Twitter reciprocal reply networks exhibit assortativity with respect to happiness

Catherine A. Bliss*, Isabel M. Kloumann, Kameron Decker Harris, Christopher M. Danforth, Peter Sheridan Dodds

Department of Mathematics and Statistics, Vermont Complex Systems Center & the Vermont Advanced Computing Core, University of Vermont, Burlington, VT 05405, United States

ABSTRACT

The advent of social media has provided an extraordinary, if imperfect, 'big data' window into the form and evolution of social networks. Based on nearly 40 million message pairs posted to Twitter between September 2008 and February 2009, we construct and examine the revealed social network structure and dynamics over the time scales of days, weeks, and months. At the level of user behavior, we employ our recently developed hedonometric analysis methods to investigate patterns of sentiment expression. We find users' average happiness scores to be positively and significantly correlated with those of users one, two, and three links away. We strengthen our analysis by proposing and using a null model to test the effect of network topology on the assortativity of happiness. We also find evidence that more well connected users write happier status updates, with a transition occurring around Dunbar's number. More generally, our work provides evidence of a social sub-network structure within Twitter and raises several methodological points of interest with regard to social network reconstructions.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Social network analysis has a long history in both theoretical and applied settings [1]. During the last 15 years, and driven by the increased availability of real-time, in-situ data reflecting people's social interactions and choices, there has been an explosion of research activity around social phenomena, and many new techniques for characterizing large-scale social networks have emerged. Numerous studies have examined the structure of online social networks in particular, such as blogs, Facebook, and Twitter [2–19].

In a series of analyses of the Framingham Heart Study data and the National Longitudinal Study of Adolescent Health, Christakis, Fowler, and others have examined how qualities such as happiness, obesity, disease, and habits (e.g., smoking) are correlated within social network neighborhoods [20–25]. The authors' additional assertion of contagion, however, has been criticized primarily on the basis of the difficulties to be found in distinguishing these phenomena from homophily [26–28]. The observation that social networks exhibit assortativity with respect to these traits evidently requires further study and leads us to explore potential mechanisms. Advances would naturally provide further insight into the nature of how social groups influence individual behavior and vice versa.

Our focus in the present work is the social network of Twitter users. With the abundance of available data, Twitter serves as a living laboratory for studying contagion and homophily [29]. As a requisite step toward these goals, we first define sub-networks of Twitter users suitable to such study and, second, examine whether assortativity is observed in these sub-networks. Before describing our methods, we provide a brief overview of Twitter, related work, and the challenges associated with social network analysis in this arena.

Twitter is an online, interactive social media platform in which users post tweets, micro-blogs with a 140 character limit. Since its inception in 2006, Twitter has grown to encompass over 200 million accounts, with over 100 million of these accounts currently active as of October 2011, and with some users having garnered over 10 million followers [30]. Tweets are open online by default, and are also broadcast directly to a user's followers. Users may express interest in a tweet by retweeting the message to their followers. Alternatively, followers may reply directly to the author.

Understanding the topology of the Twitter network, the manner in which users interact and the diffusion of information through this media is challenging, both computationally and theoretically. One of the central issues in characterizing the topology of any network representation of Twitter lies in defining the criteria for establishing a link between two users. The majority of previous studies have examined the topology of and information cascades on the Twitter follower network [7,10,15], as well as on networks derived from mutual following [8]. However, the follower network

* Corresponding author.
  E-mail addresses: Catherine.Bliss@uvm.edu (C.A. Bliss),
Isabel.Kloumann@uvm.edu (I.M. Kloumann), Kameron.Harris@uvm.edu
(K.D. Harris), Chris.Danforth@uvm.edu (C.M. Danforth), Peter.Dodds@uvm.edu
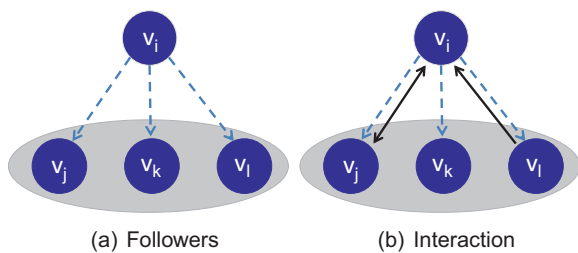(P.S. Dodds).

Fig. 1. (a) Follower network: The follower network is generated by declared following choices, absent any messages being sent. If user v_i broadcasts tweets to followers v_j, v_k and v_ℓ (represented by the dashed, blue arrow) v_i would be connected to each of v_j, v_k and v_ℓ by a directed link in a follower network. (b) Reciprocal-reply network: Directed replies are represented by a solid black arrow. When considering the interaction between users, a reply (i.e., v_ℓ replies to v_i) provides evidence of a directional interaction between nodes. We mandate a stronger condition for interaction, namely reciprocal replies (i.e., v_j replies to v_i and vice versa) over a given time period. Thus v_i and v_j are connected in the reciprocal reply network that we construct. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

is not the only representation of Twitter's social network, and its structure can be misleading [31]. For example, in a study of over 6 million users, Cha et al. [10] found that users with the highest follower counts were not the users whose messages were most frequently retweeted. This suggests that such popular users (as measured by follower count) may not be the most influential in terms of spreading information, and this calls into question the extent to which users are influenced by those that they follow [32]. Of further concern is the finding of low reciprocity within follower networks. Kwak et al. found very few individuals who followed their followers [15]. As a result, trying to infer meaningful influence and contagion in such a network is difficult.

While popular users and their many followers clearly exhibit an affiliation, they do not necessarily interact, as there are different relationships implicated by broadcasting (tweeting), sending a message (@someone), and replying to a message. As an example, we consider a user represented by node v_i which has three followers, represented by v_j, v_k, and v_ℓ as shown in Fig. 1a. When a user broadcasts tweets to their many followers, as represented by the directed arrow in Fig. 1a, this does not imply that followers read or respond to these tweets. Followers v_j, v_k, and v_ℓ receive all tweets broadcast by node v_i, but this provides no guarantee of interaction. Suppose, though, that we observe that v_ℓ replies to v_i as shown in Fig. 1b. This provides evidence (but not proof) that the user represented by v_ℓ has indeed received a tweet from v_i and is sufficiently motivated to create a response to v_i. Although a directional network based on these replies can be created, such a directional interaction, however, does not suggest reciprocity between the nodes. In this example, we have no evidence that v_i has, in any way, considered or even read such a response from his/her follower.

We conclude that following and unreciprocated replies are not sufficient for interaction and present an alternative means by which to derive a social network from Twitter messages, via reciprocal replies. In our reciprocal-reply network, two nodes, v_i and v_j, are connected if v_i has replied to v_j and v_j has replied to v_i at least once within a given time period of consideration. In Fig. 1b, the nodes v_i and v_j meet this criterion. Another challenge in characterizing the topology of any network representation of Twitter concerns determining how long a link between two users in the network should persist. Including stale user–user interactions in the network mistakenly creates an inaccurate portrayal of the current state of the system; this is typically referred to as the "unfriending problem" [26]. Not only will network statistics such as the number of nodes, average degree, maximum degree and proportion of nodes in the giant component be artificially inflated due to superfluous,

no-longer-active links [26,33], but the degree distribution will also be distorted. Kwak et al. [15] found that the degree distribution for a Twitter follower network deviated from a power law distribution due to an overabundance of high degree nodes resulting from an accumulation of "dead-weight" in the network.

