Center of Attention: How Facebook Users Allocate Attention across Friends

Lars Backstrom
Facebook
lars@fb.com

Eytan Bakshy
Univ. of Michigan & Facebook
ebakshy@fb.com

Jon Kleinberg
Cornell University
kleinber@cs.cornell.edu

Thomas M. Lento
Facebook
tlento@fb.com

Itamar Rosenn
Facebook
itamar@fb.com

Abstract

An individual's *personal network* — their set of social contacts — is a basic object of study in sociology. Studies of personal networks have focused on their size (the number of contacts) and their composition (in terms of categories such as kin and co-workers). Here we propose a new measure for the analysis of personal networks, based on the way in which an individual divides his or her attention across contacts. This allows us to contrast people who focus a large fraction of their interactions on a small set of close friends with people who disperse their attention more widely.

Using data from Facebook, we find that this balance of attention is a relatively stable property of an individual over time, and that it displays interesting variation across both different groups of people and different modes of interaction. In particular, activities based on communication involve a much higher focus of attention than activities based simply on observation, and these two modalities also exhibit different forms of variation in interaction patterns both within and across groups. Finally, we contrast the amount of attention paid by individuals to their most frequent contacts with the rate of change in the actual identities of these contacts, providing a measure of *churn* for this set.

1 Introduction

People maintain a broad range of personal relationships. In the language of social networks, these relationships can be thought of as the links connecting an individual to her network neighbors, a set of people we will refer to as her *contacts*. A significant body of research in sociology has focused on an individual's contacts — her *personal network* — as an important attribute in settings that range from professional opportunities (Granovetter 1973; Burt 1992) to social support and advice on important matters. (Fischer 1982; McPherson, Smith-Lovin, and Brashears 2006; Wellman and Wortley 1990).

This line of work has considered variations in both the *size* and the *composition* of personal networks. Size is most naturally defined simply as the number of contacts (Killworth et al. 1990). Composition has generally been studied in terms of discrete variables that include the number of kin and non-kin contacts, and the distinction between close friends and more distant acquaintances. Earlier research has considered how the composition of personal networks differs across attributes including age, race/ethnicity, gender, and educational level (Hartup and Stevens 1997; Marsden 1987; McPherson and Smith-Lovin 1993; Moore 1990) while more recent work has examined personal network composition within the context of social media (Chang et al. 2010; Gilbert, Karahalios, and Sandvig 2008).

In this paper, we propose a new measure for analyzing personal networks that addresses a dimension distinct from network size and composition. This measure expresses the way in which an individual divides his or her attention across contacts. Everyday experience suggests that some people focus most of their attention on a small circle of close friends, while others disperse their attention more broadly over a large set. As a specific property of an individual, this contrast between focus and dispersion — the individual's *balance of social attention* — is distinct from the properties discussed above: two people with personal networks of similar size and composition can differ greatly in the extent to which their attention is focused on a small or large subset of their personal network. Furthermore, the balance of social attention is not a purely structural measure, since it takes into account both the links in the underlying social network and the amount of time that an individual allocates to these links.

We believe this type of measure can play a useful role in illuminating the fine structure of an individual's personal network in both on-line and off-line settings. An understanding of how the balance of attention varies across individuals can also help to inform the design of social media applications, many of which must manage a tradeoff between diversity and relevance. These applications attempt to avoid stale content, while at the same time ensuring that everything that appears is personally relevant. Designers of such social products can use an individual's balance of social attention to help customize this tradeoff on a per user basis. For example, stratification of users by a measure capturing balance of social attention recently led to increased interaction with the Facebook News Feed.

Although a metric for balance of social attention is potentially useful and theoretically interesting it has been difficult to study empirically. Even the size and composition of friendship networks are notoriously difficult to measure, and generally have been captured through self-reports aided by elicitation mechanisms (Campbell and Lee 1991). Measuring the balance of attention requires an even higher resolution, as it depends on a careful estimation of the volume of interaction between an individual and each member of her personal network. In order to overcome these measurement difficulties we use data from Facebook to analyze the interaction volume. After reviewing further related work, we turn in the next section to a precise formulation for the balance of social attention. Subsequent sections present analysis that shows how this measure exhibits interesting patterns of variation across groups of people and across different modalities of interaction.

Further Related Work. Recent work in on-line social networks has articulated the contrast between the links in a network and the activity that takes place on these links. This is also the distinction that motivates our work, although our focus differs from earlier papers to address this issue: Kossinets et al. (2008) study how link activity can lead to different pathways for information flow over multi-step paths, and Wilson et al. (2009) focus on aggregate measures for how activity is distributed, and the network structures that result from thresholding the links by activity level. In contrast, we are interested in the distribution of attention levels as an attribute operating at the individual level — in understanding how this attribute varies across people and groups, and how it relates to other individual attributes.

