The AAAI 2017 Spring Symposium on
Wellbeing AI: From Machine Learning to Subjectivity Oriented Computing
Technical Report SS-17-08

Loneliness in a Connected World: Analyzing Online Activity and Expressions on Real Life Relationships of Lonely Users

Camille Ruiz, Kaoru Ito, Shoko Wakamiya, Eiji Aramaki
Nara Institute of Science and Technology
8916-5 Takayama-cho
Ikoma, Nara 630-0192, JAPAN
camille.ruiz.co3@is.naist.jp

Abstract

Although loneliness is a very familiar emotion, little is known about it. An aspect to explore is the prevalence of loneliness in the connected world that social media sites like Twitter provide. In light of this, this study investigates the Twitter data of users that have expressed loneliness to understand the phenomenon. Since our primary material are tweets, we developed various indices that can measure social activities reflected in online relationships and real life relationship solely through online Twitter data. Through these indices, the relations between social activity and loneliness were investigated. The results show that high lonely users seem to have low online activity, high positive expressions on real life relationships, and narrow ingroups.

Introduction

Most people have felt lonely at least once in their lives. Through modern technology like the internet, we could imagine that people would feel less lonely in general since social media sites like Facebook and Twitter can keep us connected with others. After all, the technology can allow individuals to express themselves and have their loved ones respond no matter how far apart they may be from each other. Nonetheless, people still express feelings of loneliness on social media. Given the connectivity available to many people today, expressions of being lonely in sites like Twitter are interesting to explore because it may give insight on the phenomenon of loneliness.

To be more precise, loneliness in this paper is described as a distressing emotion that springs from poor or inadequate social relations (Perlman and Peplau 1981). While depression is the usual topic in the mental disease field, loneliness is not well studied in comparison. Nonetheless, studying loneliness is important especially since loneliness is considered as a strong factor of depression (Stravynski and Boyer 2001). In fact, loneliness has been deemed by studies a risk factor for mortality, some even claiming that loneliness can be pandemic by 2030 if nothing is done to attenuate it (Holt-Lunstad et al. 2015).

Previous studies suggest that social media use can help attenuate loneliness. For instance, users who participate in both Facebook and Twitter are less lonely compared to those who only use Facebook according to their UCLA Loneliness scores (Petrocchi et al. 2015). Another study found that interacting and browsing in Instagram—instead of simply broadcasting—is related in easing loneliness (Yang 2016). While agreeing that social media help lower loneliness, some studies argue that image-based sites such as Instagram and Snapchat are better in this task than text-based like Twitter(Pittman and Reich 2016). Nonetheless, all these studies agree that social media helps in attenuating loneliness.

The question remains: why is loneliness prevalent today even if social media should already aid in lowering it? The intuitive answer is that loneliness likely springs from events outside online interactions. If used to understand the phenomenon of loneliness, social activity should then be explored not just in the online sphere but also in the sphere of real life. Moreover, social activity in terms of relationships are relevant in such a study since loneliness is rooted in the deficiency of social relations.

In light of this, this paper investigates the relation between loneliness and social activity. Social activity can be reflected in two types of relationships: online relationships and real life relationships. These were measured through several indices (see Figure 1). Online relationships have two indices: Passive Online Relationship Index and Active Online Relationship Index. On the other hand, real life relationships are estimated through three measurements: Family-Sentiment Index, Friend-Sentiment Index, and Family-Friend Mention Balancing Index.

Online Relationships Online relationships here pertains to the connection a person has through communicating and relating to others on the internet. For instance, participation in social media sites like Twitter allows a person to have various online relationships. Given this example, we are then able to determine a Twitter users online relationships with his or her tweets, followers, and followees.

