Emotions in Social Networks: Distributions, Patterns, and Models

Shengmin Jin
Data Lab, EECS Department
Syracuse University
shengmin@data.syr.edu

Reza Zafarani
Data Lab, EECS Department
Syracuse University
reza@data.syr.edu

ABSTRACT

Understanding the role emotions play in social interactions has been a central research question in the social sciences. However, the challenge of obtaining large-scale data on human emotions has left the most fundamental questions on emotions less explored: How do emotions vary across individuals, evolve over time, and are connected to social ties?

We address these questions using a large-scale dataset of users that contains both their emotions and social ties. Using this dataset, we identify patterns of human emotions on five different network levels, starting from the user-level and moving up to the whole-network level. At the user-level, we identify how human emotions are distributed and vary over time. At the ego-network level, we find that assortativity is only observed with respect to positive moods. This observation allows us to introduce *emotional balance*, the "dual" of structural balance theory. We show that emotional balance has a natural connection to structural balance theory. At the community-level, we find that community members are emotionally-similar and that this similarity is stronger in smaller communities. Structural properties of communities, such as their sparseness or isolatedness, are also connected to the emotions of their members. At the whole-network level, we show that there is a tight connection between the global structure of a network and the emotions of its members. As a result, we demonstrate how one can accurately predict the proportion of positive/negative users within a network by only looking at the network structure. Based on our observations, we propose the *Emotional-Tie* model – a network model that can simulate the formation of friendships based on emotions. This model generates graphs that exhibit both patterns of human emotions identified in this work and those observed in real-world social networks, such as having a high clustering coefficient. Our findings can help better understand the interplay between emotions and social ties.

CCS CONCEPTS

• **Human-centered computing → Social networks**;

KEYWORDS

Emotions, Sentiments, Signed networks, Network Models

CIKM'17 , November 6–10, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4918-5/17/11...$15.00
https://doi.org/10.1145/3132847.3132932

1 INTRODUCTION

Social media has become the primary online venue for users to express their emotions. Emotions are shared via posts, comments, photos, tweets, likes, among other interactions. Emotions expressed by users are tightly coupled with their social relationships [31]. Social relationships are one of the strongest forces behind the formation of emotions; in return, emotions are known to regulate social life [31]. On most social media sites, users develop social relationships by befriending or following others, interacting with them, or forming small communities or joining larger ones. The abundance of emotion-carrying data and social relationships on social media facilitates the study of human emotions (1) across users, (2) over time, and (3) with respect to social ties. The findings of such a study have various implications:

Implications of Studying Emotion. Findings on how emotions are linked with social ties can be harnessed for emotion prediction. Traditionally, emotion prediction and sentiment analysis models have utilized textual data to predict emotions [18]. By identifying the connections between emotions and social ties, one can take an alternative route to emotion prediction by using social network information. Furthermore, studying emotions with respect to social ties can help answer questions such as: how correlated are my emotions to that of my friends [6]? are emotions linked to the structure of social networks? are happy users friends with other happy users (i.e., indicating *assortative mixing* [21])? or are sad users more likely to be found in sparse, or dense communities?

Studying emotions across social media users can also help better understand the types of emotions that are expressed online. For instance, one can investgiate whether social media is mostly used to share joy, or to vent frustration. Moreover, links between various user activities and emotions can be identified, e.g., is extensive social media usage correlated to negative emotions and depression [27]?

Pursuing a systematic study on how emotions are connected to social networks is challenging. The difficulty in conducting such a study lies in obtaining large-scale longitudinal data on users, containing both their social networks and emotions, where emotions are directly provided and not subjectively predicted (e.g., by sentiment classification [18]).

The Present Work. In this paper, using a large-scale longitudinal network of users and their emotions, we aim to identify patterns of emotions. Our goal is to study how emotions (1) vary across users, (2) evolve over time, and (3) are connected to social ties. Based on the identified patterns, we aim to propose a network model that can properly simulate friendship formations based on emotions.

To study emotions systematically, we analyze emotions on five different levels and make the following contributions:

I. User-Level Analysis (Section 3). At the most basic network level, we study users and their emotions. We identify (1) emotions that users express more, (2) how emotions are distributed across users (Section 3.1), and (3) how emotions vary over time or with increased activity (Section 3.2).

II. Ego-Level Analysis (Section 4). At the ego-level, we identify directed and undirected relationships that are formed between users depending on their emotions. We also study whether the emotions of one's friends are correlated to her future emotions.

III. Triads and Emotional Balance (Section 5). A natural extension to the ego-level analysis (i.e., two users) is to study emotions in sets of three connected users (a triad). We identify configurations of triads that are more common in social networks depending on the emotions expressed by their three members. Studying emotions in triads naturally connects this study to *structural balance theory*, which studies triads with friendly/antagonistic relationships, i.e., signed edges. We introduce *emotional balance*, which looks at triads of users with different emotions (signed nodes). We show that structural balance theory and emotional balance are connected.

IV. Community-Level Analysis (Section 6). We identify situations in which community members exhibit similar emotions. We investigate whether (1) such similarities are stronger in smaller communities and whether (2) structural properties of a community (e.g., its sparsity or isolatedness) are related to the emotions expressed by its members.

V. Network-Level Analysis (Section 7). Finally, we analyze user emotions at the whole network-level. We find that users exhibit distinct network-level patterns based on their emotions. Exploiting such patterns, we show that one can accurately predict the emotion that is dominant in the network using only the graph information.

The findings at different network levels allow us to study the relationship between friendship formations and user emotions. By modeling tie strength as a function of user emotions, we propose the *Emotional-Tie model* – a network model that can properly simulate friendship formations based on emotions. In Section 8, we will discuss this model along with its properties and limitations.

Before we delve into the details, we discuss our experimental setup and how data was prepared for our experiments next.

2 EXPERIMENTAL SETUP

To understand emotions at different network levels and over time, proper data is required. This data should have users and their emotions at different times. Such emotions can be directly provided by the users or can be indirectly obtained via sentiment classification on the content users generate (e.g., posts) [24]. As sentiment classification can be subjective and imprecise [18], it is preferable that emotion data is provided directly. The dataset should also contain directed and undirected relationships such as friendships between users and follower/followee relationships. Finally, for community-level analysis, community membership information is required. Community membership information can be provided directly or can be obtained using a community detection algorithm. Explicit community membership information is preferred as community detection can be subjective [7] and imprecise community membership information may impact our experimental outcomes. The LiveJournal website (http://www.livejournal.com/) provides data satisfying all of the above constraints and is an appropriate candidate for data collection for our experiments.