From a theoretical perspective the balance of social attention is related to the distinction between strong and weak ties (Granovetter 1973), but this is not simply a different measure of tie strength. Although tie strength is ultimately a synthesis of several factors, including volume of interaction and affective closeness (Marsden and Campbell 1984), our measure begins from the aspects of tie strength related to volume and synthesizes them into a node-level measure in the network that takes into account an individual's full set of ties. Furthermore, our approach also relates to arguments by Milgram (1970) and Mayhew and Levinger (1976) that settings such as dense urban areas, which produce many interactions ought to result in less time spent on any one of these interactions. Our measure enriches these considerations by formulating multiple ways in which an individual can manage a large personal network: either by slicing her attention relatively evenly over all contacts, or by focusing on a few at the expense of the others. Finally, our measure is related to other quantitative trade-offs between focus and dispersion in an individual's personal network, such as the geographic spread of one's friends and the searchability of social networks (Kleinberg 2006; Backstrom, Sun, and Marlow 2010). The focus of our work is to quantify this trade-off in terms of the volume of interaction, rather than embedding the analysis in external frames of reference such as geography or social categories.

2 The Balance of Social Attention

Consider a population of n individuals, and a person i in this population who sends messages to her contacts. (Later we will consider a range of different interaction modalities, but for purposes of exposition it is useful to think about messages.) Suppose m_j is the number of messages sent by person i to person j in her set of contacts. If the total number of messages sent by i (over all contacts) is m, we say that the fraction of i's attention that she devotes to j is $a_j = m_j/m$.

As a function of k, what fraction of i's attention does she devote in total to her k most frequent contacts? We sort all of i's contacts j in order of decreasing a_j, and we say that i's *top k contacts* are the people corresponding to the first k positions in this sorted list. The fraction of i's attention devoted to her top k contacts, denoted f_k, is the sum of a_j over all individuals j in this set of top k contacts. If i has n contacts, then the vectors $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ are each complete descriptions of how i divides her volume of interaction across her contacts, and these vectors serve as our starting point for measuring the balance of social attention.

The full vectors turn out to be a highly redundant representation. For much of our analysis, we find that individual coordinates of the vector $\mathbf{f}$ can serve as relatively stable summaries of aggregate properties computed from the full vector. Specifically, if we compare individuals simply by the single number f_k, we get extremely similar aggregate comparison results for all k in a broad middle range where most of the volume of interaction takes place, i.e., in the interval between $k = 5$ and $k = 25$. There is a natural reason for this: in general, if user A has lower f_k value than user B, then A will also typically have a lower f_ℓ value compared to B, when k and ℓ are in this middle range from 5 to 25. As a result, any coordinate from this range produces roughly similar results, which allows us to collapse a collection of measurements to something that is effectively a single-dimensional question.

3 Balance of Attention Across Modalities

To examine how an individual balances her attention across her friends, we compute metrics for a number of different modalities of attention. These modalities can be divided into two distinct groups: communication and viewing. The communication modalities encompass directed interaction, such as sending a private message or posting a public comment on a photo, while viewing behavior is derived from users visiting pages on Facebook. Thus, in the communication modalities the target is aware of the user's actions (since they receive the communication), but in the viewing modalities they are not; only the user is aware of the viewing activity. (See also Jiang et al. (2010) for further discussion of this contrast in on-line social networks.)

- *Messages.* Individuals can send each other private messages similar to email.
- *Comments.* When a user shares a piece of content, such as posting a photo or a link, other users can typically leave a public comment on the item.
- *Wall Posts.* A user's Facebook profile includes a publicly viewable 'wall', on which other users can post content.
- *Profile Views.* This measures how many times one user views another's profile page.

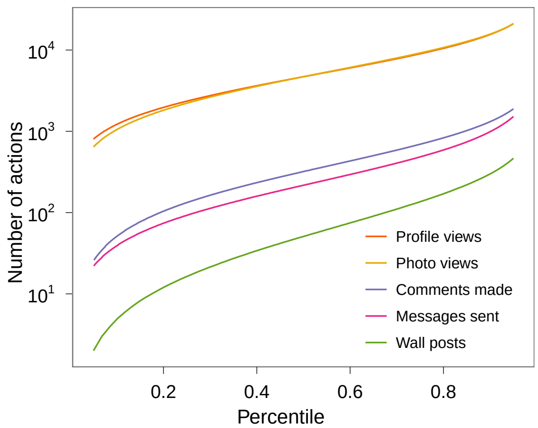

Figure 1: Distribution of volume of activity per modality.

