the diagonal line, where the actual audience is equal

to the perceived audience. The cluster of points near the x-axis indicates that the majority of participants significantly underestimated their audience.

For participants considering a specific post in the past (Figure 1, left column), the median perceived audience size was 20 friends ($mean = 60$, $sd = 163$); the median actual audience was 78 friends ($mean = 99$, $sd = 84$). Transformed into percentages of network size, the median post reached 24% of a user's friends ($mean = 24\%$, $sd = 10\%$), but the median participant estimated that it only reached 6% ($mean = 17\%$, $sd = 31\%$). In fact, most participants guessed that no more than fifty friends saw the content, regardless of how many people actually saw it. We note that this data is skewed and long-tailed, as is common with many internet phenonema: as such, we rely on medians rather than means as our core summary statistic.

We quantify the $relative\ error$ as $1 - \frac{perceived\ audience\ size}{actual\ audience\ size}$. The median relative error is 0.73, meaning the median estimate was just 27% of the actual audience size. In other words, the median participants underestimated their actual audience size by a factor of four.

Participants in the general survey also underestimated their audiences (Figure 1, right column). The median perceived audience size in the general survey was 50 ($mean = 137$, $sd = 236$), while the median actual audience (number of friends who saw any post the user produced in the previous month) was 180 ($mean = 283$, $sd = 302$). The median relative error was 0.68, indicating that participants underestimated their general audience by roughly a factor of three.

Figure 1 (bottom row) shows the distribution of errors in both surveys. In both cases, participants tended to underestimate their audiences, and there was greater variance in the general version of the survey.

Not surprisingly, estimates of audience size "in general" were significantly larger than those for a specific post in the past ($mean = 137$ vs. 60, $p < 0.001$). As disjoint sets of friends may see different pieces of content, it makes sense that in a month, more people would see a given user's content than would see any single post she made. Survey participants correctly accounted for this.

To further quantify this relationship, we can fit a linear model to measure the correlation between estimated and actual audience size. The model does not explain much variance at all: $R^2 = 0.04, p < .001$.

Folk theories of audience

What heuristics are guiding users' estimates, and why are users underestimating so much? To answer these questions, we investigated the theories that participants reported when estimating their audience size. We performed a content analysis on the survey responses for how participants came up with their estimates. The authors inductively coded a subset of the responses to generate categories, then iterated on the coding scheme until arriving at sufficient agreement (Fleiss's Kappa = 0.72).

Theory	Prevalence
Guess	23%
Based on likes and comments	21%
Portion of total friend count	15%
How many friends might log in	9%
Who they regularly see on the site	5%
Number of close friends and family	3%
Who might be interested in the topic	2%
Based on privacy settings	2%
Another explanation given	8%

Table 1: The relative prevalence of folk theories for estimating audience size. Participants most often used heuristics based on the amount of feedback they get on their posts or their number of friends.

The categories and their relative popularities across both surveys appear in Table 1. Nearly one-quarter of participants said they had no idea and simply guessed. The magnitude of this number indicates how little understanding users have of their audience size. The most popular strategy (other than guessing) was based on feedback: the number of likes and comments on a post. Participants explained, *"I figured about half of the people who see it will 'like' it, or comment on it,"* or *"number of people who liked it x4."*

Others made a rough guess at how many people might log in to Facebook: e.g., *"I'm guessing one hundred of my frieds [sic] actually read facebook daily"* and *"not a lot of people stay up late at night."* Many respondents based their estimates on a fixed fraction of their friend count: *"figure maybe a third of my friends saw it."* Others assumed that the people they typically see in their own feeds or chat with regularly are also the audience for their posts: *"Judging by the number of people that regularly share with me"*, *"I assume the number of people who see me are the same people that show up on my news feed."* Finally, some respondents mentioned their close friends or family, or friends who would be interested in the topic of the post: *"I'm sure a lot of people have blocked me because of all the political memes I've been putting up lately,"*, *"Friends that are involved with paintball would catch a glimpse of me mentioning MAO."* A small number mentioned their privacy settings: *"Based on the privacy and sharing rules I have set-up, I would imagine it's close to that number."* Particularly savvy users transferred knowledge from other domains, like *"based on the FB page of a business that i'm admin on."*

We compared the audience estimation accuracy of each group. Despite the variety of theories of audience size, no theory performed better than users who responded "guess".