LiveJournal is a popular blogging and social networking site, where users can maintain a blog, journal, or a diary. When posting blogs, users have the option of reporting their emotion by selecting a *mood*. This mood can be selected from a predefined list of 132 common moods such as `happy` or `angry`, or can be entered as free-text, e.g., `:-)))`. The mood list is sorted alphabetically and no mood is preselected (i.e., no default mood). LiveJournal social network has both directed and undirected user-user relationships. Users on LiveJournal can be (1) friends (undirected/mutual relationship) or can have (2) follower or following relationship (directed relationship). LiveJournal has explicit communities. Users can choose to create or join a community. The site enables one to study emotions, how emotions evolve over time, and how emotions vary with respect to social network structure.

We have crawled LiveJournal using a BFS crawler starting from central nodes in the largest connected component. We collect more than 14.7 million posts, spanning more than 10 years of LiveJournal data. Each post is assigned a mood by the user publishing it. In addition, we collect around 1.13 million friendships (undirected), 14.1 million followers/followees (directed), and community memberships for all users. Our dataset is released for research purposes.[1]

Data Limitations. Unfortunately, LiveJournal does not provide timestamps on when social ties are formed or when users join communities. Given such timestamps, this work could be naturally extended to consider the causality (or pseudo-causality) relationships between emotions and social ties.

Data Preprocessing. We perform a set of preprocessing steps on the dataset. In particular, for consistency in sentiment analysis and removing meme-type moods, we only retain the posts that have their moods selected from the predefined list provided by LiveJournal. These posts account for the majority of posts in the dataset (85.96%). In addition, we only retain users that have 10 or more posts to exclude occasionally active or inactive users. Finally, we manually convert each mood in our dataset to its polarity (positive, negative, or neutral). After this step, all moods in our datasets are either positive (+), negative (−), or neutral (0).

Following data preparation, we conduct the following analyses.

3 USER-LEVEL ANALYSIS

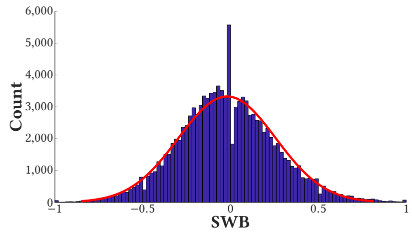
We begin by analyzing user-level emotions. In particular, we look at the (1) distribution of users with positive or negative emotions and the (2) longitudinal dynamics of emotions.

3.1 Distribution of Emotion

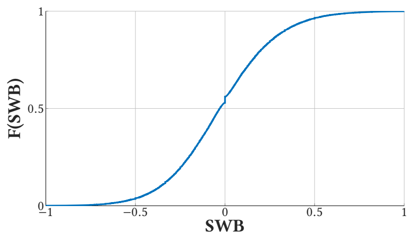
First, we need to determine how positive or negative a user is in general. We use the previously proposed Subjective Well-Being (SWB) [3] to achieve this goal. Let $S(u)$ denote the subjective well-being of user u. Subjective well-being is defined as the fractional difference between the number of positive and negative posts:

$$S(u) = \frac{N_p(u) - N_n(u)}{N_p(u) + N_n(u)}, \tag{1}$$

[1] http://data.syr.edu/get/EmotionPatterns

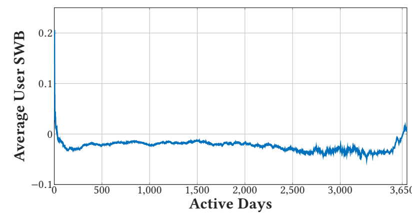


(a) Emotions Histogram with Normal Fit (Red Line)

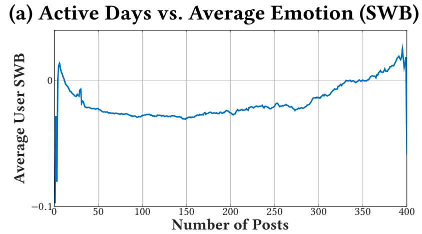


(b) Emotions Empirical CDF

Figure 1: User Emotion Distribution



(a) Active Days vs. Average Emotion (SWB)



(b) Number of Posts vs. Average Emotion (SWB)

Figure 2: Emotion Change with Time or Activity

Table 1: Distribution of Positive, Negative, and Neutral Users

Users	Number	Proportion
Positive (+)	50,705	43.92%
Negative (−)	61,066	52.90%
Neutral (0)	3,673	3.18%
Total	115,444	100.00%

where $N_p(u)$ and $N_n(u)$ represent the number of positive and negative posts for user u, respectively. Figure 1a plots the distribution of $S(u)$ values for all users. The distribution exhibits a spike at $S(u) = 0$ (users that have equal number of positive and negative posts). The distribution is approximately normal, as can be observed by the normal fit in the figure. However, the empirical Cumulative Distribution Function (CDF) of the $S(u)$ values (Figure 1b) reveals a slight skew towards users with more negative posts, i.e., $P(S(u) < 0) > 0.5$. Thus, we consider users with $S(u) > 0$ as positive (+), with $S(u) < 0$ as negative (−), and with $S(u) = 0$ as neutral (0) individuals.

Table 1 provides the distribution of positive, negative, and neutral users. The majority of users express negative emotions most of the time and negative users are almost 20% more than positive users. Neutral users account for 3% of the population.

Observing more negative users in social networks is in line with findings on *negativity bias* in psychology literature, indicating that humans are more likely to respond to negative events [2, 26, 29]. It also confirms recent discoveries indicating that (1) negative posts are expected to receive more feedback [28] and that (2) social media is used to vent negative emotions and frustrations [25].

3.2 Longitudinal Dynamics

Do users get sadder or happier over time? Does more activity on social media lead to sadness? To answer such questions, we investigate how emotions change (1) over time and (2) with increased activity. Figure 2a shows the average SWB versus the number of active days for users that have been active for up to 10 years on the site. The average SWB decreases until around 200 days and stays stable between 200 to 3,000 days. After that, the SWB value increases again. To summarize, (1) users who leave the site in a short time tend to leave positive posts; (2) Users that are very loyal to the site — who are active over 3,000 days — are also more positive; and (3) the remaining users maintain a slightly negative emotion. Similarly, to determine how emotions vary with increased activity, we plot the relationship between number of posts and the average SWB in Figure 2b. A similar trend is observed where people with limited or many posts are more positive.

4 EGO-LEVEL ANALYSIS

To understand how ego networks are formed with respect to emotions, we study undirected and directed ego networks.

Undirected Ego Networks. To study emotions in undirected ego networks, we look at the friendships that are between negative and positive users. In large-scale networks, assortativity is a common pattern [21], where users of the same type connect to one another. Due to assortativity, we expect positive users to be more connected to positive users and negative users to be more connected to negative ones. Recent studies have shown evidence of assortativity with respect to sentiment [3]. Hence, we expect to see edges $(+, +)$ and $(-, -)$ more than expected.