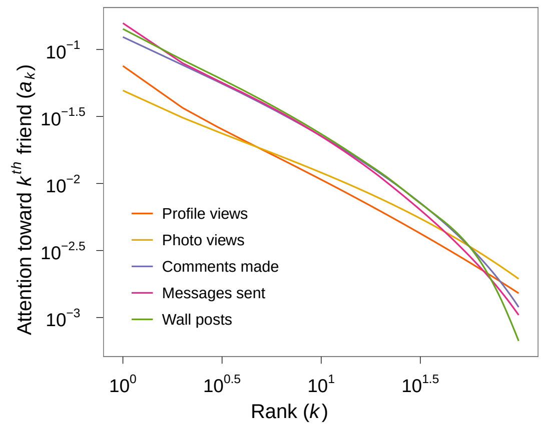

Figure 2: Fraction of attention devoted to a given contact, based on the contact's rank in terms of overall volume of attention received.

- *Photo Views.* This measures how many times a user views photos posted by another user.

An individual might focus her attention toward differing subsets of her contacts through each of the modalities mentioned above. Therefore, we compute attention measurements independently for each modality by collecting the sum of all actions for each user-target pair within each modality from January 2010 to December 2010. All data has been anonymized and aggregated prior to analysis. For comments and messages, each individual post counts as a single action directed at a given target. For wall posts we consider only the subset of items posted outside of the target user's birthday window, defined as the time span from two days prior through one day after the target's birthday.[1]. We exclude birthday wall posts from our measurements because they are typically triggered by a birthday reminder on Facebook rather than some user-specific mechanism, and are therefore not representative of the directed attention the communication modalities generally capture. The viewing modalities require the user to make a direct navigation to view a target user's profile page or a photo owned by the target user. Simply encountering a target user or a target user's photo in the News Feed or on another user's profile does not constitute a view. Henceforth, when we talk about a user's level of *activity* in a given modality, we refer to the number of discrete actions the user performed in this modality, as measured according to the definitions above.

In order to minimize the impact of behavioral trends related to the overall growth of Facebook, we restrict our analysis to users who were already members as of January 1, 2009. In addition, we are interested in measuring the behavior of users for whom Facebook represents a non-trivial medium of communication and social attention, so that we can see the balance of attention among people in a context where this is a relevant quantity. Therefore, we select only those Facebook users who have visited the site on at least 80% of the days in 2009 and 2010. This user sample represents a population of 16 million heavily active Facebook users.

Figure 1 shows the distribution of activity for each modality, with the percentile rank within the modality along the x-axis and the total number of actions within each modality on the y-axis. A given user's volume of viewing actions is likely to be an order of magnitude higher than her volume of communication actions, while non-birthday wall posts are least common. As even active users may not use some of the features (for instance, 27% of these active users users sent less than 100 messages in a year), some of our subsequent analyses will further restrict the user set studied.

The average balance of attention. Figure 2 shows the fraction a_k as a function of k for the five different modalities. We only consider the users in the 70^{th} to 95^{th} percentiles of activity level for each modality. In doing so, we filter out the individuals who do not significantly use each modality as well as the extreme outliers at the top end.

The communication curves begin somewhat higher than the viewing curves, but tail off more quickly at the higher ranks. This happens because many users, even within this relatively active set, have not communicated with more than 50 unique targets, causing us to average in zeros. All of the communication modalities and profile viewing have very similar slopes for low k in a log-log plot, each fitting Cx^{α} for α between 0.75 and 0.78. Thus, while the viewing modalities account for an order of magnitude more activity than the communication modalities, and the quantities $a_1, a_2, a_3, \ldots$ are smaller in absolute terms, they fall off at a very similar rate (proportional to about $k^{-3/4}$) for both viewing and communication. The one modality that behaves differently here is photo views and one possible explanation is that the viewing target is less clear: user A might look at a photo created by user B not because of interest in B, but because of interest in one of the photo's subjects.

[1]This time span represents the average time window in which the number of wall posts received is significantly higher than normal

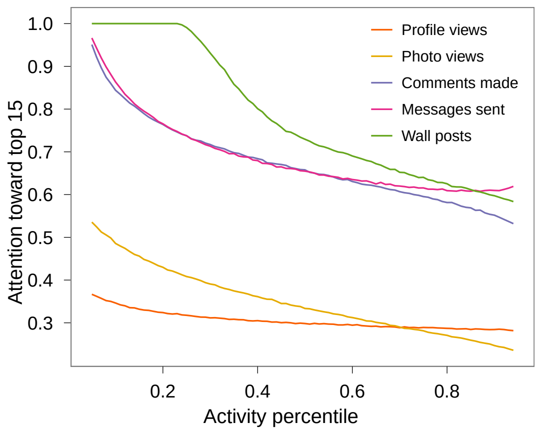

Figure 3: Average fraction of attention devoted to top 15 contacts within a given modality, against level of activity

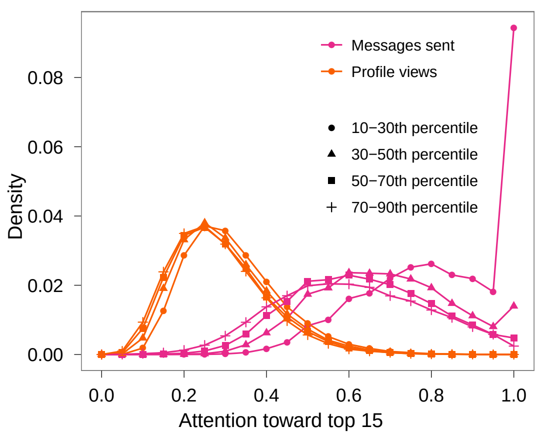

Figure 4: Distribution of fraction of attention devoted to top 15 contacts for messaging and profile views, broken down by activity level.