Perceptions of audience size and satisfaction

Though users have limited insight into their own audience size, their perceptions may play a strong role in their satisfaction with their posts' reach. In this section, we investigate survey participants' responses to the question, *"How many people do you wish saw this piece of content?"*

Table 2 reports summary statistics for our survey results. Only 3% of participants wanted a smaller audience than they

Desired audience	Count	Median perceived audience	Median actual audience
Far fewer people	6	1%	8%
Fewer people	9	19%	28%
About the same	295	9%	26%
More people	145	5%	23%
Far more people	125	5%	22%

Table 2: Summary statistics for survey participants who desired smaller, the same, or larger audiences. Perceived and actual audiences shown as a percentage of friend count.

thought they had, while slightly less than half (270) wanted a larger audience and another half (295) were satisfied with their audience size. The fifteen individuals who wanted smaller audiences had large variance in their responses: those who wanted "far fewer people" had the smallest perceived audience (1%) and those who wanted "fewer people" had the largest perceived audience (19%).

Given the small number of people who wanted smaller audiences, we combined the two groups that desired larger audiences and ended up with two comparison groups: those who were satisfied with their audience size (N=295), and those who wanted a larger audience (N=270). Those who desired the same audience size typically estimated that 9% of their friends would see a post ($mean = 20\%, SD = 32\%$), while those who wanted "more people" or "far more people" typically estimated 5% of their friends (combined $mean = 12\%, SD = 28\%$). A nonparametric Wilcoxon rank-sum test comparing these means is significant ($W = 47179, p < .001$), indicating that those who wanted to reach a larger audience estimated that their reach was smaller.

UNCERTAINTY IN ACTUAL AUDIENCE SIZE
The survey results suggest that people do a poor job of estimating the size of the invisible audience in social networks. Is this inaccuracy inevitable? Is audience size too uncertain to estimate accurately?

In this section, we investigate sources of uncertainty when estimating the invisible audience, broadening our view from survey users to our entire dataset of over 220,000 users. We investigate the audience for individual posts in our dataset, then the cumulative audience size for each user over a month. We demonstrate that there is a great deal of variability in audience size, and that audience size is difficult to infer using visible signals. These results suggest that audience size can be quite unpredictable and that users simply do not receive enough feedback to predict their audience size well.

Audience per post
Each time a Facebook user creates a new post, there is a new audience. This section investigates the variance in actual audience sizes for individual posts. We consider how well we can predict a new post's audience using the visible folk theory signals such as fraction of friends and feedback.

Relationship between friend count and audience size
A user's total friend count is the upper limit on their audience size. Is friend count strongly related to audience size?
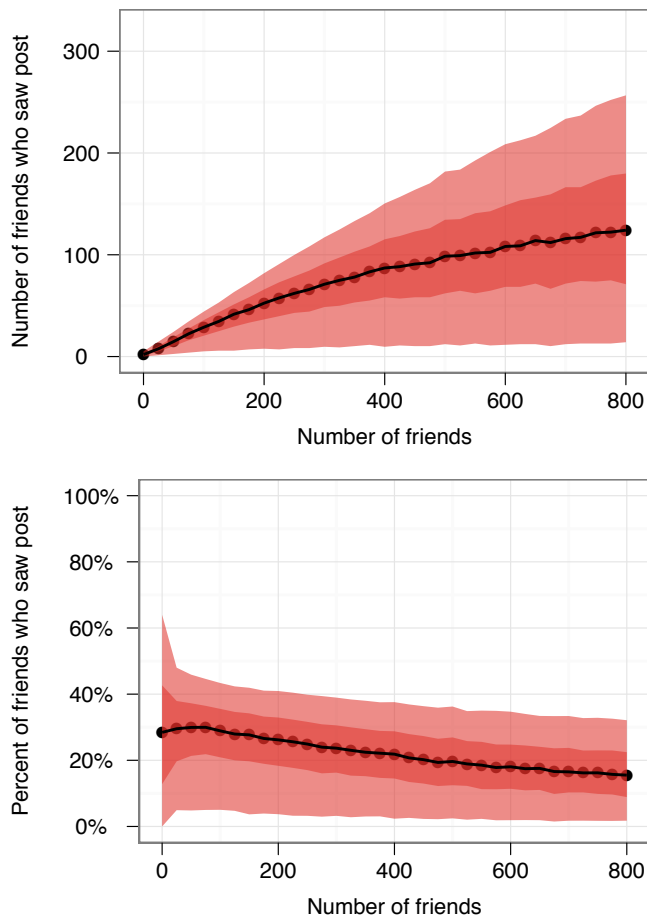