Additional problems are encountered if one uses accumulated network data to measure assortativity with respect to a trait (e.g., happiness). As an example, consider a network in which two users are connected because they interacted during the last week of a year-long study. Including this user–user pair in the list of pairs to compute assortativity for the entire network blurs the relationship between more consistent and repeated interactions that occurred throughout the timespan of the study. Further complications arise when averaging a user's trait over a large time scale (i.e., averaging happiness over a 6 month or 12 month timespan). Detecting changes in users' traits over time and how these may (or may not) be correlated with nearest neighbors' traits is of fundamental importance; accumulated network data occludes exactly the interactions we are looking to understand. Recognizing that, due to practical limitations, accumulation of network data must occur on some scale, we analyze users in day, week, and month reciprocal reply networks. By examining networks constructed at smaller time scales and calculating users' happiness scores based on tweets made only during that time period, we aim to take a more dynamic view of the network.

In addition to defining reciprocal reply networks and advocating for their use, we also seek to describe how happiness is distributed in the reciprocal reply networks of Twitter. Previous hedonometric work with Twitter data has revealed cyclical fluctuations in average happiness at the level of days and weeks, as well as spikes and troughs over a time scale of years corresponding to events such as U.S. Presidential Elections, the Japanese tsunami and major holidays [11,34,35]. Other studies have examined changes in valence of tweets associated with the death of Michael Jackson [14], changes in the U.S. Stock Market [9], the Chilean Earthquake of 2010, and the Oscars [16]. In the present work, we seek to understand localized patterns of happiness in the Twitter users' social network.

Understanding how emotions are distributed through social networks, as well as how they may spread, provides insight into the role of the social environment on individual emotional states of being, a fundamental characteristic of any sociotechnical system. Bollen et al. [8] examine a reciprocal-follower network using Twitter and suggest that Subjective Well-Being (SWB), a proxy for happiness, is assortative. Building on their work, we address whether happiness is assortative in reciprocal-reply networks. We also test the hypothesis of Christakis and Fowler [25] who find evidence that the assortativity of happiness may be detected up to three links away. In doing so, we raise an additional point which is not specific to Twitter networks, but rather relates to empirical measures of assortativity in general. Relatively few studies have employed a null model for calculating the pairwise correlations (e.g., happiness–happiness). We devise a null model which maintains the topology of the network and randomly permutes happiness scores attached to each node. By randomly permuting users' happiness scores, we can detect what effect, if any, network structure has on the pairwise correlation coefficient.

We organize our paper as follows: In Section 2, we describe our data set, the algorithm for constructing reciprocal-reply networks, network statistics used for characterizing the networks, and our measure for happiness. We propose an alternative means by which to detect social structure and argue that our method detects a large social sub-network on Twitter. In Section 3, we describe the structure of this network, the extent to which it is assortative with respect to happiness and the results of testing assortativity against a null model. In Section 4, we discuss these findings and propose further investigations of interest.
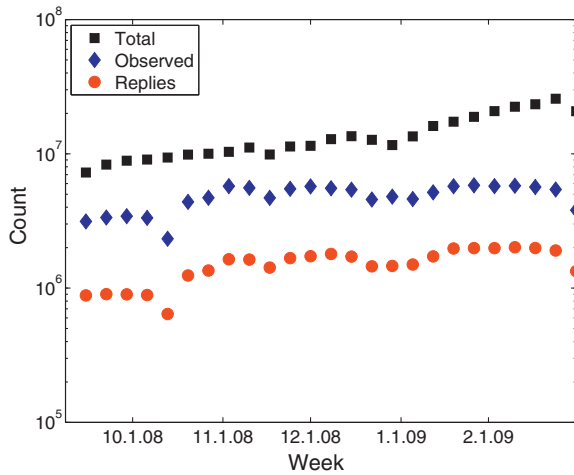
Fig. 2. Tweet counts are plotted for the weeks between September 2008 and February 2009. The three curves represent the total, those that we observed and the number of the observed tweets that constituted replies.
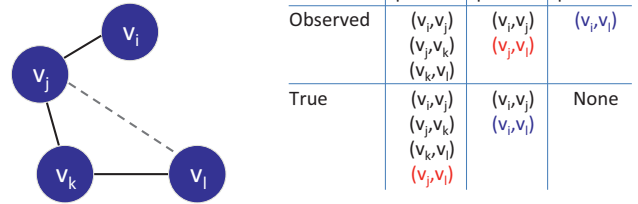

Fig. 3. The effect of missing links in the reciprocal reply network is depicted where observed links are shown as a solid line and an unobserved link is shown as a dashed line. The effect of unobserved links is twofold: (1) some connections between nodes are missed (e.g., v_j and v_ℓ are not connected in the observed reciprocal reply network) and (2) some path lengths between nodes are artificially inflated (e.g., the distance from v_i to v_ℓ is 3 in the observed reciprocal-reply network, however in reality the path length is 2).

2. Methods

2.1. Data

From September 2008 to February 2009, we retrieved over 100 million tweets from the Twitter streaming API service.[1] While the volume of our feed from the Twitter API increased during this study period, the total number of tweets grew at a faster rate (Fig. 2). During this time period, we estimate that we collected roughly 38% of all tweets.[2] The number of messages and percent of which were replies are reported in Table A4. For the remainder of this paper, we restrict our attention to the nearly 40 million message-reply pairs within this data set and the users who authored these tweets.

The data received from the Twitter API service for each tweet contained separate fields for the identification number of the message (message id), the identification number of the user who authored the tweet (user id), the 140 character tweet, and several other geo-spatial and user-specific metadata. If the tweet was made using Twitter's built-in reply function,[3] the identification number of the message being replied to (original message id) and the identification of the user being replied to (original user id) were also reported.

We acknowledge two sources of missing data. First, the Twitter API did not allow us access to all tweets posted during the 6 month period under consideration. Thus, there are replies that we have not observed. As a result, some users may remain unconnected or connected by a path of longer length due to missing intermediary links in our reciprocal-reply network (Fig. 3). Secondly, we acknowledge that users may be interacting with each other and not using the built-in reply function. We discuss this further in the next section.

2.2. Reciprocal-reply network

In keeping with terminology used in the field of complex networks, the terms *nodes* and *links* will be used henceforth to describe users and their connections. Define $\mathcal{G} = (V, E)$ to be a simple graph which contains, $N = |V|$ nodes and $M = |E|$ links. We construct the reciprocal-reply networks in which users are represented by nodes, $v_i \in V$, and links connecting two nodes, $e_{ij} \in E$, indicate that v_i and v_j have made replies to each other during the period of time under analysis (Fig. 1). For each network, we remove self-loops (i.e., users who responded to themselves). We characterize the reciprocal-reply network for each week by the calculation of network statistics such as N (the number of nodes), $\langle k \rangle$ (average degree), k_{max} (maximum degree), the number of connected components and S (proportion of nodes in the giant component). We calculate clustering, C_G, according to Newman's global clustering coefficient [36]:

$$C_G = \frac{3 \times (\text{number of triangles on a graph})}{\text{number of connected triples of nodes}}.$$

Assortativity refers to the extent to which similar nodes are connected in a network. Often, degree assortativity is quantified by computing the Pearson correlation coefficient of the degrees at each end of links in the network [37]. Since we are interested in quantifying the extent to which the highest degree nodes are connected to other high degree nodes, as defined by the rank of their degrees, we instead measure degree assortativity by the Spearman correlation coefficient.[4] Thus for each link that connects nodes v_i and v_j, we examine the ranks of k_{v_i} and k_{v_j}. The Spearman correlation coefficient, which is the Pearson correlation coefficient applied to the ranks of the degrees at each end of links in the network, is a non-parametric test that does not rely on normally distributed data and is much less sensitive to outliers.[5]