Although the curves in Figure 2 are restricted to users in the upper percentiles of activity level, they still aggregate over users with varying activity levels, which may hide the importance of a user's overall activity level in impacting the shape of her attention curve. Indeed, the impact of overall activity on these curves is not immediately clear. It may be that people who communicate more are doing so because they communicate more with their lower ranked contacts. But it could also be the case that most people are only capable of maintaining a small number of direct contacts, and increased activity occurs mostly within this fixed set.

To understand the impact of activity on attentional balance, we consider the fraction of activity f_{15} as a function of activity level. (Results for f_k are very similar for all k in the range between 5 and 25, and somewhat beyond this as well.) In order to enable comparisons between modalities like messaging (which a typical user performs a few hundred times a year) and activity types like profile viewing (which occur in the thousands), we examine f_{15} as a function of a user's

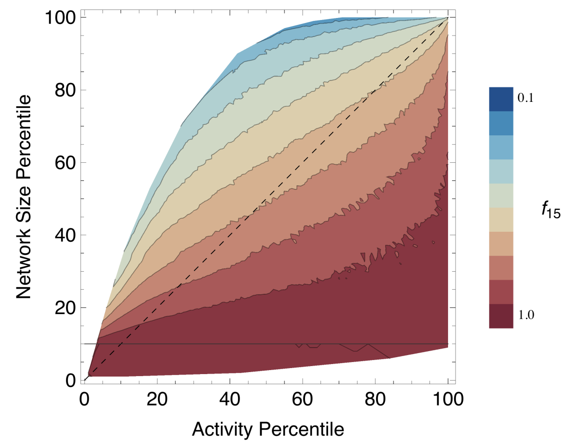

Figure 5: Average fraction of messages sent to top 15 contacts as a function of network size and activity. Horizontal line indicates the percentile at which the number of contacts exceeds 15.

percentile rank of activity level within each modality. Figure 3 shows that within each modality, there is a sharper initial decrease for users at low activity levels, but then a long section that is relatively more gradual. Indeed, while users with low activity necessarily have high f_{15} (those with 15 or fewer contacts have $f_{15} = 1$), the middle region of activity levels decreases quite slowly, and in the case of messaging and especially profile views is approximately flat. One can also look at the distribution of f_k values at a given activity level rather than their mean: Figure 4 shows this for varying levels of activity within messaging and profile viewing. While higher activity groups are slightly left-shifted, the distributions are qualitatively similar once we move beyond the 30^{th} percentile (many below the 30^{th} percentile have messages to 15 or fewer people), and are nearly identical for profile views.

These graphical comparisons show evidence of a broad distinction between viewing and communicating modalities. In general, communication is much more focused, with a high fraction going towards top contacts, while viewing is significantly more dispersed across contacts.

In addition to a user's activity level within a given modality, the size of her personal network within that modality will also affect the value of f_{15}. Naturally, as the network size increases for a fixed activity level, f_{15} tends to decrease, since the individuals added to the network must receive some share of this fixed activity. On the other hand, Figure 5 shows that among users with comparable personal network size, those with higher activity level are more focused. Thus, larger networks tend to lead to smaller f_{15}, while more activity tends to lead to larger f_{15}. Due to the high correlation (0.83 for messaging, 0.91 for profile viewing) between network size and activity level, this effect is lost when looking at f_k only as a function of activity level. The relationship between f_k and individual characteristics, including network size and activity level will be discussed in the following section.

4 Variation By Individual Characteristics

Variation Across Individuals The distributions in Figure 4 show that, even for a fixed activity level, some individuals seem significantly more focused than others in their attention. Although it's possible that this variation arises primarily from the inherent randomness in all of our interactions over time, the more intriguing possibility is that some individuals are genuinely more focused or dispersed than others, and that these differences persist over time.