Among + and − users, three types of friendships are possible: $(+, +)$, $(+, -)$, and $(-, -)$. Table 2 provides the distribution of these types of friendships (edges) for our dataset. To understand the significance of these values, one has to compare them with the expected values. To compute expected values, we maintain the proportions of + and − users (Table 1) and the network structure, but shuffle user (node) emotions randomly. After shuffling, to compute the expected values, we count each type of edge again.[2] Clearly, when expected values are larger than observed values, the type of friendship is underrepresented and when the opposite takes place, the

[2] Expected values can also be computed from Table 1.

Table 2: Friendships (Undirected Edges) Distribution

Edge	Number	Proportion	Expected	Surprise
$(+,+)$	70,081	30.35%	20.60%	115.92
$(+,-)$	110,084	47.68%	49.57%	-18.20
$(-,-)$	50,736	21.97%	29.83%	-82.56
Total	230,901	100.00%	100.00%	

Table 3: Follower/Followee (Directed Edges) Distribution

Edge	Number	Proportion	Expected	Surprise
$\xrightarrow{(+,+)}$	415,310	29.10%	20.57%	251.88
$\xrightarrow{(+,-)}$	334,455	23.43%	24.78%	-37.25
$\xrightarrow{(-,+)}$	345,818	24.23%	24.79%	-15.59
$\xrightarrow{(-,-)}$	331,739	23.24%	29.86%	-172.64
Total	1,427,322	100.00%	100.00%	

friendship type is overrepresented. To understand if the difference between expected values and those that are observed is significant, we compute *surprise* (see Ref. [15] for details). A surprise value on the order of tens is highly significant; hence, all observed values are significant with p-values almost equal to zero.

From Table 2, we observe that though there are more negative users than positive users in the network, friendships among positive users $(+,+)$ are far more than those between negative users $(-,-)$. Moreover, while $(+,+)$ friendships are significantly overrepresented, $(-,+)$ and $(-,-)$ are significantly underrepresented. This indicates that with respect to emotions, *only positive users demonstrate assortativity* and negative users are neither connected [as expected] to other negative or positive users. This contradicts the common perception that networks with respect to emotion and sentiment are assortative [3]. Note that friendships are mutual, where both parties need to approve of the friendship. For instance, one can speculate that $(+,-)$ are far less than expected as positive users do not approve of such friendships, whereas negative users find the friendship of interest. However, to systematically investigate such speculations, we can analyze directed ego networks.

Directed Ego Networks. We extend our previous study to directed networks where follower/followee types of edges are present. Table 3 provides the distribution of the four types of directed edges. The links from positive to positive users $\xrightarrow{(+,+)}$ are slightly higher than other types. Other links are almost equally likely. After comparing with the expected values, we find $\xrightarrow{(+,+)}$ links to be significantly overrepresented, and the other three links are significantly underrepresented. The results is in accord with our observations in undirected ego networks that only positive users demonstrate assortativity. We also observe that $\xrightarrow{(+,-)}$ links are much more underrepresented than the opposite direction $\xrightarrow{(-,+)}$, which shows that $+$ users are less likely to follow $-$ users than the opposite.

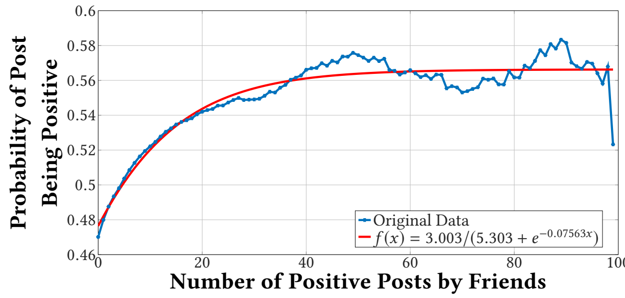


Figure 3: The Probability of a Post being Positive as a Function of Number of Positive Posts by Friends.

Emotion Correlations to Friends. We studied emotions commonly shared among ties. Our static analysis did not consider emotion temporality or how emotions of friends (ties in the ego network) are correlated to those expressed by the user. Here, we bridge this gap and study temporal dynamics of emotion in ego networks.

A user's emotion can be correlated to the users he or she follows. Hence, for each post of a user, we gather all posts by his or her friends (i.e., *the user's feed*) published in the week prior to the post. The choice of one week was based on previous studies [10]. Following the tradition in diffusion of innovation studies [1, 12], we enumerate the number of positive[3] posts in this one-week feed. Hence, we identify the number of positive posts that the user may have read during the week before he or she published the post. As we know whether each post is positive or negative, we can calculate the probability of a post being positive given the number of positive posts published by the user's friends one week prior to the post getting published. Figure 3 shows that with the increase in the number of positive posts in the feed, the probability of a post being positive also increases. This result is in line with previous diffusion of innovations studies that have shown that the adoption rate of an innovation as a function of number of friends adopting it follows an S-shaped curve [1, 12]. In fact, a logistic S-curve, $f(x) = 3.003/(5.303 + e^{-0.07563x})$, fits the plot in Figure 3 with $R^2 \approx 0.89$. While this result neither indicates causality nor proves an emotional contagion, it demonstrates the predictive power of friends' emotions in predicting future user emotions. In fact, a confounding factor (e.g., an earthquake) might impact the emotions of friends and the user simultaneously.

5 EMOTIONAL BALANCE

Consider the following scenario: A and B are friends, and A is a positive person and B is a negative. They have a mutual friend C. Is C more likely to be positive or negative? And if A and B are both positive (or both negative), what will C be? To answer these questions, we define *emotional balance*. Similar to the structural balance theory, we consider all the possible ways in which the three users in a triangle can be signed. Unlike structural balance theory, we assign signs to nodes with user-level emotions, but not to edges. We analyze both undirected and directed networks and identify more-frequent-than-expected (i.e., *emotionally balanced*) triads.

[3] The same experiment was performed for negative posts with similar results which we do not include for brevity.

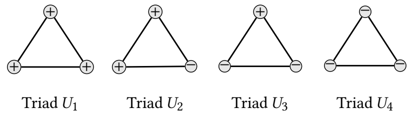
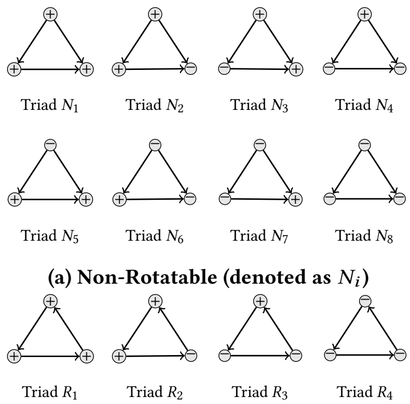


Figure 4: Undirected Signed Triads (denoted as U_i)



(a) Non-Rotatable (denoted as N_i)

(b) Rotatable (denoted as R_i)

Figure 5: Directed Signed Triads

Undirected Networks. As shown in Figure 4, four types of undirected triads (up to symmetry) exist in undirected networks. We count each triad. To assess the significance of these counts, we shuffle the signs of all nodes, while maintaining the fractions of positive and negative users, and count these triads again after shuffling. We shuffle for 1,000 times and get the average expected triad counts and their standard deviations. The results are provided in Table 4.

