Sickness and Health: Homophily in Online Health Forums

Brant Chee
University of Illinois at Urbana-Champaign
Graduate School of Library and Information Science
chee@uiuc.edu

ABSTRACT
This work explores the link between health and social relations by creating an automated metric of similarity of positive or negative affect (sentiment) between peers in online health forums. We analyze textual communication between peers and demonstrate that those who communicate often have similar average sentiment scores. Sentiment is the author's immediate affective state, their positive or negative orientation. We hypothesize that average sentiment over time indicates overall happiness or sadness as similar analysis has been utilized to identify depression and depression at risk college students [19]. These results follow the analysis of Framingham study data demonstrating that happy people tend to associate with one another and that happiness spreads within social networks [4].

Categories and Subject Descriptors
J.4 [Social and Behavior Sciences]: Sociology

General Terms
Measurement

Keywords
Sentiment Analysis, Social Networks, Tie Strength Indicator, Homophily

1. INTRODUCTION
Happiness and depression are important factors for chronic illness patients. The Centers for Disease Control (CDC) currently uses a 14 item measure based upon patient response to simple questions including topics on stress, depression and problems with emotions [13]. The CDC believe that health-related quality of life (HRQOL) is important in the measurement of effects of chronic illness on patient's lives. HRQOL is important in tracking patient's perceived physical and mental health over time and tracking the effects of multiple diseases and disabilities within patient populations. We suggest that with further work this average sentiment metric could augment existing HRQOL measurement tools.

Homophily is the principle that similar people – in many regards to socio-demographic, behavioral, and intrapersonal characteristics including race and ethnicity, interact more than those who are dissimilar [14]. The goal of this work is to derive a quantifiable metric for automatically measuring the similarity of people with respect to their attitudes, abilities, beliefs and aspirations. Specifically, we are looking at features pertaining to value homophily. This includes the wide variety of internal states presumed to shape our orientation toward future behavior. Examples of values might include higher education attainment, social characteristics that can be correlated with political similarity, etc. People tend to assume that their friends are like them [14]. People's social network, their relationships and interactions with other people, is formed by whom they choose to interact with.

Within the social networking paradigm people represent the nodes in a social network diagram or sociogram and their interactions represent the edges connecting the nodes. Patterns of interaction are used to demonstrate an effect, such as, people whose friends become obese tend to also become obese [2]. In many cases the network and effect are hard to construct due to the difficulty in ascertaining necessary information about social interaction and the hypothesized social effect. Special datasets such as the Framingham study are often used. This dataset includes manually collected data over a period of 20 years, preventing the analysis from being performed on many people [2,3,4]. We would like to perform similar sociological analysis on new segments of the population automatically.

In this work we define a similarity metric for a person's cognitive model and overall sentiment utilizing the words they use when constructing messages to other people. This sentiment metric is used as a measure of a type of value homophily within the Yahoo Health forums. We demonstrate that people's affective or emotional state, specifically their positive or negative orientation, is likely to be similar to others' they choose to associate with.

For our analysis, we use messages within online Yahoo Health forums, constructing a social network through people's message exchanges. Email is used as a proxy for measuring relationship strength, which others have done previously [20]. We demonstrate that people's affective emotional state is similar to others they communicate frequently with.

2. RELATED WORK
The words people use in conversation correlate to physical and mental health [19]. Research in content analysis introduced in the 1960's detected a person's affective or immediate feeling state based solely on variations in the content of verbal communications [8]. The same language processing technique was used in the late 1970's to differentiate between people with schizophrenia and those without [18]. Following related work focused on written text, finding variations in language usage between depressed, depression-vulnerable and non-depressed students [19]. Much work has been done on the automatic

detection and analysis of sentiment [15]. One way to think of sentiment is an author's attitude; the positive or negative polarity apparent through the author's writing.

Similar work by Fowler and Christakis addressed the spread of happiness within a large social network. The same dataset has been used to demonstrate smoking cessation through one's social network [3]. Our goal is to automatically compute results for large populations of users and datasets. Little work has been done on the construction of social networks within the context of online health forums. Furthermore, this work specifically addresses the idea of automatically constructing and analyzing these networks utilizing the concept of homophily.

Adamic and Adar look at websites, specifically MIT and Stanford student home pages to programmatically create a social network, and from it predict friendships [1]. The social network is automatically constructed from web pages (not messages). Verification of friendship is accomplished manually. While homophily is not addressed, the idea that, the "more similar a person are, the more likely they are to be a friend" is. The measure of friendship is determined by text, links, and mailing lists.

Work by Golder et al used Facebook messages to find temporal rhythms consistent across university campuses and seasons [7]. They demonstrate that students at the same university have similar messaging habits. No content analysis was performed on the messages and the resulting data was not used to construct a social network.