In order to examine whether users who are active in two distinct time periods have consistent attention patterns across both observation windows, we compare data from early 2010 and late 2010. A simple regression that attempts to predict a user's f_5 value in Oct-Dec 2010 from just her f_5 in Jan-Mar 2010 yields R^2 values of 0.45 and 0.23 for viewing and messaging that — while relatively modest in absolute terms — show a non-trivial level of stability in this quantity over time. This is all the more notable given that this computation has access only to this single f_5 number for predicting the corresponding value close to a year later. Moreover, using only the user's activity level and personal network size in Jan-Mar 2010 performs worse at predicting f_5 in Oct-Dec 2010 than simply using the Jan-Mar f_5 by itself for this prediction (0.23 vs. 0.19 and 0.45 vs. 0.31).

Age and Gender. Figure 6 shows the average value of f_{15} for users between the ages of 13 and 60. We restrict to users in 70-95$^{\mathrm{th}}$ percentiles of activity in each modality, for the reasons discussed above. Each modality exhibits a roughly monotonic relationship with age, but the relationship for viewing moves in the opposite direction of communication: we find that older users are more focused in their viewing behavior, but more dispersed in their communication behavior. Moreover, these two directions of change appear at different rates as we consider older users: the decreasing focus in communication is rapid over ages ranging roughly from 13 to 30, with slower changes beyond this point, while the increasing focus in viewing is much steadier over the full range of ages considered.

Compared to males, females tend to focus more of their attention toward their top k friends in all modalities (Figure 7). This difference may be partly explained by differences in the underlying distribution of activity and network size for each gender. For example, female Facebook users tend to maintain larger active networks than their male counterparts.[2] To adjust for this, we perform a regression analysis to explain the relationship between f_5, number of contacts, activity level, age, and gender (Table 1), using data from all users that self-report their gender and fall within the $5^{th} - 95^{th}$ percentile in terms of total activity and number of contacts. The regression indicates that the fraction of attention allocated to the top five friends depends to a large extent on the number of contacts and activity level of an individual: more contacts are associated with lower values of f_k, and this effect is balanced by higher levels of activity (as seen in Figure 5). For a given level of activity and number of contacts, gender and age both have a small but significant effect. For example, a male user with the same activity level and number of contacts as a female would be expected to have a f_5 score that is 0.02 higher than that of a female. Thus, while there are significant differences in f_5 that depend solely on age and gender, the primary effect on f_5 seems to stem from the fact that total activity and number of contacts vary significantly with age and gender. Note that in some cases, the coefficients on age and gender appear to contradict Figure 7 and Figure 6. However, this is not a contradiction and shows that, for example, while older users tend to be less focused in their messaging overall, comparing users with the same activity level and number of contacts, the older ones will be slightly more focused.

[2]See http://www.facebook.com/note.php?note_id=55257228858 for a comparison

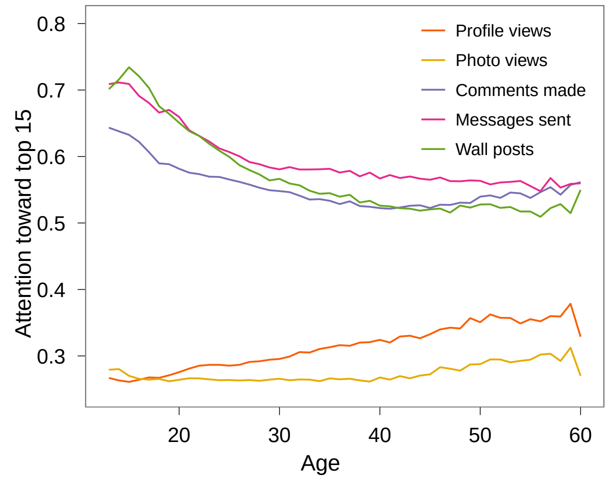

Figure 6: f_{15} as a function of age, for users in the 70th to 95th percentile of activity in the given modality.

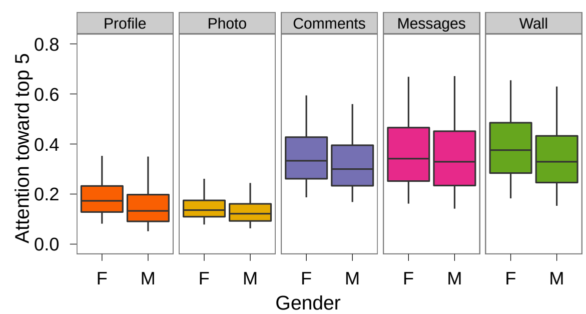

Figure 7: Distribution of attention given to top 5 friends for females and males, for users in the 70th to 95th percentile of activity in the given modality.