Figure 2: Users with the same number of friends have highly variable audience sizes. Panels show the distribution of the number of friends (top) and fraction of friends (bottom) who saw a post as a function of friend count. The line and bands indicate the median, interquartile range, and 90% region.

Across all posts in our sample, the mean and median fraction of friends that see a post among users is 34% and 35%, respectively. However, this quantity is highly variable. As Figure 2 demonstrates, the interquartile range for the fraction of friends that see a post is approximately 20%, and the 90th percentile range can be as large as 84% (12–96% of the friend count). While the fraction of friends who see a post remains relatively stable, it is important to note that if we consider a user with a fixed number of friends, the actual number of friends who see the post is quite variable. Figure 3 illustrates this relationship: the standard deviation in the audience size increases with the number of friends. As a result, post produced by a user with many friends has more variability in the audience size than one produced by a user with few friends.

Audience size grows most rapidly as a function of number of friends for users with fewer than 200 friends, then tapers off for users with more friends (Figure 2, top). Likewise, the fraction of friends who see a user's post is greatest when users have fewer friends (Figure 2, bottom). This result might be explained by the fact that users with many friends are likely to
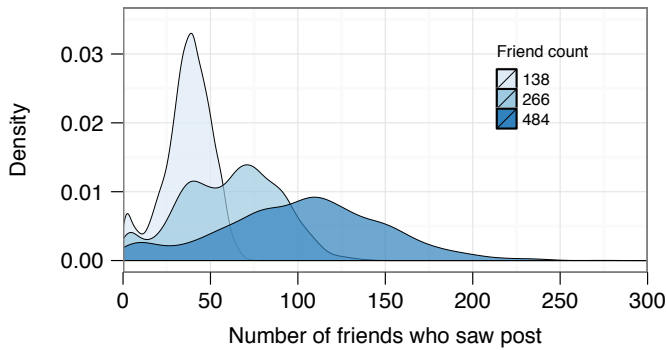
Figure 3: The distribution of audience sizes for users at the 25th, 50th and 75th percentile of friend count within our sample, corresponding to 138, 266, and 484 friends.

have friends who also have large networks [37]; in a system with limited attention, highly connected users may have less opportunity to see all of their friends' content.

To quantify this uncertainty, we can fit a linear model to predict the fraction of friends viewing a post using the number of friends as input. This model explains relatively little variance ($R^2 = .12, p < .001$). The mean absolute error of the model is .08, meaning that the average actual audience is 8% of friends away from the prediction.

Informativeness of feedback
Feedback is the main mechanism that users have to understand their audience's reaction to a post. It was also one of the most popular folk theories in our survey. Does feedback actually help users understand their audience size?

Figure 4 reports the median audience sizes for posts, depending on the amount of feedback they get from friends liking and commenting on the post. Audience size grows rapidly as posts gather more feedback, though this growth slows when a post has feedback from five unique friends. Posts with no likes and no comments had an especially large variance in audience size: the median audience was 28.9% of the user's friends, but the 90% range was from 1.9% to 55.2% of the user's friends. So, while users may be disappointed in posts that receive no feedback, the lack of feedback says little about the number of people it has reached.