In addition, we also investigate user pairs which are connected by a minimal path length of two (or three) in the reciprocal reply networks. We define $d(v_i, v_j)$ to be the path length (i.e., number of links) between nodes v_i and v_j such that no shorter path exists. As a consequence of missing messages, we recognize that some users will appear to remain unconnected or connected by a path of longer length. Fig. 3 depicts the effect of missing links on inferred path lengths between nodes in the network. Nodes v_j and v_ℓ are adjacent
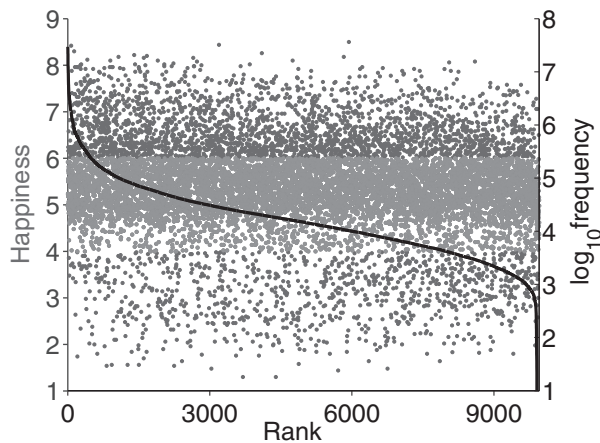

Fig. 4. The happiness scores of words are plotted as a function of their rank (dots), with the stop words (words within $\pm\Delta h = 1$ of $h_{avg} = 5$) depicted in light gray [38]. These words were excluded from the happiness score computation. The frequency of words and their rank (1 = most frequent, 9956 = least frequent) are plotted (solid curve). Not all 10,222 labMT words were observed during the time period from September 2008 to February 2009.

in the network, however, due to the missing link represented by the dashed line, these nodes are inferred to be two links apart.

2.3. Measuring happiness

To quantify happiness for Twitter users, we apply the real-time hedonometer methodology for measuring sentiment in large-scale text developed in Dodds et al. [11]. In this study, the 5000 most frequently used words from Twitter, Google Books (English), music lyrics (1960–2007) and the New York Times (1987–2007) were compiled and merged into one list of 10,222 unique words.[6] This word list was chosen solely on the basis of frequency of usage and is independent of any other presupposed significance of individual words. Human subjects scored these 10,222 words on an integer scale from 1 to 9 (1 representing sad and 9 representing happy) using Mechanical Turk. We compute the average happiness score (h_{avg}) to be the average score from 50 independent evaluations. Examples of such words and their happiness scores are: $h_{avg}(love) = 8.42$, $h_{avg}(special) = 7.20$, $h_{avg}(house) = 6.34$, $h_{avg}(work) = 5.24$, $h_{avg}(sigh) = 4.16$, $h_{avg}(never) = 3.34$, $h_{avg}(sad) = 2.38$, $h_{avg}(die) = 1.74$. Words that lie within $\pm\Delta h_{avg} = 1$ of $h_{avg} = 5$ were defined as "stop words" and excluded to sharpen the hedonometer's resolution.[7] The result is a list of 3,686 words, hereafter referred to as the Language Assessment by Mechanical Turk (labMT) word list [11]. See Tables A1 and A2 for additional example word happiness scores.

Fig. 4 presents word happiness as a function of usage rank for the roughly 10,000 words in the labMT data set. This figure reveals a frequency independent bias toward the usage of positive words (see [37] for further discussion of this positivity bias). Proceeding with the labMT word list, a pattern-matching script evaluated each tweet for the frequency of words. We compute the happiness of each user by applying the hedonometer to the collection of words from all tweets authored by the user during the given time period. Note that each users' collection of words likely reflects messages that were not replies. The happiness of this collection of words is taken

Table 1
Happiness scores are computed as a weighted average of words' h_{avg} scores. Since "starts" is a stop word, it is not included in the calculation of $h_{avg}(T) = 7.07$. This example serves is included as a means to illustrate the methodology; in practice, the average is calculated over a much larger word set.

w_i	$h_{avg}(w_i)$	labMT?	f_i	p_i
Vacation	7.92	Yes	1	1/2
starts	5.96	Yes	n/a	n/a
today	6.22	Yes	1	1/2
yeahhhhh	n/a	No	n/a	n/a

to be the frequency weighted average of happiness scores for each labMT word as $h_{avg}(T) = \sum_{i=1}^{N} h_{avg}(w_i)f_i / \sum_{i=1}^{N} f_i = \sum_{i=1}^{N} h_{avg}(w_i)p_i$, where $h_{avg}(w_i)$ is the average happiness of the ith word appearing with frequency f_i and where p_i is the normalized frequency $\left(p_i = f_i / \sum_{j=1}^{N} f_j\right)$. As a simple example, we consider the phrase: *Vacation starts today, yeahhhhh!* in Table 1. In practice, though, the hedonometer is applied to a much larger word set and is not applied to single sentences.

Having found happiness scores for each node (user), we then form happiness–happiness pairs (h_{v_i}, h_{v_j}), where h_{v_i} and h_{v_j} denote the happiness of nodes v_i and v_j connected by an edge. The Spearman correlation coefficient of these happiness–happiness pairs measures how similar individuals' average happiness is to that of their nearest neighbors'. Lastly, we investigate the strength of the correlation between users' average happiness scores and those of other users in the two and three link neighborhoods.

3. Results

3.1. Reciprocal-reply network statistics

Visualizations of day and week networks were created using the software package Gephi [39]. Figs. 5 and A6 show a sample week and day network, respectively. All layouts were produced using the Force Atlas 2 algorithm, which is a spring based algorithm that plots nodes together if they are highly connected (see [40] for more details). The sizes of the nodes are proportional to the degrees.

Network statistics, such as the number of nodes (N), the average degree $\langle k \rangle$, the maximum degree (k_{max}), global clustering C_G, degree assortativity (Assort), and the proportion of nodes in the giant component (S) are summarized in Fig. 6. Several trends are apparent.

Throughout the course of the study, the number of users in the observed reciprocal-reply network shows an increase, whereas the average degree, degree assortativity, and proportion of nodes in the giant component remain fairly constant. The fluctuations in maximum degree are the result of celebrities or companies having bursts of high volume reply exchanges with their fans during a particular week, for example Bob Bryar, Drummer for the band *My Chemical Romance* ($k_{max} = 1244$, Week 12), *Namecheap* domain registration company ($k_{max} = 1245$, Week 13), Twitter's own *Shorty Awards* ($k_{max} = 1456$, Week 14), and Stephen Fry, actor and megablogger ($k_{max} = 1718$, Week 22). This observation highlights the importance of examining network data on the appropriate time scale, otherwise information about these kinds of dynamics would be lost. The clustering coefficient shows a slight decrease over the course of this period. This is most likely due to an increasing number of nodes, which results in a smaller proportion of closed triangles in the network. The degree distribution, P_k, for a sample week (week beginning January 27, 2009) is presented in Fig. 7. Using the approach outlined by Clauset, Shalizi, and Newman [41], we find a lower bound for the scaling region to be $k_{min} \approx 34$ and a very steep scaling exponent of $\alpha = 3.5$. This suggests a constrained variance and mean. We test whether the empirical distribution is distinguishable

---

[6] We provide a brief summary of this methodology here and refer the interested reader to the original paper for a full discussion. The supplementary information contains the full word list, along with happiness averages and standard deviations for these words [11].