Interactions Within and Between Genders. In general, the structure of social ties among people of the same gender is quite different than the structure of social ties across genders (McPherson, Smith-Lovin, and Cook 2001). Thus,

Modality	$inter$	con	act	age	$male$	R^2
Profile	0.18	-0.53	0.44	0.03	0.02	0.38
Photo	0.20	-0.47	0.21	-0.01	0.01	0.53
Comment	0.43	-0.81	0.41	-0.03	-0.01	0.67
Message	0.44	-0.87	0.48	0.03	0.00	0.59
Wall	0.51	-1.48	0.92	-0.02	0.00	0.62

Table 1: Regressions explaining the variation in the fraction of f_5 as a function of individual characteristics ($N = 1,037,885$) for different modalities. Activity (act) and number of contacts (con) are centered percentiles within each modality, and range between -0.45 and 0.45. Age is given in terms of centered percentiles, with -0.5, -0.25, 0, -0.25, and 0.5 corresponding to 13, 21, 25, 33, 65 years, respectively. The intercept ($inter$) shows the expected f_5 score for a 25 year old female with a median number of contacts and activity level. All coefficients are significant at the $p < 10^{-16}$ level and have standard errors that are at least two orders of magnitude less than the coefficient.

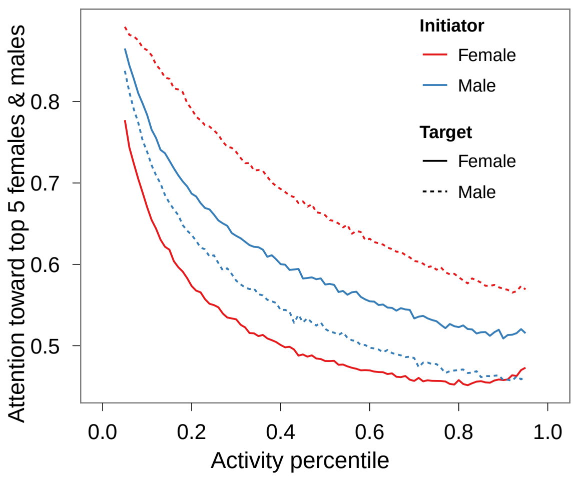

Figure 8: f_5 for messages, fixing the gender of both the initiator and the target of the action.

we further refine the gender analysis to separately consider the interaction of users within their own gender and across genders. We find that women send 68% of their messages to women, while men send only 53% of their messages to women. This distinction is consistent with *gender homophily* — in which each gender has a bias toward within-gender communication — modulated by the overall distribution of Facebook messages. On the other hand, we see much smaller differences in viewing: for typical activity levels, both women and men direct roughly 60% of their profile viewing activity to female users.

We then examine how a person balances their social attention separately to their contacts of each gender. That is, we partition each user's set of actions into two subsets — one for the actions directed at women, and one for the actions directed at men — and then compute the quantities a_k and f_k separately for these subsets. Figure 8 shows the results for messaging: the average f_5 value for

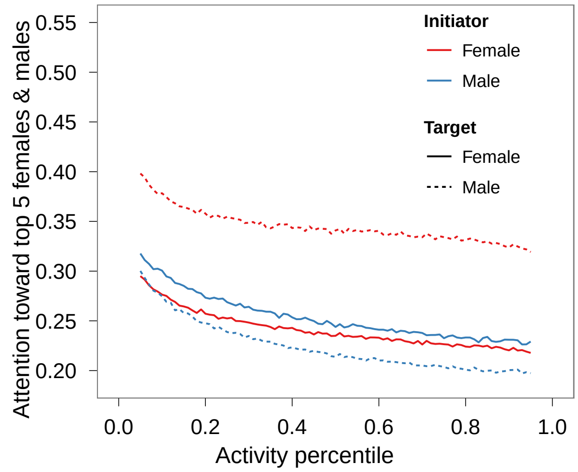

Figure 9: f_5 for profile views, fixing the gender of both the initiator and the target of the action.

Gender Type	Rel. Status	f_5 (Msg.)
women-women	married	0.476
women-women	single	0.523
women-women	relationship	0.524
men-men	married	0.570
men-men	relationship	0.578
men-men	single	0.584
men-women	single	0.631
men-women	married	0.637
women-men	single	0.663
men-women	relationship	0.678
women-men	relationship	0.700
women-men	married	0.715

Table 2: Focus in messaging, grouped by gender and relationship status.

actions by users of gender X toward users of gender Y, for each choice of X and Y. We see that there is greater concentration in across-gender communication than within-gender communication. Furthermore, women are more concentrated than men with respect to across-gender communication, and more dispersed than men with respect to within-gender communication. Viewing behaviors provide (in Figure 9) an interesting contrast with messaging: women and men have roughly equivalent levels of focus in viewing profiles of female users, but markedly differing levels of focus in viewing profiles of male users, where female viewers are much more focused.