Like with friend count, we can fit a linear model that uses feedback indicators such as unique commenters and unique likers to predict the fraction of friends that see a post. This model also explains very little variance ($R^2 = .13, p < .001$). Like the previous model, its mean absolute error is 8%, so the average prediction is off by 8% from the actual percentage of friends who saw the post. These results are nearly identical to those for the linear predictor based on friend count.

Joint model
Neither friend count nor feedback individually are very accurate predictors of audience size. However, they may be contributing data that is jointly informative. We fit a linear model using all three predictors from before — friend count, unique commenters, unique likers — to predict the fraction of
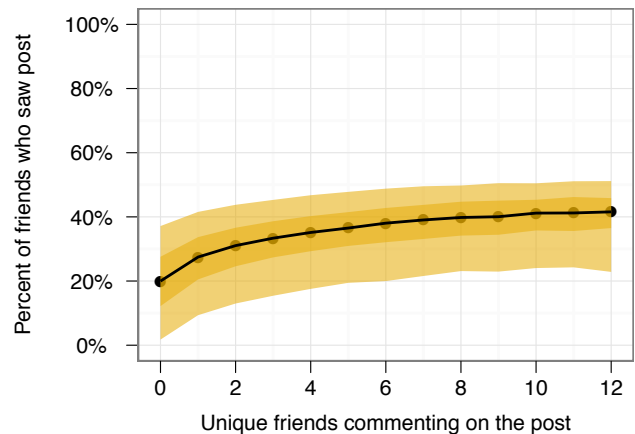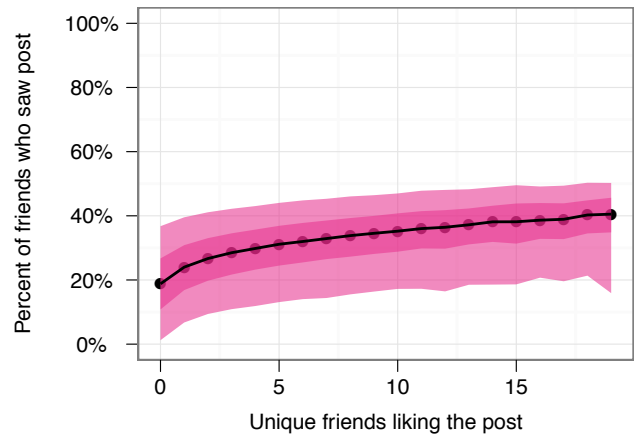

Figure 4: The number of unique friends leaving likes (top) and comments (bottom) is positively associated with audience size, but has large variance.

friends in the audience. This model explained around twice as much variance (*adj.* $R^2 = .27, p < .001$), but this is still not a highly predictive model. The mean absolute error for this model was 7% of the friend network.

So, even using all three signals that users can view, it is difficult to predict actual audience size.

Cumulative audience
Audience size estimates for the general survey were significantly higher than those specific to a single post. Likewise, users evolve their understanding over extended periods of time. To capture these longer-term processes, we analyzed *cumulative audience*: the total number of distinct friends who saw or interacted with each user's content over the span of an entire month. Three quarters of the users in our sample produced more than one piece of content during the month, and half the users in our sample produced five or more pieces of content.