Table 4 shows that triads U_1 and U_2 are significantly overrepresented (balanced), while U_3 and U_4 are significantly underrepresented (unbalanced). Comparing the surprise values, we identify that *triads with two or more positive users are emotionally balanced.* This observation shows that positive users are inclined to form groups with other positive users; however, as negative members increase, groups are less likely to form. Interestingly, U_3 is more underrepresented than U_4. We speculate that positive users are less willing to be part of groups formed mostly from negative users.

We can now answer the questions asked at the beginning of this section: if A and B are both positive, or at least one of them is positive, C is more likely to be positive; if A and B are both negative, C is slightly more likely to be negative.

Directed Networks. In directed networks, nodes are connected via follower-following relationships. Different from undirected networks, directed networks have 12 types of triads, eight of which are non-rotatable and the other four can be rotated (see Figures 5a and 5b). We count the triads and compute the expected number of triads via shuffling. The results are provided in Tables 5 and 6.

The results show that triads N_1, N_2, N_3, N_5, R_1, and R_2 are significantly overrepresented (balanced) and the others are significantly underrepresented. Similar to undirected networks, triads with two or more positive users are balanced.

Connection to Structural Balance Theory. Emotional balance has a direct connection to structrual balance theory [5, 9]. Structural balance theory affirms that in networks where positive/negative edges indicate friendly/hostile relationships, there are more triangles with even number of negative edges. Thus, balance theory confirms that "a friend of a friend is a friend" or "a friend of an enemy is an enemy." Structural balance theory discusses more frequent triangles in networks with positive and negative edges and emotional balance discusses more frequent triangles with positive and negative nodes. To connect the two theorems, one has to identify the sign of an edge, given the sign of the nodes it connects. The edge signs can be identified from our results provided in Table 2. We can see that only edges between two positive nodes, i.e., (+,+), are overrepresented. All other edges that include a negative node (−) are underrepresented. Hence, we can consider edges between two positive nodes as positive edges and all other edges where one endpoint is a negative node as a negative edge. Thus,

> *"Connections between two positive users are positive.*
> *Connections involving a negative user are negative."*

Following this approach, we see that all triangles in Figure 4 that are considered balanced by emotional balance, are also considered balanced by structural balance theory. For instance, Triad U_2 which is balanced by emotional balance, is also balanced by structural balance theory as after edge sign assignments it contains 1 positive edge and 2 negative edges. Similarly, Triad U_3 which is considered imbalanced by emotional balance, is also imbalanced by structural balance as after assigning edge signs it contains 3 negative edges.

6 COMMUNITY-LEVEL ANALYSIS

To investigate emotions in communities, we investigate (**I**) whether community members are emotionally similar and (**II**) whether structural properties of communities are related to the emotions of their members. To study the former, labeled community membership information is required and to investigate the latter, metrics that describe structural properties of communities are needed.

I. Are Community Members Emotionally-Similar? Here, we investigate two questions. First, we identify whether a user's emotion is connected to the overall emotion of the communities she has joined. Second, we determine whether this connection depends on the size of the community. We speculate that members of smaller communities to be more emotionally similar to each other.

We first define a community's subjective well-being S_C to be the average subjective well-being of its members ($S(u)$ values):

$$S_C = \frac{1}{|C|} \sum_{u \in C} S(u), \tag{2}$$

where C is the community and $|C|$ denotes its size.

For each user, we calculate the average SWB of the communities that the user has joined, after removing the user from these communities. We consider a user to be emotionally-similar to her communities, when the sign of this average value and that of her SWB match. The signs match for 58% of the users, indicating that most community members are, on average, emotionally similar.

Next, we follow a similar approach and compute the proportion of users for which their SWB's sign matches that of the average SWB of all communities of size k that they have joined. Figure 6

Table 4: Distribution of Undirected Triads

Triad U_i	$\lvert U_i \rvert$	E($\lvert U_i \rvert$): **Expected** $\lvert U_i \rvert$ **after shuffling (std. dev.)**	**P**(U_i)	E(**P**(U_i)) **(std. dev.)**	**Surprise**
U_1: $+++$	60,384	24,433.81 (2,179.87)	23.04%	9.32% (0.83%)	241.66
U_2: $++-$	106,245	88,329.49 (3,083.68)	40.55%	33.71% (1.18%)	74.02
U_3: $+--$	74,399	106,483.39 (2,120.52)	28.39%	40.63% (0.81%)	-127.54
U_4: $---$	21,008	42,789.31 (3,204.84)	8.02%	16.33% (1.22%)	-115.12
Total	262,036				

Table 5: Distribution of Rotatable Directed Triads

Triad R_i	$\lvert R_i \rvert$	E($\lvert R_i \rvert$): **Expected** $\lvert R_i \rvert$ **after shuffling (std. dev.)**	**P**(R_i)	E(**P**(R_i)) **(std. dev.)**	**Surprise**
R_1: $+++$	913,683	378,049.32 (15,917.34)	22.61%	9.36% (0.39%)	915.00
R_2: $++-$	1,595,178	1,364,136.18 (21,500.88)	39.48%	9.36% (0.53%)	243.05
R_3: $+--$	1,176,753	1,640,888.97 (15,863.67)	29.12%	40.61% (0.39%)	-470.15
R_4: $---$	355,344	657,883.50 (22,220.46)	8.79%	16.28% (0.55%)	-407.65
Total	4,040,958				

Table 6: Distribution of Non-Rotatable Directed Triads

Triad N_i	$\lvert N_i \rvert$	E($\lvert N_i \rvert$): **Expected** $\lvert N_i \rvert$ **after shuffling (std. dev.)**	**P**(N_i)	E(**P**(N_i)) **(std. dev.)**	**Surprise**
N_1: $+++$	1,316,055	552,917.31 (20,755.11)	22.24%	9.34% (0.35%)	1077.5
N_2: $++-$	765,223	665,765.25 (11,463)	12.93%	11.25% (0.19%)	129.3
N_3: $+-+$	775,262	665,405.19 (8,718)	13.10%	11.24% (0.15%)	142.9
N_4: $+--$	571,829	801,288.96 (10,310.46)	9.66%	13.54% (0.17%)	-275.6
N_5: $-++$	790,960	665,551.50 (11,885.61)	13.37%	11.25% (0.20%)	163.1
N_6: $-+-$	581,457	801,385.11 (6,394.29)	9.83%	13.54% (0.11%)	-264.2
N_7: $--+$	584,889	800,928.03 (9,984.96)	9.88%	13.53% (0.17%)	-259.6
N_8: $---$	532,109	964,542.66 (29,768.46)	8.99%	16.30% (0.50%)	-481.3
Total	5,917,784				