In this study, we divide online relationships under two categories: **passive online relationships** and **active online relationships**. Passive online relationships center on the ability of users to observe each other online. On Twitter, passive online relationships concerns the followers, followees, and mutuals of a user. On the other hand, active online relationships involve mentions by the user and mentions at the user since this category hinges on the users' activity of express-

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

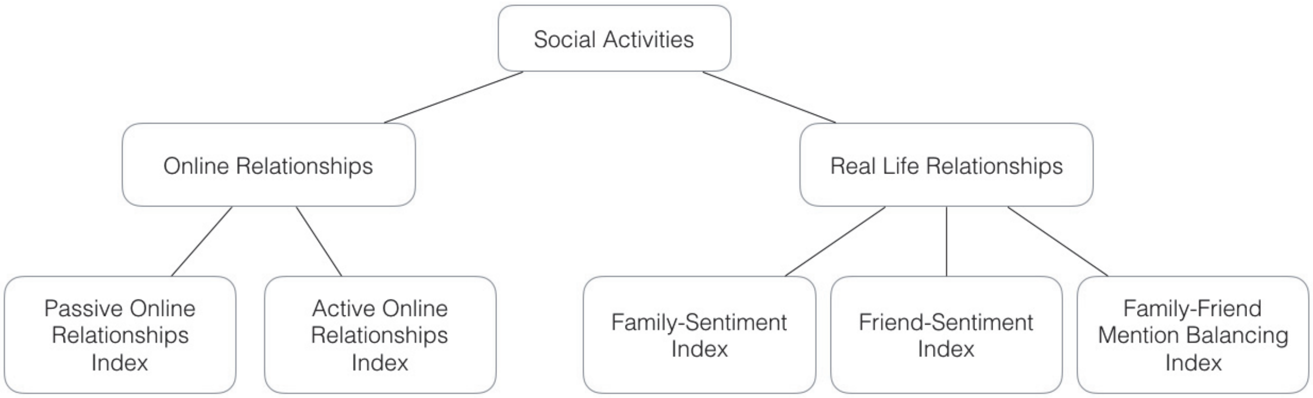

Figure 1: *Index Structure*: The designed indices that were used for online relationships and real life relationships.

ing or communicating to others.

Real Life Relationships Unlike online relationships, real life relationships cannot be easily measured or estimated. Real life relationships in this paper are defined as meaningful connections a person would have that often develop beyond online interactions. A clear example of a real life relationship is a relationship with a **family member**–like a mother or sibling. Another real life relationship is a relationship a person who would have with someone he or she considers a **friend**.

It is intuitive to suppose that having a positive or negative relationship with someone affects whether or not a person would feel lonelier or not. For instance, having an argument with one's friend or parent will tend to make the person feel more isolated; therefore, feel lonelier. However, the state of a person's real life relationship is difficult, if not impossible, to measure based on what we see on the person's online profile. What we can measure is the *expressed sentiment* about such relationships online. In light of this, we estimate real life relationships through what we call **Family-Sentiment Index** and **Friend-Sentiment Index**.

We may also be able to estimate the size of the real life ingroup of a user with their online data. An ingroup is "a group with which one feels a sense of solidarity or community of interest"(ing 2017). The idea is that the balance of tweets mentioning family and friends can capture whether a user has a narrow or wide ingroup. We coined this estimation method as **Family-Friend Mention Balancing Index**.

Objective The objective of this paper is to investigate the phenomenon of loneliness in social media through the data of Twitter users. Focusing on social activity in term of relationships, we explore possible measurements of both online and real life relationships and find their relations to loneliness. In other words, this research faces the following challenges:

- From only online data, we aim to obtain various estimates of social activities including real life related indices.
- We, then, investigate the relation between loneliness and these social activities.

Materials

Twitter Crawling

We crawled tweets of users to be closely observed and tweets for training our sentiment classifier. From 7 December 2016 to 6 January 2017, we attempted to gather 3 Twitter usernames every 15 minutes that appeared from a query of the keyword "lonely." We then collected the last 3,200 tweets of each user. We also gathered public tweets that mention the users by querying "@username."

From 17 January to 23 January 2017, we also crawled tweets as training data for sentiment analysis. For this, we used ':)' and ':(' as query for positive and negative tweets, respectively.

Filtering While gathering information, we ignored users that had more than 500 followers so that the data will have regular users as opposed to famous people. We also ignored users that had the string "bot" in their user names as a simple way to prevent obvious bots. Moreover, users with less than 100 tweets were taken out of the data set to lessen the possibility of outliers.