[7] For notational convenience, we henceforth use Δh in lieu of Δh_{avg}.
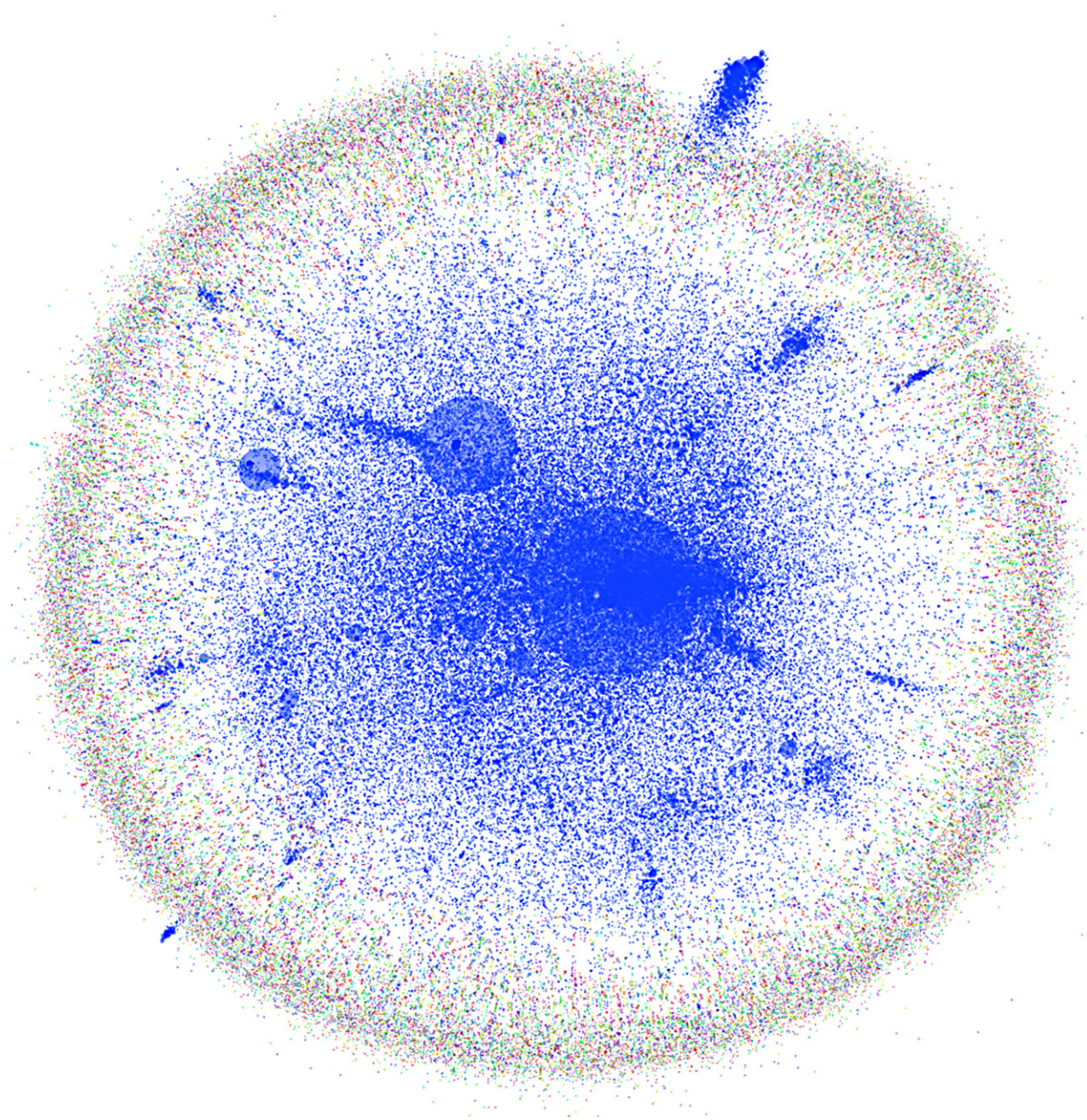
Fig. 5. A visualization of the 162,445 nodes in the reciprocal reply network for the week beginning December 9, 2008 (Week 14) is depicted. Node colors represent connected components, a total of 15,342, with the giant component (shown in blue) comprising 76 % of all nodes. The size of each node is proportional to its degree. The visualization was made using Gephi [39]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

from a power law using the Kolmogorov–Smirnov test and find no evidence against the null hypothesis for the week ($D = 2.28 \times 10^{-2}$, $p = 0.095$, $n = 203,852$). We find the same exponent and statistically stronger evidence of a power law for a sample month (see Fig. A1). This suggests that these distributions' tails may be fit by a power law.

3.2. Measuring happiness

The application of the hedonometer gives reasonable results when applied to a large body of text, but can be misleading when applied to smaller units of language [11]. To provide a sense of how sensitive this measure is to the number of labMT words posted

by users, we sampled happiness–happiness pairs, (h_{v_i}, h_{v_j}) whose respective users, v_i and v_j, had posted at least α total labMT words during a sample week (week beginning January 27, 2009). For these users, we compute happiness assortativity and show the variation with α in Fig. 8. For $\Delta h = 0$, there is less variation due to the numerous words centered around the mean happiness score regardless of the threshold, α. Tuning both parameters too high results in few sampled words and corrupts the interpretation of the results.

Figs. 9 and 10 reveal a weakening happiness–happiness correlation for users in the week networks as the path length between nodes increases. All correlations, for each week, were significant ($p < 10^{-10}$). This suggests that the network is assortative with
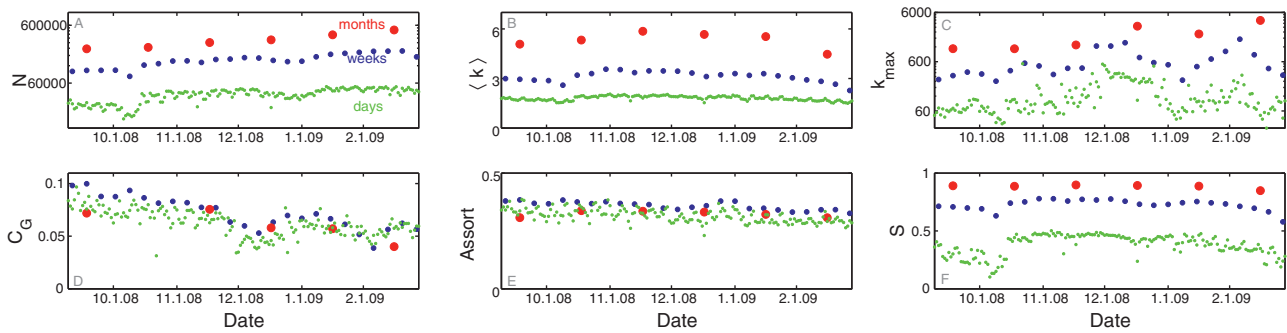
Fig. 6. Network statistics for the reciprocal-reply network are constructed at the scale of days (green), weeks (blue), and months (red). (A) The number of users (N) engaged in reciprocal exchanges when viewed at the level of days, weeks, or months increases over the study period. (B) The average degree ($\langle k \rangle$) remains fairly constant throughout the study period, with higher values detected for larger interaction time periods. (C) The maximum degree (k_{max}) shows variability throughout the study period. (D) Clustering decreases quite likely resulting from the inability of the networks' closed triangles to keep up with the growing number of nodes. (E) Degree assortativity remains fairly constant throughout the study period, and shows little sensitivity to the time period over which the networks represent interactions. (E) The proportion of nodes in the giant component (S) remains fairly constant for week and month networks, however, shows some variability during the first month of the study for day networks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

respect to happiness and that user happiness is more similar to their nearest neighbors than those who are 2 or 3 links away.