Relationship with friend count
Cumulative audiences are much larger than those for a single post. The median size of a user's cumulative audience is
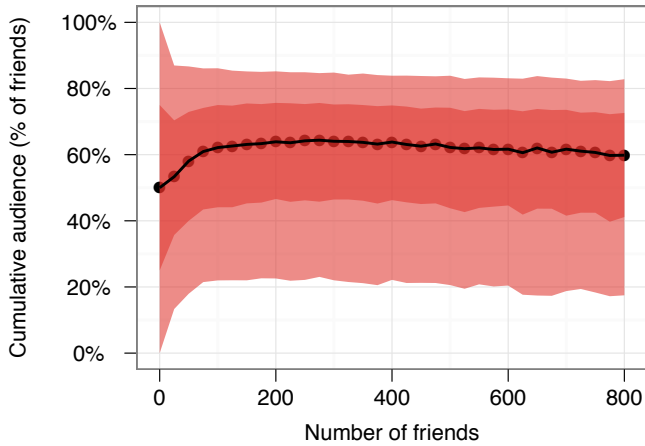
Figure 5: Roughly 60% of users' friends see at least one item they share each month. Cumulative audience size is aggregated over all content produced by a user over 28 days.
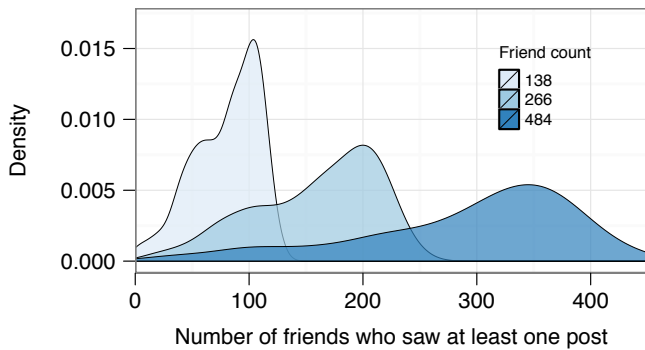

Figure 6: The distribution of cumulative audience sizes across all content produced over one month for users at the 25th, 50th and 75th percentile of friend count in our sample (138, 266, and 484 friends).
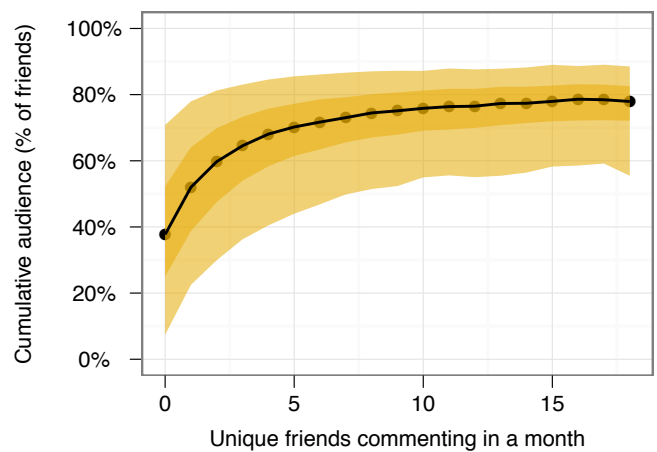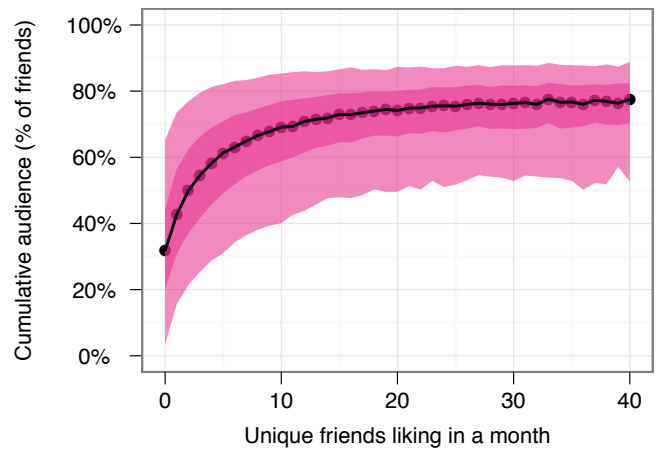

Figure 7: The distribution of the fraction of friends who had seen at least one post by a user as a function of the number of distinct users who had liked (top) or commented on (bottom) at least one piece of the user's content during a one month period. Horizontal axis extends to the 95th percentile of likes and comments for users in our sample.

61% of that user's friends, compared to 35% for a single post. The relationship between friend count and monthly cumulative audience as a fraction of friend count appears in Figure 5. Again, this relationship can be quite variable: the interquartile range is 31% and the 90% range is 65%. The more friends the user has, the larger the variation (Figure 6).