With regard to the training data for sentiment analysis, tweets that did not contain certain emoticons were ignored. Specifically, we ignored tweets without any of the strings in the following array: [‘ :D ’, ‘ D: ’, ‘ :P ’, ‘ :p ’, ‘ :-) ’, ‘ :-(’, ‘ :) ’, ‘ :)) ’, ‘ :))) ’, ‘ :(’, ‘ :((’, ‘ :(((’] After collecting and filtering out the tweets, we removed the emoticons based on the list of emoticons that we made. We also masked the usernames and links contained in the tweets by replacing them with <@USER>and <LINK >.

Corpus Statistics

After the filtering, 7,787 users with an average of 1,636 tweets per user were left. This amounts to 12,735,749 user tweets and 70,208 public tweets mentioning the users. With regard to sentiment analysis, we were able to gather 34,938 tweets to serve as training data.

Data Structure

After the crawling of our target users' tweets, we created a data file that summarizes information about each user and their tweets. This includes information such as:

- total number of *tweets* collected (maximum 3200)
- number of *followers*, *followees*, and *mutuals*
- *lonely tweets*: total number of tweets containing either "lonely", "alone", "lonesome", and "loneliness"
- *mentions by user*: number of tweets that contain "@string."
- *mentions at user*: number of public tweets that mentions the Twitter usernames, gathered through querying "@username" for every user.
- *friends count*: number of tweets containing at least one of the keywords from the family category in LIWC2015. The LIWC contains 94 keywords, such as "bestie", "mate" (James W. Pennebaker and Blackburn 2015).
- *family count*: number of tweets containing at least one of the keywords from the friends category in LIWC2015. The LIWC contains 118 keywords, such as "mum", "dad" (James W. Pennebaker and Blackburn 2015).
- *Family Friend Keyword Frequencies*: an integer array in descending order containing the frequencies of the family and friend keywords that the users tweeted at least once

  Our data is available on the web [1].

User Grouping

With this data file, we analyzed different aspects of the users with respect to loneliness. To do this, the users were ordered according to the ratio between lonely tweets and total tweets. This ordered set of users were then divided into three sets. The users in the first, second, and third set are respectively labeled as **low lonely** users, **neutral** users, and **high lonely** users. In our study, only two categories—low lonely and high lonely—are compared and investigated.

Method

This study investigates various indices, ranging from online relationships (2 indices) to real life relationships (3 indices). Among them, the Family-Friend Mention Balancing Index is newly introduced. The *t*-test were performed on all comparisons discussed in the following sections.

Online Relationships

Online relationships of users can be measured under two categories: passive online relationships and active online relationships.

[1] http://sociocom.jp/data/en/

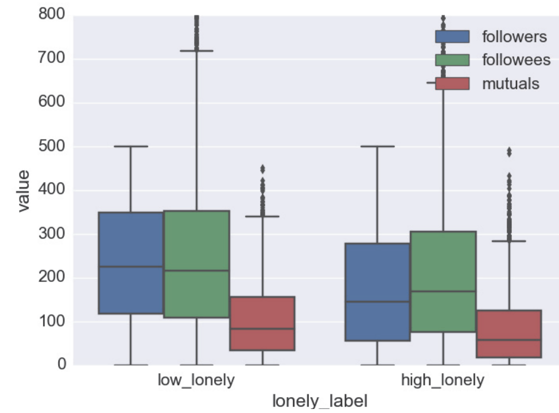

Figure 2: *Relation of Passive Online Relationship and loneliness*: The average number of followers (blue), followees (green), and mutuals (red) per user group. In all three indices, low lonely users have higher values compared to high lonely users.

Passive Online Relationship Index This index concerns how a user is related to other users passively through the ability of observing the online activity of the other user and vice versa. On Twitter, passive online relationships are measured through the number of *followers, followees, and mutuals* of a user. We regard these numbers as representing passive online relationships as is.

The act of following a user means that the tweets of the user you are following will be included in one's timeline. A user that follows another user is called a *follower*, while the followed user is coined as *followee* (see Figure 3). If a user you are following also follows you, that user is called a *mutual*.