Relationship Status. The effect of gender on interaction patterns is further influenced by factors such as marital status — unmarried people display different network structures than married ones (McPherson and Smith-Lovin 1993). To understand the effect of these factors, we consider the subset of active users in our population whose listed *relationship status* on Facebook remained unchanged throughout 2010 and was set to one of the following three values: single, in

Gender Type	Rel. Status	f_5 (View)
women-women	married	0.225
women-women	relationship	0.225
men-men	married	0.225
men-men	relationship	0.227
men-men	single	0.228
men-women	single	0.232
men-women	married	0.242
women-women	single	0.244
men-women	relationship	0.274
women-men	single	0.311
women-men	married	0.329
women-men	relationship	0.364

Table 3: Focus in profile viewing, grouped by gender and relationship status.

a relationship, or married.[3] Therefore, we consider 12 categories of behavior for each modality: for users of gender X and relationship status S we look at the balance of attention in their interactions with users of gender Y. Table 2 shows the average f_5 values for these 12 categories for messaging, and Table 3 shows them for profile viewing. For messaging, we see a refinement of the gender homophily effects in Figure 8. For viewing, a striking "non-monotonic" effect shows up clearly in interactions across the genders: for both women viewing men's profiles and men viewing women's profiles, the level of focus for single and married users is roughly the same, while the focus for users in a relationship is significantly higher than either of these.

5 Attention Over Time

We have shown that the fraction of attention to users' top k contacts, f_k, decreases as a function of activity when averaged over all individuals. One might expect that as a result the top k contacts will tend to change more rapidly over time for those with higher activity. However, we find that increased levels of activity are actually associated with higher levels of stability over time. We examine the overlap between a user's top k contacts in two consecutive time periods; Jan-Feb 2010 constitute the first time period, and Mar-Apr 2010 form the second period. We find the relationship between activity level, number of contacts, and overlap to be qualitatively similar for ranges of k between 1 and 20; we report on $k = 10$ for concreteness.

Stability and Activity Figure 10 shows the overlap between the two time periods as a function of activity. Although we found in Section 3 that users' attention to their top k contacts was lowest for profile and photo views, we find that they are among the highest in terms of overlap. We also find differences between modalities of similar total volume and aggregate f_k values; for example, commenting exhibits significantly higher overlap than messaging, and in fact its

[3]Note that since most relationships in a broad population are heterosexual (Black et al. 2000), such relationships will be the bulk of our computed averages.

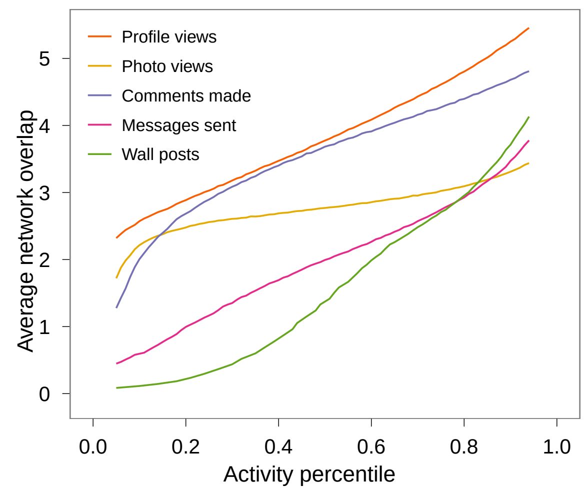

Figure 10: Overlap between top 10 users from January & February to March & April as a function of activity. Higher activity levels are associated with greater stability in users' networks over time.

overlap is quite similar to that of profile viewing. It is an interesting question to consider the possible bases for these contrasts and similarities; one possibility is that a large fraction of profile views are initiated from the news feed. While the news feed may in part be responsible for the stability of top contacts, the act of sending messages or leaving wall posts are not directly affected by stability introduced by the Facebook news feed. We find that the messages and wall posts tend to have the greatest churn, although higher levels of activity lead to relatively large increases in stability.

Regression Analysis of Modal Stability To further explore the predictability of network churn and the relationship between network size and activity, we perform a regression analysis for each modality shown in Figure 10. We consider modalities independently, and attempt to predict the fraction of each user's top 10 neighbors that persist between two time periods. We use two basic properties of users' networks that factor into attention: activity level and network size.

The regressions are summarized in Table 4. One can see that in all regressions, the con coefficient is negative, meaning an increase in the number of contacts decreases the expected amount of overlap for a user with median activity level. Conversely, an increase in activity generally increases the expected degree of overlap, as act is positive.

We find that the stability of users' networks can depend significantly on the tradeoff between network size and activity level. For example, wall posts exhibit a relatively large and positive interaction term, meaning that high levels of activity mitigate the negative effect of additional contacts. In other cases, such as comments or profile views, highly active users are even less likely to retain top contacts as they interact with more contacts. The models explain between 7% and 62% of the variance, which suggests that activity and net-

Modality	$intercept$	act	con	$act \times con$	R^2
Profile	0.39	0.83	-0.56	-0.11	0.31
Photo	0.24	0.38	-0.31	-0.02	0.07
Comment	0.38	0.65	-0.40	-0.26	0.30
Message	0.20	0.53	-0.24	-0.09	0.26
Wall post	0.15	1.07	-0.64	0.20	0.62

Table 4: Regressions explaining the variation in the fraction of top 10 contacts that persist over time. Independent variables are given as percentiles, centered at zero, so that the intercept captures the expected level of overlap between the two time periods for users with a median level of activity (act) and median number of contacts (con). N = 103, 058 All coefficients are significant at the $p < 10^{-16}$ level and have standard errors that are at least an order of magnitude smaller than the coefficient itself.

work size are useful but not sufficient for predicting shifts in attention over time.