In Fig. 11 we provide a visualization of an ego-network for a single node, including neighbors up to three links away. Nodes are colored by their h_{avg} score, illustrating the assortativity of happiness. Fig. A5 visualizes the happiness assortativity for an entire week network.

In Fig. 12, we show the average happiness score as a function of user degree k for all week networks. The average happiness score increases gradually as a function of degree, with large degree nodes demonstrating a larger average happiness than small degree nodes. Large degree nodes use words such as "you," "thanks," and "lol" more frequently than small degree nodes, while the latter group uses words such as "damn," "hate," and "tired" more frequently. A word shift diagram, comparing nodes with $k < 100$ vs. nodes with $k \geq 100$ is included in Fig. A7. Fig. 12 also reveals that the number of large degree nodes is fairly small. Our results support recent work

showing that most users of Twitter exhibit an upper limit on the number of active interactions in which they can be engaged [31]. This may provide further evidence in support of Dunbar's hypothesis, which suggests that the number of meaningful interactions one can have is near 150 [42].

3.3. Testing assortativity against a null model

To further examine these findings, we create a null model which maintains the network topology (i.e., adjacency matrices for one link, two link, and three link remain intact), but randomly permutes the happiness scores associated with each node. The Spearman correlation coefficient shows no statistically significant relationship for the null model applied to a sample week of the data set. Fig. 13 shows the results of 100 random permutations applied to nodes' associated happiness scores. The Spearman correlation coefficients for the observed data are shown as blue
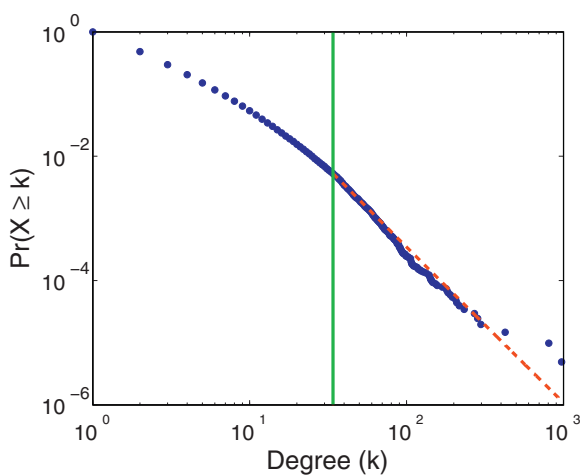

Fig. 7. Log–log plot of the complementary cumulative distribution function (CCDF) of the degree distribution for a sample week (week of January 27, 2009) network is shown (blue), along with the best fitting power law model ($\alpha = 3.50$ and $k_{min} = 34$) using the procedure of Clauset et al. [41]. We test whether the empirical distribution is distinguishable from a power law using the Kolmogorov–Smirnov test and find no evidence against the null hypothesis ($D = 2.28 \times 10^{-2}$, $p = 0.095$, $n = 203,852$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)
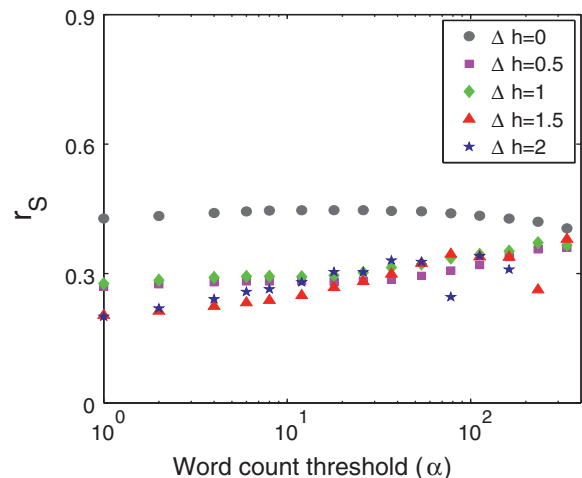

Fig. 8. Nearest neighbor happiness assortativity as a function of the number of labMT words required per user is displayed for a sample week reciprocal-reply network. Notice that when $\Delta h = 0$, there is less variation due to the numerous words centered around the mean happiness score regardless of the threshold, α. While this stability is desirable, tuning Δh allows us to sharpen the resolution of the hedonometer. This tuning, however, must be balanced with the appropriate choice of α.
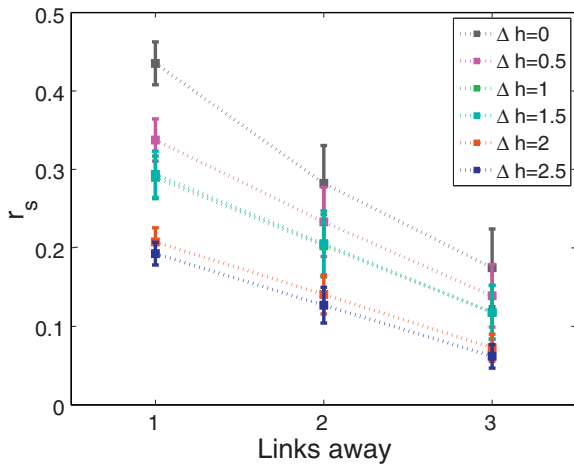
Fig. 9. Average assortativity of happiness for week networks measured by Spearman's correlation coefficients as Δh is dialed from 0 to 2.5, with $\alpha = 50$. As Δh increases, the average correlation decreases. For large Δh the resulting words under analysis have more disparate happiness scores and thus the correlations between users' happiness scores are smaller. Similarly, choosing Δh to be too small (e.g., $\Delta h = 0$) could result in an over estimate of happiness–happiness correlations because of the uni-modal distribution of h_{avg} for the labMT words. Thus a moderate value for Δh is chosen (Δh is set to 1 for this study).

squares ($\Delta h_{avg} = 0$) and green diamonds ($\Delta h_{avg} = 1$). The average and standard deviation of the Spearman correlation coefficient calculated for the 100 randomized happiness scores (null model) are shown as red circles with error bars (the error bars are smaller than the symbol). This data supports the hypothesis that happiness is less assortative as network distance increases.

Lastly, we explore whether these correlations are due to similarity of word usage. For this analysis, we compute the similarity of word bags for users connected in the reciprocal reply networks. We compare the distribution of observed similarity scores to similarity

scores obtained by randomly reassigning word bags to users. Fig. A8 shows that both distributions are of a similar form, with the randomized version exhibiting a slightly lower mean similarity score ($\overline{D_{i,j}} = .167$) as compared to the mean of the observed similarity scores for users ($\overline{D_{i,j}} = .267$). If users were tweeting similar words with a similar frequency, we would expect a much larger mean similarity score for the observed data. Thus, we do not find evidence suggesting that the happiness correlations are due to similarity of word bags.

4. Discussion

In this paper, we describe how a social sub-network of Twitter can be derived from reciprocal-replies. Countering claims that Twitter is not social a network [15], we provide evidence of a very social Twitter. The large volume of replies (millions every week) and assortativity of user happiness indicates that Twitter is being used as a social service. Furthermore, conducted at the level of weeks, our analysis examines an in the moment social network, rather than the stale accumulation of social ties over a longer period of time. A network in which edges are created and never disintegrate results in dead links with no contemporary functional activity. This problem of unfriending has been noted [26] and can greatly impact conclusions drawn when observational data are used to infer contagion.