3. EXPERIMENT

3.1 People and Data

The social forums we explored consist of 27,290 public Yahoo Health groups. 12,519,807 messages exist within these groups. These groups range from illness based support groups focusing on Multiple Sclerosis to groups focusing on herbal home remedies. For this study, we looked at the 10 largest message groups by file size from this Yahoo corpus.

The messages in these forums consist of informal, often emotional text dealing with feelings of hopelessness, depression or bereavement, for example, "My doctor told me that it works for both depression and as an antianxiety drug… I was in such adepressed stat that I had to go for counselling. [taken verbatim]" Recent studies have shown that the expression of emotional experiences either verbally or in a written context leads to improved physical and psychological health [16]. These texts can also provide emotional insights about the author's mental state at that point in time [19].

Words are not the only elements of analysis that provide necessary emotional insights. People augment computer mediated communication to mimic face to face interactions through the use of nonverbal elements [21]. Emoticons are nonverbal expressions and are often textual representations of writer's facial expressions [5]. For example :) or :-) would correspond to a smile indicating happiness. These cues indicate to the reader the author's intensions which can be hard to determine in informal written communication.

3.2 Experimental Design

We seek to quantitatively determine if a person's affective or emotional state, specifically their positive or negative orientation, is likely to be similar to others' they choose to associate with.

A social network from the messages within the Yahoo group was constructed for this experiment. From this network we extract pairs of people who have numerous interactions and calculate the difference in sentiment scores between the nodes. The same number of pairs is selected at random from the network and the difference in sentiment scores between these pairs of random nodes is calculated. Statistical tests on the two groups (random pairs and interacting pairs) of differential sentiment scores are performed to determine if there is a significant difference between them.

The Yahoo groups were first parsed such that non-text including images or attachments and replies were removed. From the messages, we create a social network. Within the network, nodes are email addresses, which serve as a unique identifier for a person. An edge is formed between nodes *A* and *B* if *B* responds to a message posted by *A* or vice versa. We arbitrarily decided that an interaction of ten messages between two people implies a strong tie. Figure 1 demonstrates an example of a social network. The strong ties are solid black and consist of ten or more interactions. The dotted lines represent weak ties.

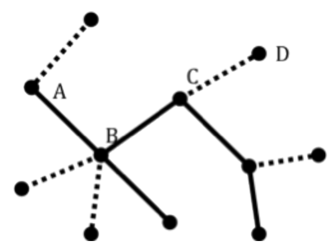

Figure 1: Example social network, strongly connected nodes are represented by solid black lines e.g. nodes A and B. Weakly connected ones e.g. nodes C and D are represented by dotted lines.

The correlation between strong ties, in this case numerous communications between *A* and *B* and sentiment is interesting. If they co-vary together, this suggests homophily. While replies help us understand the context of a message, a message's emotional circumstance should not be based on what other people write, only on the authors' text. Therefore, we remove replies from the content of messages.

Portions of the lexicon in the Linguistic Inquiry and Word Count (LIWC) were used [17]. Specifically, the words corresponding to the following categories: positive emotion, negative emotion, anxiety, anger and sadness. We have augmented the LIWC lexicon to include a wide range of emoticons such as :) :(:P ^__^ LOL ROFL. The resulting messages were matched against the LIWC lexicon categories and emoticons. Counts containing number of positive emotion words, and negative ones, and total number of words were recorded.

A score for each message was calculated. In the score calculation shown below, the set of positive terms denoted as *p* consist of the positive emotion words including positive emoticons. The set of negative terms denoted *n* consists of the negative emotion, anxiety, anger, and sadness terms from LIWC as well as negative emoticons such as :(.

Each message m consists of the set of t white space delimited tokens such that $m = \{t\}$.

$$score = \frac{\sum_i (\alpha \mid i \in p) - (\alpha \mid i \in n)}{|m|}$$

For each token i in a message m if i is in the set of positive LIWC lexicon α is added to the score and if i is in the negative LIWC lexicon α is deducted from the aggregate score. The aggregate score is then divided by the number of tokens. Here α is a constant in this case $\alpha = 1$; thus a message consisting primarily of positive lexicon will have a score > 0, whereas a purely informational message will have a score = 0, and a message that is predominantly negative will have a score < 0. For each person within the network, an aggregate score is calculated by taking the average score of each of the messages they have written.

The absolute value of the difference in average sentiment between people with strong ties was calculated. The difference in average sentiment provides a distance metric between the two people. We call this metric sentiment distance. We hypothesize that pairs of nodes for example (*A,B*) and (*B,C*) in Figure 1 will have a smaller sentiment distance than randomly chosen nodes. Intuitively, this means that because people, (nodes) *A* and *B* chose to communicate more, they are more likely to have similar positive or negative alignment.