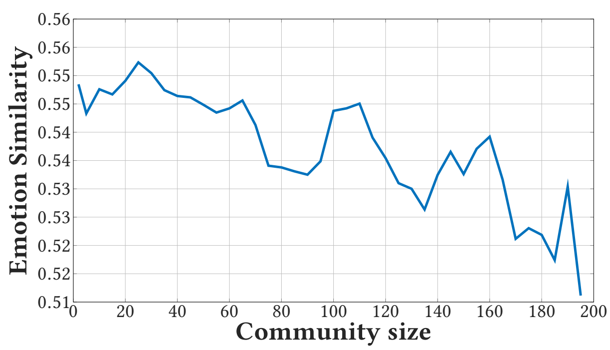
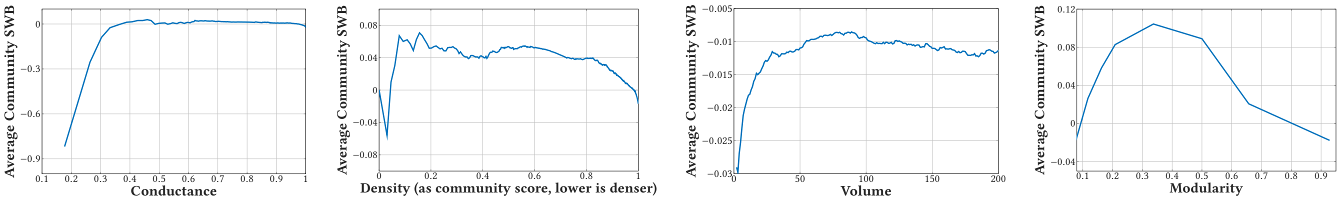


Figure 6: User-Community Emotion Similarity

provides the results for different k. The figure shows that members of smaller communities are more emotionally similar, confirming our initial speculation.

II. Are Structural Properties of Communities Related to their Members' Emotions? We observed that community members are emotionally similar. In this section, we investigate whether communities with different network structures have members with different types of emotions. This allows us to reach insights such as "dense subgraphs contain positive users."

To assess the structure and quality of a community, community detection metrics are often employed. Here, we use well-known measures such as (1) conductance, (2) internal density, (3) volume, and (4) modularity to describe communities [16]. Conductance of a community is the fraction of total edge volume that points outside the community. Low conductance indicates that the community is separated from the rest of the graph whereas high conductance indicates that the community is well-connected to the rest of the graph. Internal density, as the name suggests, is the edge density of the community. Here, for consistency reasons with other measures, we use (1−density) as our measure such that tightly-knits communities have low density values and sparse communities have high density values. Volume is the sum of degrees (i.e., number of friends) of members of the community. Finally, modularity quantifies how different the community is from a community formed with the same degree sequence in a random graph. A low modularity



(a) Conductance **(b) Internal Density** **(c) Volume** **(d) Modularity**

Figure 7: Community Measures vs. Emotion of Community Members

value indicates that a community appears to be random and a high value indicates that a community is statistically significant.

To assess the relationship between community structure and member emotions, for each community, we compute both its community Subjective Well-Being (Equation 2) and the values of the aforementioned measures. Figure 7 shows the relationship between community measures and the SWB values. From the figure, we observe the following community-level patterns (denoted as *C1-3*):

C1. Members of Isolated Communities are Extremely Negative. Figures 7a and 7b show that when conductance or volume is low; that is, when the community is isolated from the rest of the network, the average community SWB is extremely low. When communities are well-embedded (conductance > 0.4), members are mostly neutral.

C2. Members of Super Dense (or Super Sparse) Communities are Mostly Negative. Figure 7b shows that when the internal density of a community is close to 0 (a dense community) or 1 (a sparse community), community members are mostly negative. We observe a similar pattern in Figure 7d, where most members of communities with very low or very high modularity values are negative. Very low modularity indicates sparse communities formed in random graphs [21].

C3. Communities with Moderate Densities contain Positive Members. Figure 7b also shows that for most communities with a moderate density (not too sparse or dense), the community subjective well-being is more positive.

These observations have further implications in community detection. Community detection algorithms often tend to minimize/maximize a community detection measure. The connections between subjective well-being of a community and some of these measures indicate that user emotions may prove useful as an alternative data source to detect communities. However, we have not pursued this path further and leave it as a part of our future work.

7 NETWORK-LEVEL ANALYSIS

In Section 4, specifically in Table 2, we showed that: (1) friendships among positive users $(+,+)$ are significantly more frequent than $(-,+)$ and $(-,-)$; (2) $(-,+)$ and $(-,-)$ are both significantly underrepresented; and (3) $(-,+)$ friendships are much more frequent, and less underrepresented, than $(-,-)$, indicating that negative users prefer befriending positive users rather than other negative users.

These observations indicate that there might be a connection between user emotions and a network property known as *core-periphery* [4]. Social networks often exhibit a core-periphery structure [32], where they consist of a dense cohesive core and a sparse, loosely connected periphery. Here, we speculate that positive users act as the core, and negative users form the periphery. However, to systematically investigate this speculation, we need a method that can capture the core-periphery structure of a network.

Stochastic Kronecker graphs can effectively model core-periphery structure in real-world networks [13]. Stochastic Kronecker graph is a generative model that can capture a long list of properties of real-world networks, including the core-periphery structure, using Kronecker graph product. In Stochastic Kronecker graphs, given the adjacency matrix of a graph $A \in \mathbb{R}^{n^k \times n^k}$, we aim to learn a small probability matrix $P \in \mathbb{R}^{n \times n}$, known as the *Kronecker initiator matrix*, such that the k^{th} Kronecker power of P (i.e., $\underbrace{P \otimes P \cdots \otimes P}_{k \text{ times}}$) is most likely to have generated A, i.e., $P(A|P)$ is maximized (for further details refer to Ref. [13]). The KronFit algorithm can estimate the Kronecker initiator matrix for a real-world graph using the maximum likelihood principle [13]. Assume that KronFit is used to fit a 2×2 initiator matrix $I = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ to a network that exhibits a core-periphery structure. Then, a represents the core strength and is large, as most edges are inside the core; by contrast, d is small, as very few edges exist among the periphery nodes. In an undirected friendship network, where the adjacency matrix is symmetric, the Kronecker initiator matrix learned is also symmetric, i.e., $b = c$.