Our characterization of the reciprocal reply network reveals several trends over the 25 week period from September 2008 to February 2009. The number of nodes, N, in a given week network increased as time progressed, which is undoubtedly due to Twitter's enormous growth in popularity over the study period. Similarly, with an increasing number of nodes, we observe a smaller proportion of closed triangles (i.e., clustering shows a slight decrease). This may be due in part to sub-sampling effects or due to an increasing N, with which the number of closed triangles (i.e., friends of friends) cannot keep up. The proportion of nodes in the giant component remains fairly constant, as does degree
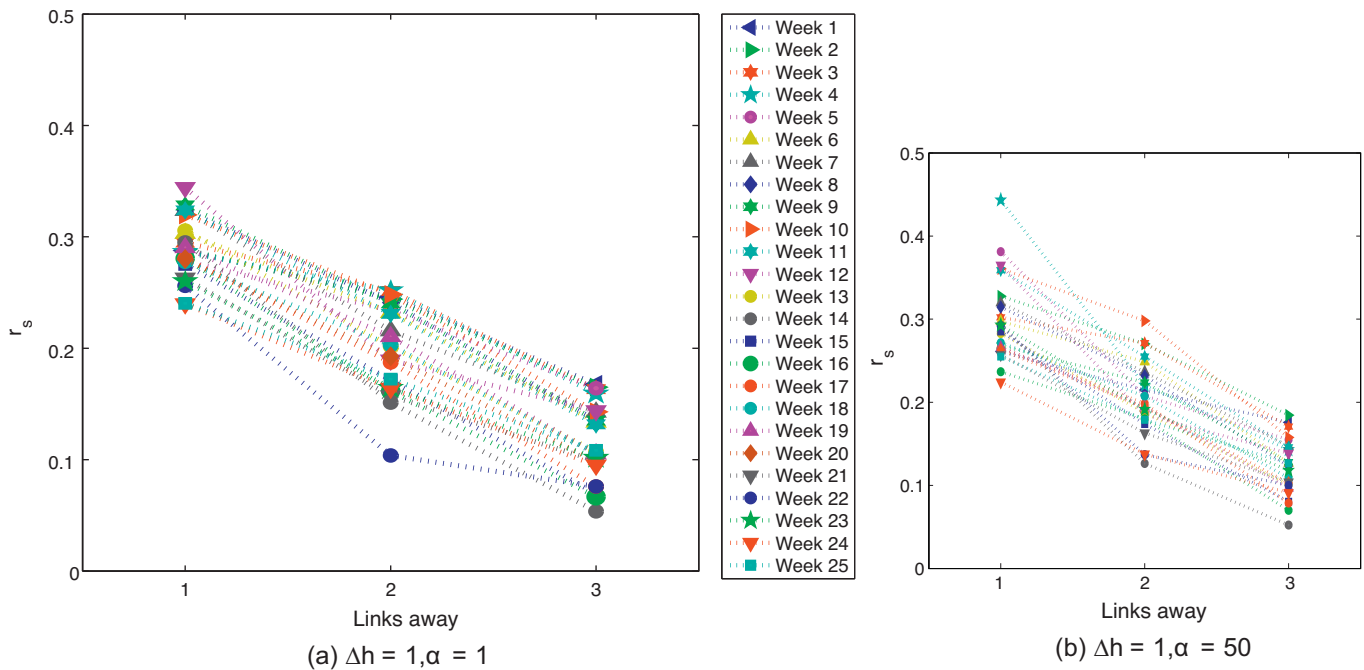

Fig. 10. Happiness assortativity as measured by Spearman's correlation coefficients is shown for week networks, with $\Delta h = 1$ and (a) the threshold of labMT words written by users set to $\alpha = 1$ and (b) $\alpha = 50$. The dashed lines indicate weakening happiness–happiness correlations as the path length increases from one, two, and three links away, for each week in the data set.
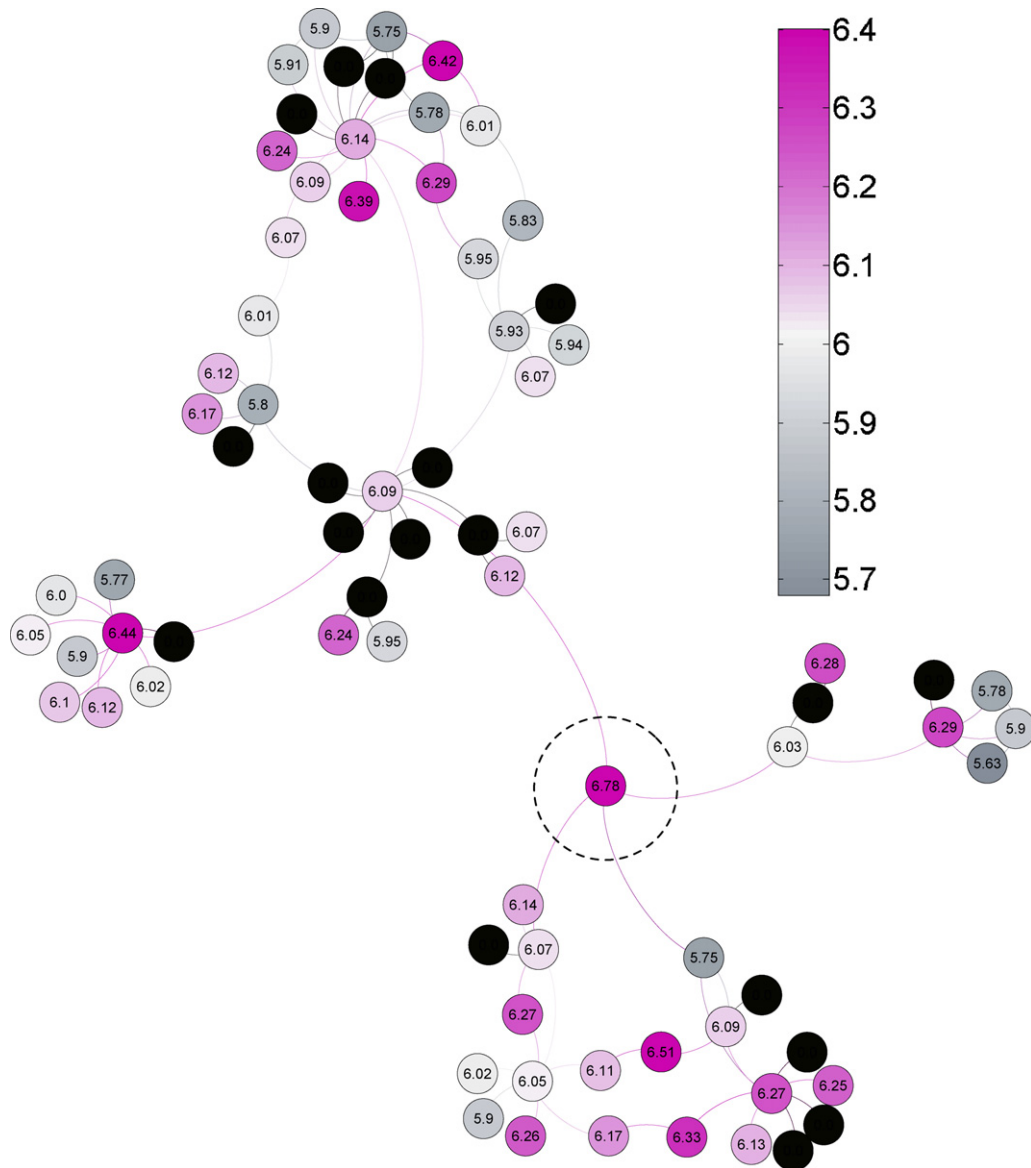
Fig. 11. A visualization of a user and its neighbors 3-links away for a week beginning September 9, 2008 (Week 1). Colors represent happiness scores for users posting more than $\alpha = 50$ labMT words. Nodes depicted with the color black are nodes for which the user's wordbag did not meet our thresholding criteria.

assortativity as measured by Spearman's correlation coefficient. Had we used the Pearson correlation coefficient, degree assortativity would have been highly variable (Fig. A1) due to the extreme values of maximum degree ($k_{\max}$) during weeks 12–14 and 22. Using the Spearman rank correlation coefficient, which is less sensitive to extreme values, we find that the degree assortativity is fairly constant.