However, in calculating the sentiment for each person who has met the communication threshold or "strong tie", we do not use the messages where the people communicated. We did this so that the language and topic of the threaded messages did not bias the average score.

For every pair of strongly connected nodes, for example (*A*,*B*) and (*B*,*C*) in Figure 1, we pick two other nodes at random which are not strongly connected, for example (*C*,*D*) or (*A*,*D*). We calculate the sentiment distance between the pairs of nodes. A sentiment distance can be calculated between every pair of nodes within the network. The nodes do not need to be directly connected in the network in order to calculate the sentiment distance since it is based on the messages of the people represented by those nodes.

This process of choosing pairs of nodes and calculating the sentiment distance creates two distributions named strong and weak. The strong distribution is composed of the pair wise sentiment distances between people who have had more than ten communications. The weak are the same number of pairs of randomly selected nodes. [11] show that for a student's t-test sample populations should be approximately equal. Considering all nodes n, in a graph, and s is the set of strongly connected pairs, the number of random pairs, $nr = (n\text{-}1)! - |s|$ where $(n\text{-}1)! >> |s|$ in the general case, so sampling was used.

4. RESULTS

The mean message size for each group is 160,984 messages. The table below lists the statistics for the different groups. The number of pairs of nodes with numerous interactions (10 or more) is shown in the column second from the left. The average distance between those nodes with numerous interactions is in the Strong column followed by the average distance between random pairs of nodes in the Weak column. The p-value is the result of a T-test comparing the distribution of the Strong column values to the Weak column ones. The last column shows the ratio of the average difference between the two populations to demonstrate the quantifiable difference between the two. For each group the mean difference between the sentiment of strong pairs and weak ones appear to be statistically significant with p-values much lower than .05. The people who have numerous interactions are much closer in average sentiment than pairs chosen at random, thus indicating their mental model and happiness levels are similar.

The average ratio of the mean differences between the population of random people versus those with numerous interactions is 2.152, not only is the average difference between nodes with numerous interactions significant, it is less than half of that between those with few interactions.

Of particular interest is that the network was automatically generated from people's behavior within groups. The metrics we use to define similarity are automatically derived and demonstrate that such automatic metrics are still able to detect these similarities with great reliability.

Table 1: Table displaying the means and p-values for T-tests comparing strongly connected people and weakly connected ones within ten Yahoo! Health forums.

Group	Pairs	Strong	Weak	p-value	Weak/ Strong
1	888	0.0082	0.0144	2.20E-16	1.756
2	505	0.0104	0.0324	2.20E-16	3.115
3	463	0.01	0.0243	2.20E-16	2.430
4	398	0.0171	0.0281	1.37E-11	1.643
5	380	0.0058	0.0129	2.20E-16	2.224
6	341	0.0173	0.0207	3.69E-03	1.197
7	306	0.0096	0.0169	2.28E-09	1.760
8	262	0.0112	0.031	2.20E-16	2.768
9	237	0.0096	0.0199	2.20E-16	2.073
10	233	0.0074	0.0189	2.96E-12	2.554

4.1 Limitations

These finding validate the idea that people who interact frequently have similar average sentiment and therefore similar mental models, however these results do not indicate causality. Our current work only explored the differences between an arbitrarily defined strong tie and a weak one. It is not known if the current metric is a continuous one, which would prove more useful as Gilbert's work suggests [6]. We do not currently have any information on how sentiment between people is affected by the strengthening or dissolution of ties.

4.2 Implications

4.2.1 Practical Implications

People with strong ties have a small difference in their average sentiment scores. A potential implication is use of average sentiment difference as a feature in the calculation of tie strength. Gilbert and Karahalios used words in inbox messages and Facebook wall posts to quantify tie strength, however they did not look at average sentiment of posts and message [6].

Average sentiment analysis is more computationally expensive than finding people with numerous communications. Utilizing our findings, it is possible to create groups of people utilizing this cheaper distance metric and verify their average emotional distance to create groups of people with similar value homophily. Implications of similar value and emotional/psychological based groups include targeted advertising, identification of depression at-risk populations. Previous work by Rude et al. shows that automatic detection of people with depression is possible [19].

4.2.2 Theoretical Implications

Hancock et al. demonstrate that emotional contagion, the mood of one person can change the mood of others interacting with that person in text-based communication [10]. Similar results were demonstrated for groups of people in a Social Network and shifts in happiness of people within that group [4].

People who are optimistic tend to be healthier and live longer than those who are pessimistic and cynical. A long term study started in 1921, of 1,500 pre-adolescent boys demonstrated that expecting the worst was linked to a 25-percent higher risk of dying before age 65 [12]. Over 1300 people in a 10 year Harvard study showed cardio-protective effects of optimism; the risk of coronary death or disease, Angina, or non-fatal Myocardial infarction was reduced by half [9].