6 Conclusion

We have provided a way of analyzing individuals' personal networks in terms of the way they balance their attention across social contacts. This measure exposes properties that are distinct from traditional analyses of personal networks based on size and composition, and it enables a comparison of different interaction modalities and different patterns within and between groups. In addition, the measure has important practical implications: by modeling an individual's balance of social attention, product designers can properly tailor that individual's experience to match her preferences for keeping in touch mostly with her top contacts, or with a more diverse set of people.

While our analysis here is based on Facebook data, the framework is very general, and can be applied to any context where detailed interaction data is available, including other social media sites as well as communication modalities such as phone and e-mail. It is an interesting open question to see how the balance of social attention varies across different domains, and in principle these measures can provide a way of categorizing such domains as more focused or more dispersed. It also becomes promising to consider using the balance of attention as a potential feature of individuals in user-based classification and learning tasks, since we have seen that it captures sources of variation among individuals in ways that other measures may miss. Finally, just as measures of network topology can be used to classify different networks into particular archetypes (Newman and Park 2003; Kwak et al. 2010), this measure might prove useful for distinguishing between different types of social environments.

References

Backstrom, L.; Sun, E.; and Marlow, C. 2010. Find me if you can: Improving geographical prediction with social and spatial proximity. In *Proc. 19th Intl. World Wide Web Conf.*

Black, D.; Gates, G.; Sanders, S.; and Taylor, L. 2000. Demographics of the gay and lesbian population in the United States: Evidence from available systematic data sources. *Demography* 37(2):139–154.

Burt, R. S. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press.

Campbell, K. E., and Lee, B. A. 1991. Name generators in surveys of personal networks. *Social Networks* 13(3).

Chang, J.; Rosenn, I.; Backstrom, L.; and Marlow, C. 2010. epluribus: Ethnicity on social networks. In *Proc. 4th International Conference on Weblogs and Social Media*.

Fischer, C. S. 1982. *To Dwell Among Friends*. University of Chicago Press.

Gilbert, E.; Karahalios, K.; and Sandvig, C. 2008. The network in the garden: An empirical analysis of social media in rural life. *Proc. 26th ACM SIGCHI Conference*, 1603–1612.

Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology* 78:1360–1380.

Hartup, W. W., and Stevens, N. 1997. Friendships and adaptation in the life course. *Psych. Bulletin* 121(3):355–370.

Jiang, J.; Wilson, C.; Wang, X.; Yafei Dai, P. H.; and Zhao, B. Y. 2010. Understanding latent interactions in online social networks. *Proc. 10th Internet Measurement Conference*.

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

Kleinberg, J. 2006. Complex networks and decentralized search algorithms. *Proc. Intl. Congress of Mathematicians*.

Kossinets, G.; Kleinberg, J.; and Watts, D. 2008. The structure of information pathways in a social communication network. In *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *Proc. 19th International World Wide Web Conference*.

Marsden, P. V., and Campbell, K. E. 1984. Measuring tie stength. *Social Forces* 63(2):482–501.

Marsden, P. V. 1987. Core discussion networks of americans. *American Sociological Review* 52(1):122–131.

Mayhew, B. H., and Levinger, R. L. 1976. Size and the density of interaction in human aggregates. *American Journal of Sociology* 82(1):86–110.

McPherson, M., and Smith-Lovin, L. 1993. You are who you know: A network approach to gender. In England, P., ed., *Handbook of Statistical Genetics*. Aldine. 223–251.

McPherson, M.; Smith-Lovin, L.; and Brashears, M. E. 2006. Social isolation in America: Changes in core discussion networks over two decades. *American Soc. Rev.* 71(3).

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–444.

Milgram, S. 1970. The experience of living in cities. *Science* 167(3924):1461–1468.

Moore, G. 1990. Structural determinants of men's and women's personal networks. *American Soc. Rev.* 55(5).

Newman, M. E. J., Park, J. 2003. Why social networks are different from other types of networks. *Phys. Rev. E* 68(3).

Wellman, B., and Wortley, S. 1990. Different strokes from different folks: Community ties and social support. *American Journal of Sociology* 96(3):558–588.

Wilson, C.; Boe, B.; Sala, A.; Puttaswamy, K. P. N.; and Zhao, B. Y. 2009. User interactions in social networks and their implications. In *Proc. 2009 EuroSys Conference*.