Informativeness of feedback
If users develop a sense of their invisible audience over time by observing the number of friends who provide feedback, how much can they learn about their audience size? Figure 7 shows the fraction of friends that consitute a user's cumulative audience over the span of one month, given a certain number of unique friends who have provided feedback. For example, approximately 60% of a typical (median) user's friends will see at least one of their posts in a month, given that exactly two distinct friends commented on their posts. While some minimal amount of feedback is indicative of wider distribution, users who have greater than 10 friends who like their posts, or 5 friends that comment, all have fairly

large audiences. Furthermore, only a small fraction of users' audiences provide feedback over the month: 95% of the users in our sample have 40 or fewer friends who like their posts, and 18 friends who comment on their posts.

Joint model
Though the cumulative audience is much larger than the audience for single posts, it exhibits no less variance. We trained a linear model to predict audience size, using the same predictors as with previous models (friend count, number of unique commenters, number of unique likers) and one new predictor: the number of stories the user produced that month. The model explains no more variance than the joint model for individual posts ($R^2 = .25, p < .001$).

DISCUSSION
Our analysis indicates that social media users underestimate how many friends they reach by a factor of four. Many users

who want larger audiences already have much larger audiences than they think. However, the actual audience cannot be predicted in any straightforward way by the user from visible cues such as likes, comments, or friend count.

The core result from this analysis is that there is a fundamental mismatch between the sizes of the perceived audience and the actual audience in social network sites. This mismatch may be impacting users' behavior, ranging from the type of content they post, how often they post, and their motivations to share content. The mismatch also reflects the state of social media as a *socially translucent* rather than *socially transparent* system [15, 16]. Social media must balance the benefits of complete information with appropriate social cues, privacy and plausible deniability. Alternately, it must allow users to do so themselves via practices such as butler lies [19].

The mismatch between estimated and actual audience size highlights an inconsistency: approximately half of our participants wanted to reach larger audiences, but they already had much larger audiences than they estimated. One interpretation would suggest that if these users saw their actual audience size, they would be satisfied. Or, these users might instead anchor on this new number and still want a larger audience.

Reasons for the audience mismatch

Why do people underestimate their audience size in social media? One possible explanation is that, in order to reduce cognitive dissonance, users may lower their estimates for posts that receive few likes or comments. A necessary consequence of users *underestimating* their audience is that they must be *overestimating* the probability that each audience member will choose to like or comment on the post. For these posts without feedback, it might be more comfortable to believe that nobody saw it than to believe that many saw it but nobody liked it. Our data lend some support for this belief: users estimated smaller audiences when we showed them a specific recent post than when they considered their posts in general, and we found the same results when we piloted a separate survey that focused users on a specific post they might create in the future.

Several of the folk theories suggest that users are influenced by whom they have interacted with recently. So, some underestimation may also be due to reliance on the *availability heuristic*, which suggests that prominent examples will impact peoples' estimates [36]. If true, the underestimation might be due to the fact that many social systems have more viewers than contributors [30]. Specifically, users might base their estimates on whom they *see* on Facebook, not accounting for those who might be reading but not responding.

Finally, users' estimates may be affected by the ranking and filtering Facebook performs on the News Feed. Other social network sites have different filter designs: Twitter shows content in an unfiltered reverse-chronological order, while Google+ gives users a slider to choose how much content to filter from each circle. One might hypothesize that not filtering users' feeds (e.g., Twitter) might increase diversity and thus increase audience size. However, we believe that filtering plays a more minor role in audience size: algorithmic feed filters might prefer content from strong ties, but the many weak ties in social networks have a large collective influence on users' feed consumption and sharing practices [6].

Limitations

Our study methodology carries several important limitations. Some concerns are associated with the limits of log data. First, techniques for estimating whether a user saw a post have a precision-recall tradeoff: depending on how the instrument is tuned, it might miss legitimate views or count spurious events as views. However, the instrument would need to be overestimating audience size by a factor of four for our conclusions to be threatened, and we believe that this is unlikely. Second, just because a user saw a status update, we cannot be certain that they focused their attention on it or would remember it [13]. Our survey attempted to be clear that we were interested in how many people *saw* the content, but some participants may have blurred the distinction. Finally, we focused on one month of data; changes to site features or norm evolution may impact audiences over longer periods.