To model core-periphery within the emotion network, we sample many subgraphs from our undirected friendship network. In sampled subgraphs, we vary (1) the number of nodes from 5,000 to 50,000 and (2) the proportion of positive nodes from 0% to 100%. For each possible set of parameters (# nodes, positive proportion), we generate 25 independently sampled subgraphs. For each subgraph, we compute the 2×2 Kronecker initiator matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. We compute the average core strengths (average a value) among the 25 subgraphs that were generated for the same set of parameters. Figure 8 provides the relationship between the average core strength and the proportion of users for graphs with various numbers of nodes. We observe that (1) increasing the proportions of positive nodes while maintaining the graph size and (2) increasing the number of nodes, both strengthen the core.

These observations link user emotions with global network structure, allowing one to offer insight into the other. For example, one can only use network information to determine whether the majority of users within a network are positive or negative. To verify this claim, we take the elements (a, b, d) of the Kronecker initiator matrix as features, and train a cross-validated linear regression model to predict the percentage of positive users. Figure 9 shows the Root Mean Square Error (RMSE) for various network sizes and positive

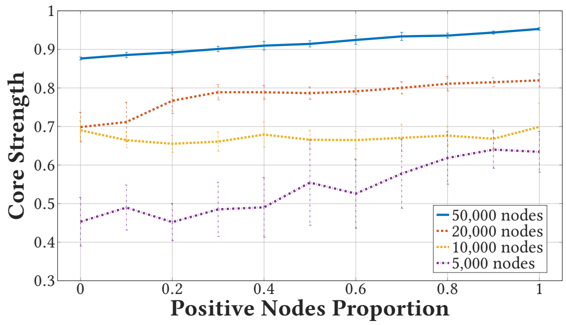


Figure 8: Positive Nodes Proportion vs. Core Strength

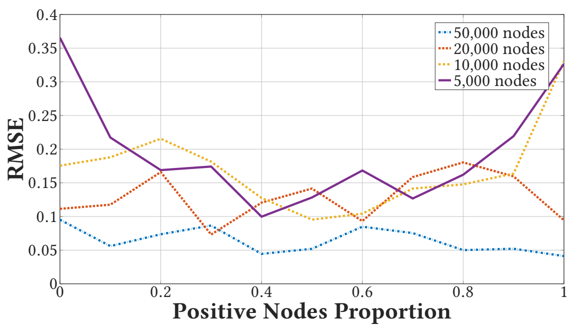


Figure 9: Positive Proportion Prediction Error

proportions. While in smaller graphs we have higher errors, in most cases the RMSE is reasonable and less than 0.2, indicating that if the percentage of positive nodes is α, our estimate is often within $\alpha \pm 0.2$. We also used our regression framework for classification purposes. Given a network, we attempted to predict whether the majority of users in the network are positive or negative (binary classification). We considered networks in which the majority of users ($\geq 70\%$) were either positive or negative. For each network, we predicted the proportion of positive or negative users using the aforementioned linear regression model. A prediction above 0.5 was considered a majority positive; otherwise, a majority negative. For each set of parameters (# nodes, positive proportion), we generated 10 randomly sampled networks and computed the fraction of networks in which the predictions were correct, i.e., the accuracy. Table 7 reports these accuracy rates. The results show that such a simple thresholding approach, using only three features, leads to high accuracy rates. We believe this leaves doors open for further investigation using more sophisticated features and classifiers.

Table 7: Accuracy of Predicting Majority Emotion

	Actual (+) Proportion	50K nodes	20K nodes	10K nodes	5K nodes
Majority +	1.0	100%	100%	80%	100%
	0.9	100%	100%	100%	100%
	0.8	100%	90%	90%	100%
	0.7	100%	90%	90%	90%
Majority −	0.3	100%	100%	80%	70%
	0.2	100%	90%	80%	80%
	0.1	100%	100%	100%	90%
	0.0	100%	100%	100%	80%

8 EMOTIONAL-TIE MODEL

Can we design a network model that explains how friendships are formed with respect to emotions and exhibits the same identified emotional patterns? The *Emotional-Tie* model is our attempt at designing such a model.

We start by capturing the core-periphery property in the model. One explanation on how core-periphery is formed within a network is that in such networks *the tie strength between two nodes is a function of how close both nodes are to the core center* [4]. Borgatti and Everett [4] propose one such tie strength function $\delta_{ij} = c_i c_j$, where δ_{ij} is the tie strength between nodes i and j, and c_i and c_j are non-negative values between 0 to 1 that denote the levels of coreness for nodes i and j, respectively. The function has high values for pairs of nodes both high in coreness, low values for pairs both in the periphery, and middling values if one is high in coreness and the other is not. Inspired by their approach and based on our findings in Section 7, we define *emotional coreness* for user u as $e_u = (S(u) + 1)/2$. Emotional coreness is a bijection that rescales SWB of a user from $[-1, 1]$ to $[0, 1]$, maintains its ordering, and still follows the same normal distribution. After this mapping, emotional coreness of negative users lies in $[0, 0.5)$ and that of positive users is in $(0.5, 1]$. Once emotional coreness is computed, we can define *emotional tie-strength* e_{ij} between users i and j as $e_{ij} = e_i \cdot e_j$. The Emotional-Tie model is simply a network model in which users i and j become friends with probability $P(i, j)$ equivalent to their emotional tie-strength e_{ij}.

The Emotional-Tie model is a variant of the recently discovered Random Dot Product Graphs [22]. In this model, an interest vector x_i, drawn from a specific distribution, is associated with node i. The probability of edge formation between nodes i and j is defined as $P(i, j) = f(x_i \cdot x_j)$, a function of the dot product of the interest vectors. It has been proven [22], that if the distribution that x_i is drawn from is uniform and function f is the identity function, the model exhibits properties such as power-law degree distribution, clustering, and a short diameter. In the Emotional-Tie model, function f is the identity, the interest vectors are 1-dimensional emotional coreness values, and the vectors are drawn according to the normal distribution.

Following the computation of emotional tie-strengths, we simulate a graph using the Emotional-Tie model. Table 8 and 9 provide the distribution of edges and triads in this simulated graph. Not only percentages are close to those of the original graph, but also the same patterns of significance are observed. In particular, edge $(+, +)$ and triads U_1 and U_2 are significantly overrepresented, while others are significantly underrepresented. However, we notice that the simulated graph is significantly denser than the original graph. This is a limitation of the model, as we are only considering the emotional compatibility between all pairs of nodes and ignore other factors such as the probability of two individuals meeting online. Next, we will discuss some of the properties and limitations of the Emotional-Tie model in more detail.