Our work is based on a sub-sample of tweets and is thus subject to the effects of missing data. The problem of missing data has been addressed by several researchers investigating the impact of missing nodes [43–47], missing links, or both [48]. More specifically, the work of Stumpf [43] shows that sub-sampled scale-free networks are not necessarily themselves scale-free. Further work which addresses the problem of missing messages and identifies the consequences of missing data on inferred network topology is needed to more fully address these questions.

We find support for the "happiness is assortative" hypothesis and evidence that these correlations can be detected up to three links away. Further, this finding does not appear to be based on

users tweeting similar words (Fig. A8). Our correlation coefficients for reciprocal-reply networks constructed at the level of weeks are smaller than those obtained by Bollen et al. [8] for a reciprocal-follower network constructed by aggregating over a six month period. This difference is likely a reflection of differences in methodologies, such as our more dynamic time scale (one-week periods vs. six month periods), our exclusion of central value happiness scores (i.e., stop words), and our use of the Spearman correlation coefficient.

While this paper does not attempt to separate homophily and contagion, future work could use reciprocal-reply networks to investigate these effects. While reciprocal-reply networks are subject to errors caused by missing data (see above discussion of this issue) they may provide a valuable framework for studying contagion effects, given that they are based on a conservative and dynamic metric of what constitutes an interaction on Twitter. A network structure in which links are known to be active and valid provides an arena in which the diffusion of information and emotion may be properly studied.
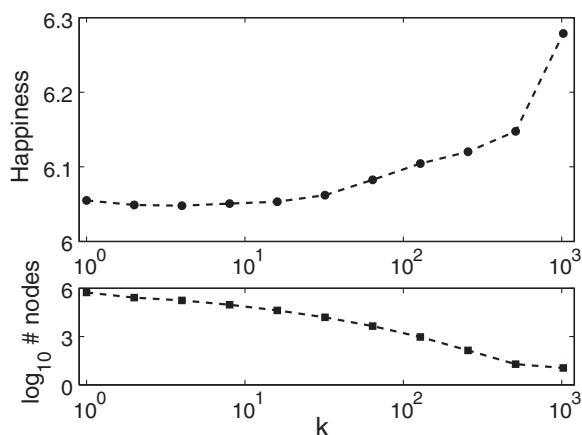
Fig. 12. Top Panel: The average happiness score as a function of user degree k for week networks is increasing, as larger degree nodes use fewer negative words (see Fig. A7). Bottom Panel: The number of unique users is reported with respect to degree k; some users appear in more than one bin because they exhibit different degree k for different weeks of the study.
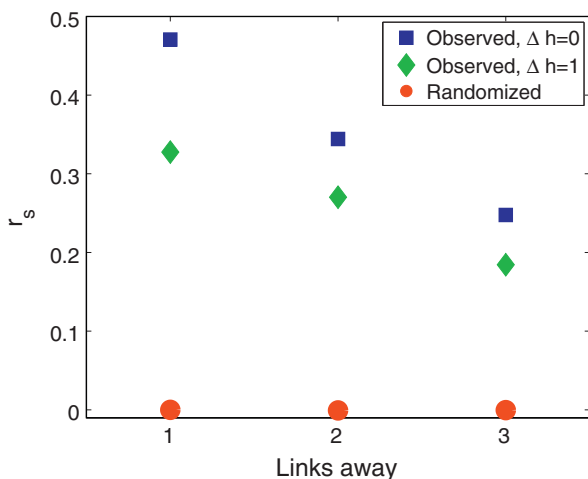

Fig. 13. One hundred random permutations were applied to the happiness scores associated with each node in a sample week network (week beginning October 8, 2008 is shown), with $\Delta h = 0$ (blue square) and $\Delta h = 0$ (green diamonds). The threshold for all cases is set to $\alpha = 50$. The Spearman correlation coefficients, r_s for the observed data are shown as blue squares. The average and standard deviation of the Spearman correlation coefficient calculated for the 100 randomized data (null model) are shown as red circles with error bars (the error bars are smaller than the symbol). The plot shows Spearman correlation coefficients for the null model to be nearly 0 and provides supporting evidence for our observed trend, namely the network is assortative with respect to happiness and the strength of assortativity decreases as path length increases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Acknowledgements

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jocs.2012.05.001.

References

[1] K.F. Stanley Wasserman, Social network analysis: methods and applications, in: Structural Analysis in the Social Sciences, vol. 8, Cambridge University Press, Cambridge, 1994.

[2] M. Gjoka, M. Kurant, C. Butts, A. Markopoulou, Walking in Facebook: a case study of unbiased sampling of OSNs, in: INFOCOM, 2010 Proceedings IEEE, 2010, pp. 1–9.

[3] B. Viswanath, A. Mislove, M. Cha, K.P. Gummadi, On the evolution of user interaction in Facebook, in: Proceedings of the 2nd ACM Workshop on Online Social Networks, WOSN '09, ACM, New York, NY, USA, 2009, pp. 37–42.

[4] Z. Papacharissi, The virtual geographies of social networks: a comparative analysis of Facebook, Linkedin and Asmallworld, New Media & Society 11 (February/March) (2009) 199–220.

[5] P. Dodds, C.M. Danforth, Measuring the happiness of large-scale written expression: songs, blogs, and presidents, Journal of Happiness Studies 11 (2010) 441–456, http://dx.doi.org/10.1007/s10902-009-9150-9.

[6] A. Java, X. Song, T. Finin, B. Tseng, Why we twitter: an analysis of a microblogging community, in: H. Zhang, M. Spiliopoulou, B. Mobasher, C. Giles, A. McCallum, O. Nasraoui, J. Srivastava, J. Yen (Eds.), Advances in Web Mining and Web Usage Analysis, Lecture Notes in Computer Science, vol. 5439, Springer, Berlin/Heidelberg, 2009, pp. 118–138.

[7] E. Bakshy, J.M. Hofman, W.A. Mason, D.J. Watts, Everone's an influencer: quantifying influence on Twitter, in: WSDM '11: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, ACM, New York, NY, USA, 2011, p. 618113.

[8] J. Bollen, B. Goncalves, G. Ruan, H. Mao, Happiness is assortative in online social networks, Artificial Life (2011) 17.

[9] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, Journal of Computational Science 2 (2011) 1–8.

[10] M. Cha, H. Haddadi, F. Benevenuto, K.P. Gummadi, Measuring user influence in twitter: the million follower fallacy, In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM), Washington DC (2010).

[11] P.S. Dodds, K.D. Harris, I.M. Kloumann, C.A. Bliss, C.M. Danforth, Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter, PLoS ONE 6 (2011) e26752.

[12] L. Guo, E. Tan, S. Chen, X. Zhang, Y.E. Zhao, Analyzing patterns of user content generation in online social networks, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09, ACM, New York, NY, USA, 2009, pp. 369–378.

[13] B.A. Huberman, D.M. Romero, F. Wu, Social networks that matter: Twitter under the microscope, CoRR abs/0812.1045, 2008.

[14] E. Kim, S. Gilbert, M.J. Edwards, E. Grae, Detecting sadness in 140 characters: sentiment analysis of mourning Michael Jackson on Twitter, Technical report, Web Ecology Project, 2009.