Our survey methodology has limitations as well. To validate the instrument, we tested different formulations of the audience estimation question. Variations included comparing "saw the content" to "read the content", as well as different wordings asking for raw numbers, percentages, and Likert-style radio button responses. The results were similar for all formulations. However, it is possible that some participants still filtered out users who they think might not have paid attention to the content. In addition, the survey format did not go into depth on participants' perceptions: interviews could further explore this phenomenon.

Our methods introduce sampling bias. Our log data selects for active users, in particular those who produce content. The survey also selects for users who produce content, those who log in to Facebook, and those who care enough about Facebook to participate in a survey advertised at the top of the News Feed. One known bias is that active users can have larger friend counts than the typical Facebook population.

Design implications

This research raises the question of whether showing actual audience information might benefit social media. This design might come in many flavors: highlighting close friends who saw the post, showing a count but no names or faces, or (in the extreme) showing a complete list of every person who saw the post. Some Facebook groups now display how many group members have seen each post, while sites such as OkCupid show who has browsed your profile. Adding audience information could certainly address the current mismatch between perceived and actual audience size.

Our results do not paint a clear picture as to whether audience information would be a good addition. Some measure of social translucence and plausible deniability seems helpful: audience members might not want to admit they saw each piece of content, and sharers might be disappointed to know that many people saw the post but nobody commented or "Liked"

it. Underestimating the audience might also be a comfortable equilibrium for some users who feel more comfortable speaking to a relatively small group and would not post if they knew that they were performing in front of a large audience. On the other hand, many users expressed a desire for a larger audience, and demonstrating that they do in fact have a large audience might make them more excited to participate.

Pragmatically, this work suggests that social media systems might do well to let their users know that they are impacting their audience. Because so many members of the audience provide no feedback, there may be other ways to emphasize that users have an engaged audience. This might involve emphasizing cumulative feedback (e.g., 400 friends saw their content last month), showing relative audience sizes for posts without sharing raw numbers, or encouraging alternate modes of feedback [3].

CONCLUSION

We have demonstrated that users' perceptions of their audience size in social media do not match reality. By combining survey and log analysis, we quantified the difference between users' estimated audience and their actual reach. Users underestimate their audience on specific posts by a factor of four, and their audience in general by a factor of three. Half of users want to reach larger audiences, but they are already reaching much larger audiences than they think. Log analysis of updates from 220,000 Facebook users suggests that feedback, friend count, and past audience size are all highly variable predictors of audience size, so it would be difficult for a user to predict their audience size reliably. Put simply, users do not receive enough feedback to be aware of their audience size. However, Facebook users do manage to reach 35% of their friends with each post and 61% of their friends over the course of a month.

Where previous work focuses on publicly visible signals such as reshares and diffusion processes, this research suggests that traditionally invisible behavioral signals may be just as important to understanding social media. Future work will further elaborate these ideas. For example, audience composition is an important element of media performance [27], and we do not yet know whether users accurately estimate which individuals are likely to see a post. We can also isolate the causal impact of estimated audience size on behavior, for example whether differences in perceived audience size cause users to share more or less. Finally, these results suggest that there are deeper biases and heuristics active when users estimate quantities about social networks, and these biases warrant deeper investigation.

ACKNOWLEDGMENTS

REFERENCES

1. A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy enhancing technologies*, pages 36–58. Springer, 2006.

2. I. Altman. *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding.* ERIC, 1975.

3. P. André, M. Bernstein, and K. Luther. Who gives a tweet?: evaluating microblog content value. In *Proc. CSCW 2012*, 2012.

4. L. Backstrom, E. Bakshy, J. Kleinberg, T. Lento, and I. Rosenn. Center of attention: How facebook users allocate attention across friends. In *Proc. ICWSM 2011*, 2011.

5. E. Bakshy, J. Hofman, W. Mason, and D. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proc. WSDM 2011*, 2011.

6. E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, 2012.

7. M. Bernstein, A. Marcus, D. Karger, and R. Miller. Enhancing directed content sharing on the web. In *Proc. CHI 2010*, 2010.