This work may contribute to the development of a quality of life metric utilizing average positive or negative orientation. Further, since one's orientation changes depending on whom one interacts with, we conjecture it is possible to change a person's orientation by changing who they interact with. Possible implementations of this include re-ordering of people's information, messages in forums to rank negative people's posts higher, or to suggest friends who are positive for negative oriented people of whom have weak ties connecting them.

5. ACKNOWLEDGMENTS

This work was an independent study oversaw by Caroline Haythornthwaite and was partially supported by a fellowship from the Center for Integration of Medicine and Innovated Technology. We thank reviewers for helpful comments.

6. REFERENCES

[1] Adamic, L.A., and Adar, E. 2001. Friends and neighbors on the web. Social Networks, 35 (2001), 211-230.

[2] Christakis N.A. and Fowler J.H. 2007. The spread of obesity in a large social network over 32 years. N Engl J Med, 357 (2007), 370-379.

[3] Christakis N.A. and Fowler J.H. 2008. The collective dynamics of smoking in a large social network. N Engl J Med, 358 (2008), 2249-2258.

[4] Fowler JH, Christakis NA. 2008. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study. BMJ, (2008) 337a2338.

[5] Gajadhar J. and Green J. 2005. An analysis of nonverbal communication in an online chat group. EDUCAUSE Quarterly 24, 4(2005), 63-64.

[6] Gilbert, E. and Karahalios, K. Predicting tie strength with social media. 2008. In Proc. CHI 2009, ACM Press (2009), 211-220.

[7] Golder, S., Wilkinson, D., and Huberman, B. 2007. Rhythms of social interaction: Messaging within a massive online network. In Proc. CT2007, Springer (2007), 41-66.

[8] Gottschalk L.A. and Gleser G.C. 1969. The Measurement of Psychological States through the Content Analysis of Verbal Behavior. University of California Press, Berkeley, CA.

[9] Kubzansky, L.D., Sparrow, D., Vokonas, P., Kawachi, I. 2002. Is the glass half empty or half full? A prospective study of optimism and coronary heart disease in the normative aging study. Psychosom Med, 63 (2002), 910-916.

[10] Hancock, J.T., Gee, K., Cicaccio, K. and Lin, J.M.H. 2008. I'm sad you're sad: Emotional contagion in CMC. In Proc. CHI 2008, ACM Press (2008), 295-298.

[11] Markowski, C. A. and Markowski, E. P. 1990. Conditions for the effectiveness of a preliminary test of variance. The American Statistician, 44, 4 (1990), 322–326.

[12] Maruta, T., Colligan, R.C., Malinchoc, M., Offord, K.P. 2000. Optimists vs pessimists: Survival rate among medical patients over a 30-year period. Mayo Clin Proc.,75, 2 (2000), 140-143.

[13] Moriarty, D. B., Zack, M. M., and Kobau, R. 2003. The centers for disease control and prevention's healthy days measures - population tracking of percieved physical and mental health over time. Health and Quality of Life Outcomes (2003), 1-37. DOI=10.1186/1477-7525-1-37

[14] McPherson M., Smith-Lovin L., and Cook J.M. 2001. Birds of a feather: Homophily in social networks. Annu. Rev. Sociol., 27 (2001), 415-444.

[15] Pang B. and Lee L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2, 1–2 (2008), 1-135.

[16] Pennebaker J.W. and Campbell R.S. 2000. The effects of writing about traumatic experience. Clinical Quarterly, 9 (2000), 17-21.

[17] Pennebaker J.W., Francis M.E., and Booth R.J. 2007. Linguistic inquiry and word count: LIWC 2007. Lawrence Erlbaum Assoc, New Jersey, 2007.

[18] Rosenberg S.D. and Tucker G.J. 1979. Verbal behavior and Schizophrenia. Arch Gen Psychiatry, 36 (1979), 1331-1337.

[19] Rude S.S., Gortner E.-M., and Pennebaker J.W. 2004. Language use of depressed and depression-vulnerable college students. Cognition and Emotion, 18, 8 (2004), 1121-1133.

[20] Tyler, J., Wilkinson, D., and Huberman, B. 2003. Email as spectroscopy: Automated discovery of community structure within organizations. In Communities and Technologies, M. Huysman, E. Wenger, and V. Wulf, Eds. Kluwer, Deventer, the Netherlands, 81-96, 2003.

[21] Walther J.B. and D'Addario K.P. 2001. The impacts of emoticons on message interpretation in computer-mediated communication. Social Science Computer Review, 19, 3 (2001),323–345.