[15] H. Kwak, C. Lee, H. Park, S. Moon, What is Twitter, a social network or a news media? in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 591–600.

[16] M. Thelwall, K. Buckley, G. Paltoglou, Sentiment in Twitter events Journal of the American Society for Information Science and Technology 62 (2011) 406–418.

[17] J. Weng, E.-P. Lim, J. Jiang, Q. He, Twitterrank: finding topic-sensitive influential twitterers, in: Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10, ACM, New York, NY, USA, 2010, pp. 261–270.

[18] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, P. Li, User-Level Sentiment Analysis Incorporating Social Networks, ArXiv e-prints, 2011.

[19] J. Ugander, L. Backstrom, C. Marlow, J. Kleinberg, Structural diversity in social contagion, Proceedings of the National Academy of Sciences of the United States of America 109 (2012) 5962–5966.

[20] N.A. Christakis, J.H. Fowler, The spread of obesity in a large social network over 32 years, New England Journal of Medicine 357 (2007) 370–379.

[21] J.H. Fowler, N.A. Christakis, Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study, British Medical Journal 337 (2008).

[22] N.A. Christakis, J.H. Fowler, The collective dynamics of smoking in a large social network, New England Journal of Medicine 358 (2008) 2249–2258.

[23] J.N. Rosenquist, J. Murabito, J.H. Fowler, N.A. Christakis, The spread of alcohol consumption behavior in a large social network, Annals of Internal Medicine 152 (2010) 426–433.

[24] A.L. Hill, D.G. Rand, M.A. Nowak, N.A. Christakis, Emotions as infectious diseases in a large social network: the SISa model, Proceedings of the Royal Society B: Biological Sciences 277 (2010) 3827–3835.

[25] N.A. Christakis, J.H. Fowler, Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior, ArXiv e-prints, 2011.

[26] H. Noel, W. Galuba, B. Nyhan, The "unfriending" problem: the consequences of homophily in friendship retention for causal estimates of social influence, Social Networks 33 (2011) 211–218.

[27] R. Lyons, The spread of evidence-poor medicine via flawed social-network analysis, Statistics, Politics, and Policy 2 (2011) 1–26.

[28] C.R. Shalizi, A.C. Thomas, Homophily and contagion are generically confounded in observational social network studies, Sociological Methods & Research 40 (2011) 211–239.

[29] D.M. Romero, B. Meeder, J. Kleinberg, Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion

on Twitter, in: Proceedings of the 20th international conference on World wide web, March 28-April 01, 2011, Hyderabad, India.

[30] Twitter api blog, 2011. http://blog.twitter.com/2011/09/one-hundred-million-voices.

[31] B. Gonalves, N. Perra, A. Vespignani, Modeling users' activity on twitter networks: validation of Dunbar's number, PLoS ONE 6 (2011) e22656.

[32] D.J. Watts, P.S. Dodds, Influentials, networks, and public opinion formation, Journal of Consumer Research 34 (2007) 441–458.

[33] R. Grannis, Six degrees of "who cares?", American Journal of Sociology 115 (2010) 991–1017.

[34] S.A. Golder, M.W. Macy, Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures, Science Magazine 333 (2011) 1878–1881.

[35] G. Miller, Social scientists wade into the tweet stream, Science Magazine 333 (2011) 1814–1815.

[36] M. Newman, The structure of scientific collaboration networks, Proceedings of the National Academy of Sciences of the United States of America 98 (2001) 404–409.

[37] M. Newman, Assortative mixing in networks, Physical Review Letters 89 (2002) 208701.

[38] I.M. Kloumann, C.M. Danforth, K.D. Harris, C.A. Bliss, P.S. Dodds, Positivity of the English language, PLoS ONE 7 (2012) e29484.

[39] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media, 2009.

[40] M. Jacomy, S. Heymann, T. Venturini, M. Bastian, Forceatlas2, A Graph Layout Algorithm for Handy Network Visualization, 2012. http://www.medialab.sciences-po.fr/publications/Jacomy_Heymann_Venturini-Force_Atlas2.pdf.

[41] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Review 51 (2009) 661–703.

[42] R. Dunbar, Neocortex size and group size in primates: a test of the hypothesis, Journal of Human Evolution 28 (1995) 287–296.

[43] M.P.H. Stumpf, C. Wiuf, R.M. May, Subnets of scale-free networks are not scale-free: sampling properties of networks, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 4221–4224.

[44] E. Sadikov, M. Medina, J. Leskovec, H. Garcia-Molina, Correcting for missing data in information cascades, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, ACM, New York, NY, USA, 2011, pp. 55–64.

[45] J. Leskovec, C. Faloutsos, Sampling from large graphs, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06, ACM, New York, NY, USA, 2006, pp. 631–636.

[46] S.H. Lee, P.-J. Kim, H. Jeong, Statistical properties of sampled networks, Physical Review E 73 (2006) 016102.

[47] T. Frantz, M. Cataldo, K. Carley, Robustness of centrality measures under uncertainty: examining the role of network topology, Computational & Mathematical Organization Theory 15 (2009) 303–328.

[48] G. Kossinets, Effects of missing data in social networks, Social Networks 28 (2006) 247–268.

Isabel Kloumann is a doctoral student of Mathematics at Cornell University holds a B.S. in Mathematics and Physics from the University of Vermont. She has explored the interplay between language and emotion by analyzing massive digital texts with a combination of human- and silicon-based supercomputers (see Amazon Mechanical Turk and the Vermont Advanced Computing Center). She is currently developing metrics for measuring happiness in digital human expressions, namely in Twitter data.

Kameron Decker Harris is a student in the Master of Science in Mathematics program at the University of Vermont studying applied mathematics. After receiving undergraduate degrees in mathematics and physics from UVM in 2009, he worked on bus transportation in Chile as a Fulbright scholar. Kameron is interested in mathematical modeling and data analysis for complex social, technological, and natural systems.

Chris Danforth received a B.S. in math and physics from Bates College in 2001, and a Ph.D. in Applied Mathematics and Scientific Computation from the University of Maryland in 2006. He is currently on the faculty of the University of Vermont where he combines mathematical modeling and big data to study a variety of complex biological, natural, and physical systems. Among other projects, he has applied principles of chaos theory to improve weather forecasts, and developed a real-time remote sensor of global happiness using Twitter. His research has been covered by the New York Times, Science Magazine, and the BBC among others. Descriptions of his projects are available at his website: http://uvm.edu/ cdanfort.

Peter Sheridan Dodds is an Associate Professor at the University of Vermont (UVM) working on system-level problems in many fields, ranging from sociology to physics. He maintains general research and teaching interests in complex systems and networks with a current focus on sociotechnical and psychological phenomena including contagion, problem-solving, and collective emotional states. His methods encompass large-scale sociotechnical experiments, large-scale data collection and analysis, and the formulation, analysis, and simulation of theoretical models. Dodds's training is in theoretical physics, mathematics, and electrical engineering with formal post-doctoral experience in the social sciences. He is Director of the UVM's Complex Systems Center, co-Director of UVM's Computational Story Lab, and a visiting faculty fellow at the Vermont Advanced Computing Center.

Catherine Bliss is a doctoral student of Mathematics at the University of Vermont, where she received the Graduate Research Fellowship from the UVM Complex Systems Center. She received her M.S. in Mathematics from the University of Vermont and her M.A. in Marine Affairs and Policy from the Rosenstiel School of Marine and Atmospheric Science at the University of Miami. She holds a B.A. in Psychology and Mathematics. Catherine is interested in computational tools to analyze complex networks.