8. d. boyd. Friends, friendsters, and myspace top 8: Writing community into being on social network sites. *First Monday*, 11(12), 2006.

9. d. boyd. Social network sites: Public, private, or what? *Knowledge Tree*, 13(1):1–7, 2007.

10. M. Burke, C. Marlow, and T. Lento. Feed me: motivating newcomer contribution in social network sites. In *Proc. CHI 2009*, 2009.

11. K. Caine, L. Kisselburgh, and L. Lareau. Audience visualization influences disclosures in online social networks. In *Ext. Abst. CHI 2011*, 2011.

12. H. Clark and G. Murphy. Audience design in meaning and reference. *Advances in psychology*, 9:287–299, 1982.

13. S. Counts and K. Fisher. Taking it all in? visual attention in microblog consumption. *Proc. ICWSM 2011*, 2011.

14. N. Ellison et al. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

15. T. Erickson and W. Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM TOCHI*, 7(1):59–83, 2000.

16. E. Gilbert. Designing social translucence over social networks. In *Proc. CHI 2012*, 2012.

17. E. Goffman. The presentation of self in everyday life. *Garden City, NY*, 1959.

18. S. Gosling, S. Gaddis, S. Vazire, et al. Personality impressions based on facebook profiles. In *Proc. ICWSM 2007*, 2007.

19. J. Hancock, J. Birnholtz, N. Bazarova, J. Guillory, J. Perlin, and B. Amos. Butler lies: awareness, deception and design. In *Proc. CHI 2009*, 2009.

20. B. Hogan. The presentation of self in the age of social media: distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, 30(6):377–386, 2010.

21. P. Killworth, E. Johnsen, H. Bernard, G. Ann Shelley, and C. McCarty. Estimating the size of personal networks. *Social Networks*, 12(4):289–312, 1990.

22. H. Kwak, H. Chun, and S. Moon. Fragile online relationship: a first look at unfollow dynamics in twitter. In *Proc. CHI 2011*, 2011.

23. C. Lampe, N. Ellison, and C. Steinfield. A familiar face (book): profile elements as signals in an online social network. In *Proc. CHI 2007*, 2007.

24. C. Lampe, N. Ellison, and C. Steinfield. Changes in use and perception of facebook. In *Proc. CSCW 2008*, 2008.

25. E. Lieberman and R. Miller. Facemail: showing faces of recipients to prevent misdirected email. In *Proc. SOUPS 2007*, 2007.

26. E. Litt. Knock, knock. who's there? the imagined audience. *Journal of Broadcasting & Electronic Media*, 56(3):330–345, 2012.

27. A. Marwick and d. boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133, 2011.

28. C. McCarty, P. Killworth, H. Bernard, E. Johnsen, and G. Shelley. Comparing two methods for estimating network size. *Human Organization*, 60(1):28–39, 2001.

29. R. Metz. Friendster outs voyeurs. *Wired*, 2005.

30. B. Nonnecke and J. Preece. Lurker demographics: Counting the silent. In *Proc. CHI 2000*, 2000.

31. L. Palen and P. Dourish. Unpacking privacy for a networked world. In *Proc. CHI 2003*, 2003.

32. B. Suh, L. Hong, P. Pirolli, and E. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Proc. SocialCom 2010*, 2010.

33. J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh. Who's viewed you?: the impact of feedback in a mobile location-sharing application. In *Proc. CHI 2009*, 2009.

34. Z. Tufekci. Can you see me now? audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society*, 28(1):20–36, 2008.

35. Z. Tufekci. Facebook, youth and privacy in networked publics. In *Proc. ICWSM 2012*, 2012.

36. A. Tversky and D. Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.

37. J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The anatomy of the facebook social graph. *CoRR*, abs/1111.4503, 2011.

38. F. Viégas. Bloggers' expectations of privacy and accountability: An initial survey. *Journal of Computer-Mediated Communication*, 10(3), 2005.

39. F. Viégas and J. Donath. Chat circles. In *Proc. CHI 1999*, 1999.

40. A. Young and A. Quan-Haase. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proc. C&T 2009*, 2009.