Table 8: Friendships Distribution in Model-Generated Graph

Edge	Number	Proportion	Surprise
(+, +)	481,928,369	32.21%	11,105.0
(+, −)	734,445,171	49.09%	-372.47
(−, −)	279,790,622	18.70%	-82.56
Total	1,496,164,162	100.00%	

Table 9: Triad Distribution in Model-Generated Graph

Triad U_i	$\lvert U_i \rvert$	$\mathbf{P}(U_i)$	Surprise
U_1: + + +	1.22×10^{12}	30.01%	1,434,400
U_2: + + −	1.81×10^{12}	44.44%	457,620
U_3: + − −	0.89×10^{12}	21.94%	-767,070
U_4: − − −	0.15×10^{12}	3.61%	-693,600
Total	4.06×10^{12}		

8.1 Properties and Limitations

As shown in Sections 3 and 8, the user SWB, and in turn emotional coreness, follow a normal distribution $\mathcal{N}(\mu, \sigma^2)$, where μ is the mean and σ^2 is the variance. In our dataset, μ is 0.5 and σ^2 is 0.019. We denote $g(x|\mu, \sigma^2)$ as the probability density function. Let n denote the number of nodes in the model-generated graph.

Theorem 1. ***(Density)*** *The expected number of edges in a graph generated by the Emotional-Tie model* $\mathbf{E}(|Edges|) = \binom{n}{2} \cdot \mu^2$.

Proof. For any two nodes i and j, the edge probability $P(i, j) = \mathbf{E}(e_i \cdot e_j) = \iint_0^1 e_i e_j g(e_i) g(e_j) \mathrm{d}e_i \mathrm{d}e_j = \mu^2$ as e_i and e_j are independent variables and $\int_0^1 x g(x) \mathrm{d}x = \mu$. Hence,

$$\mathbf{E}(|\text{Edges}|) = \binom{n}{2} \cdot P(i, j) = \binom{n}{2} \cdot \mu^2.$$

□

We have more than 1.1×10^5 nodes, so $\mathbf{E}(|\text{Edges}|) = \binom{1.1\times10^5}{2} \cdot 0.5^2 \approx 1.5125 \times 10^9$, which is very close to the number of edges in the simulated graph. The theorem shows one of the limitation of this model: graphs generated are much denser than their real-world counterparts. To mitigate this limitation, one can add a random coin flip to keep α percentages of edges that are to be formed ($\alpha = 0.1\%$ in our dataset) to achieve the same density. Our experiments show that adding such random coin flips result in graphs that have correct densities and exhibit the same patterns for edges and triads.

Theorem 2. ***(Triads)*** *The expected number of triads in a graph generated by the Emotional-Tie model* $\mathbf{E}(|Triads|) = \binom{n}{3} \cdot (\mu^2 + \sigma^2)^3$.

Proof. For any three nodes i, j, k,

$$\begin{aligned} P(ijk \text{ forms a triad}) &= \mathbf{E}(e_i^2 \cdot e_j^2 \cdot e_k^2) \\ &= \iiint_0^1 e_i^2 e_j^2 e_k^2 g(e_i) g(e_j) g(e_k) \mathrm{d}e_i \mathrm{d}e_j \mathrm{d}e_k \\ &= (\mu^2 + \sigma^2)^3, \end{aligned}$$

as e_i, e_j and e_k are independent and $\int_0^1 x^2 g(x) \mathrm{d}x = \mu^2 + \sigma^2$. So,

$$\mathbf{E}(|\text{Triads}|) = \binom{n}{3} \cdot P(ijk \text{ forms a triad}) = \binom{n}{3} \cdot (\mu^2 + \sigma^2)^3.$$

□

For our dataset, $\binom{1.1\times10^5}{3} \cdot (0.5^2 + 0.019)^3 \approx 4.318 \times 10^{12}$, which is close to the total number of triads in the simulated graph.

Theorem 3. ***(Clustering)*** *Emotional-Tie model exhibits clustering, i.e.,* $P(i, j|(i, k) \wedge (j, k)) > P(i, j)$.

Proof. For any three nodes i, j, k,

$$\begin{aligned} P(i, j|(i, k) \wedge (j, k)) &= \frac{P(ijk \text{ forms a triad})}{P((i, k) \wedge (j, k))} \\ &= \frac{\iiint_0^1 e_i^2 e_j^2 e_k^2 g(e_i) g(e_j) g(e_k) \mathrm{d}e_i \mathrm{d}e_j \mathrm{d}e_k}{\iiint_0^1 e_i e_j e_k^2 g(e_i) g(e_j) g(e_k) \mathrm{d}e_i \mathrm{d}e_j \mathrm{d}e_k} \\ &= \frac{(\mu^2 + \sigma^2)^3}{\mu^2 \cdot (\mu^2 + \sigma^2)^2} = (\mu + \frac{\sigma^2}{\mu})^2 > \mu^2 = P(i, j). \end{aligned}$$

□

Hence, in the Emotional-Tie model, two users are more likely to be friends if they have a common friend.

9 RELATED WORK

In addition to previous related research discussed throughout the paper, our work has links to the following areas:

I. Opinion Mining and Sentiment Analysis. Research in sentiment analysis and opinion mining [18, 23] has traditionally focused on means to classify, extract, or summarize opinions or sentiments. Less frequently, it has focused on mining sentiment patterns. An example includes the study by Mishne et al. [20] in which the authors also use LiveJournal as their data source. The authors built an interesting system to track the usage level of certain moods (e.g. worried) and connect mood usage to online events. Moreover, while network information has been used to classify sentiments [30], the connection between sentiments and social ties is less explored. Exceptions include studies that have utilized link information to investigate assortativity with respect to sentiments. For instance, Bollen et al. [3] show that online social networks may be subject to social mechanisms that cause assortativity at the sentiment-level. They measure correlations between sentiments expressed by twitter users and that of their friends to show sentiment assortativity. Our work extends previous research by identifying sentiment patterns within different levels of social networks.

II. Signed Networks. We looked at nodes in social networks that carried emotions. Hence, our study seems to complement studies in signed social networks where edges are signed [14]. The study by Leskovec et al. [15] is one example. The authors study relations between users and how their positive or negative relationships influence the structure of online social networks. They connect their findings to the classical theory of structural balance and discuss an alternative theory of status.

III. Cascades and Emotional Contagion. Research in information cascades [8, 11] has looked at how innovations, rumors, among other entities can propagate through a social network. Hence, cascade studies have looked at the interplay between an innovation being propagated and a social network. This indirectly connect these studies to that of ours as we are analyzing emotions and social networks. More recently, Coviello and his colleagues [6] and

Miller et al. [19] have studied sentiment propagation in online social networks. Assessing the level of influence or propagation that can happen with respect to emotions in online social networks is beyond the scope of this work, but the results obtained in this paper may help better understand emotion contagion in online social networks. Our result about emotion correlations to friends is aligned with, and also a complement of, the study by Kramer et al. [10]. While they identify posts to be positive or negative by sentiment classification, our emotion data is provided by users directly.

IV. Emotion Disclosure. Research in emotion disclosure has focused on studying the need for emotional expression on social networks [17]. As all emotions in our datasets our disclosed, our study complements such studies by looking at the interplay between expressed emotions and social ties.

10 CONCLUSIONS

We have identified the following emotional patterns at the user, ego, triad, community and the whole-network level:

▷ **User-Level Patterns.** There are more negative users that positive ones (negativity bias) and prolonged or more intense social media usage does not necessarily lead to sadness.

▷ **Ego-Level Patterns.** Assortativity is *only* observed with respect to positive moods, i.e., happy users are friends with happy users, but sad users are less often friends with other happy or sad users. Aligned with recent studies in product adoption and viral marketing, an *S*-curve pattern is observed where a user is more likely to express an emotion as more friends express similar emotions.

▷ **Triad-Level Patterns.** Most triads are formed from two or more positive users. We denote these triads as being *emotionally-balanced*. By considering connections between two positive users as positive and connections involving a negative user as negative, all emotionally balanced triads are also structurally balanced and vice versa.

▷ **Community-Level Patterns.** Community members have similar emotions. This similarity is stronger in smaller communities. A community's structure is connected to its members' emotions.

▷ **Network-Level Patterns.** Users with positive emotions form a core; users with negative emotions form its periphery. Exploiting only the core-periphery pattern, one can accurately predict whether the majority of users within a network are positive or negative.

Given user emotions, the Emotional-Tie model can simulate friendships and exhibit the aforementioned emotional patterns.

Further Directions. Findings in this study can be utilized for various applications. Examples include (1) exploiting emotion correlations to detect communities, (2) using emotional balance to recommend (or filter) friends, and (3) utilizing emotions (among other content) to predict network structures or vice versa.

Even with a detailed dataset, formulating basic questions and experiments has been challenging. We consider elaboration of further questions on emotions a future direction. In particular, emotion dynamics was less explored in this paper. Studying emotion dynamics helps answer questions such as how quickly do users recover from negative moods? or how do emotions evolve in communities?

ACKNOWLEDGMENTS

We thank Fred Morstatter and anonymous reviewers for constructive criticisms and suggestions that helped improve this work.

REFERENCES

[1] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. 2006. Group formation in large social networks: membership, growth, and evolution. In *KDD*. 44–54.

[2] Roy F Baumeister, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D Vohs. 2001. Bad is stronger than good. *Review of general psychology* 5, 4 (2001), 323.

[3] Johan Bollen, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. 2011. Happiness is assortative in online social networks. *Artificial life* 17, 3 (2011), 237–251.

[4] Stephen P Borgatti and Martin G Everett. 2000. Models of core/periphery structures. *Soc. networks* 21, 4 (2000), 375–395.

[5] Dorwin Cartwright and Frank Harary. 1956. Structural balance: a generalization of Heider's theory. *Psych. rev.* 63, 5 (1956), 277.

[6] Lorenzo Coviello, Yunkyu Sohn, Adam DI Kramer, Cameron Marlow, Massimo Franceschetti, Nicholas A Christakis, and James H Fowler. 2014. Detecting emotional contagion in massive social networks. *PloS one* 9, 3 (2014), e90315.

[7] Santo Fortunato. 2010. Community detection in graphs. *Physics reports* 486, 3 (2010), 75–174.

[8] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. 2013. Information diffusion in online social networks: A survey. *ACM SIGMOD Record* 42, 2 (2013), 17–28.

[9] Fritz Heider. 1946. Attitudes and cognitive organization. *The Journal of psychology* 21, 1 (1946), 107–112.

[10] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *PNAS* 111, 24 (2014), 8788–8790.

[11] Kristina Lerman and Rumi Ghosh. 2010. Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. *ICWSM* 10 (2010), 90–97.

[12] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. 2007. The dynamics of viral marketing. *TWEB* 1, 1 (2007), 5.

[13] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. 2010. Kronecker graphs: An approach to modeling networks. *JMLR* 11, Feb (2010), 985–1042.

[14] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Predicting positive and negative links in online social networks. In *WWW*. 641–650.

[15] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed networks in social media. In *CHI*. 1361–1370.

[16] Jure Leskovec, Kevin J Lang, and Michael Mahoney. 2010. Empirical comparison of algorithms for network community detection. In *WWW*. 631–640.

[17] Han Lin, William Tov, and Lin Qiu. 2014. Emotional disclosure on social networking sites: the role of network structure and psychological needs. *Comp. in Hum. Beh.* 41 (2014), 342–350.

[18] Bing Liu. 2012. *Sentiment analysis and opinion mining*. Vol. 5. Morgan & Claypool Publishers. 1–167 pages.

[19] Mahalia Miller, Conal Sathi, Daniel Wiesenthal, Jure Leskovec, and Christopher Potts. 2011. Sentiment Flow Through Hyperlink Networks.. In *ICWSM*.

[20] Mishne. 2006. MoodViews: Tools for Blog Mood Analysis. *AAAI Symposium: Comp. Approaches to Analyzing Weblogs* (2006).

[21] Mark Newman. 2010. *Networks: an introduction*. Oxford university press.

[22] Christine Leigh Myers Nickel. 2008. *Random Dot Product Graphs A Model For Social Networks*. Ph.D. Dissertation.

[23] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in IR* 2, 1-2 (2008), 1–135.

[24] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*. 79–86.

[25] Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. 2015. Sarcasm detection on twitter: A behavioral modeling approach. In *WSDM*. 97–106.

[26] Paul Rozin and Edward B Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and social psychology review* 5, 4 (2001), 296–320.

[27] Jaime E Sidani, Ariel Shensa, Ana Radovic, Elizabeth Miller, Jason B Colditz, Beth L Hoffman, Leila M Giles, Brian A Primack, et al. 2016. Association between social media use and depression among US young adults. *Depression and anxiety* 33, 4 (2016), 323–331.

[28] Stefan Stieglitz and Linh Dang-Xuan. 2012. Impact and Diffusion of Sentiment in Public Communication on Facebook.. In *ECIS*. 98.

[29] Stefan Stieglitz and Linh Dang-Xuan. 2013. Emotions and information diffusion in social media - sentiment of microblogs and sharing behavior. *MIS* 29, 4 (2013), 217–248.

[30] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. 2011. User-level sentiment analysis incorporating social networks. In *KDD*. 1397–1405.

[31] Gerben A Van Kleef. 2009. How emotions regulate social life the emotions as social information (EASI) model. *Current directions in psychological science* 18, 3 (2009), 184–188.

[32] Xiao Zhang, Travis Martin, and Mark EJ Newman. 2015. Identification of core-periphery structure in networks. *Physical Review E* 91, 3 (2015), 032803.