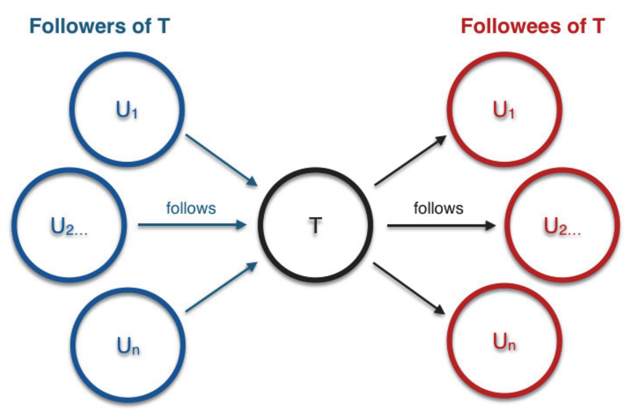

Figure 3: The relationship between users as followers and followees

To measure and compare the passive online relationships of low lonely users and high lonely users, boxplots of the users' number of followers, followees, and mutuals were drawn in a single graph (refer to Figure 2).

Active Online Relationship Index This index concerns how a user is related to other users actively through directly communicating a message specifically to another user. The material used to measure active online relationships are mentions.

We consider two forms of mentions: *mentions by the user and mentions at the user*. Given a particular user T being studied, mentions by a user T are user T's tweets directed to another user by using the @otheruser. On the other hand, mentions at the user T are another user's tweets directed at user T. In other words, these are tweets from other users that include @userT (refer to Figure 4).

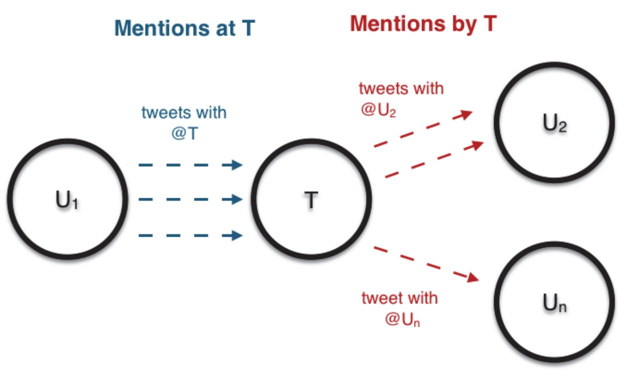

Figure 4: The way mentions can be made by and at the user

To measure and compare the active online relationships of low lonely users and high lonely users, boxplots of mentions by the user ratio (over total tweets) and the number of mentions at the user were drawn (see Figure 5 and Figure 6).

Real Life Relationships

Real life relationships are estimated through the users' expressions about their real life relationships. This includes expressions about their family and friends. Three types of measurement were explored: (1) Family-Sentiment Index, (2) Friends-Sentiment Index, and (3) Family-Friend Mention Balancing Index.

(1) Family-Sentiment and (2) Friend-Sentiment Indices These indices concern the sentiment of tweets talking about family and friends. Positive family tweets are tweets with positive sentiment that talks about family members while negative family tweets are tweets with negative sentiment. Similarly, positive friend tweets are tweets with positive sentiment that talks about friends while negative friend tweets are tweets with negative sentiment. The Family-Sentiment Index of a user is the ratio of the user's positive family tweets over total tweets while the Friend-Sentiment Index is the user's positive friends tweets over total tweets. We define this using the following formula:

$$\frac{PositiveFamilyTweets}{TotalTweets}$$

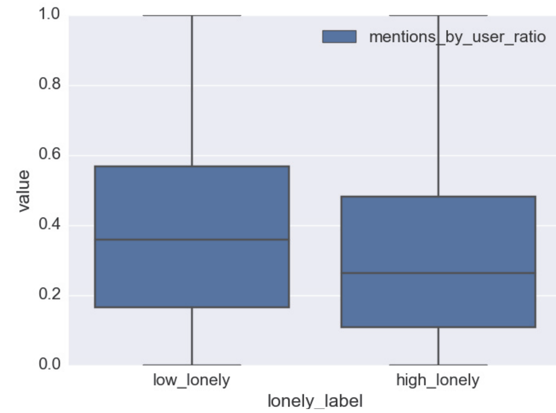

Figure 5: *Active Online Relationships - Mentions By User*: The average ratio of mentions by users over total tweets per user group

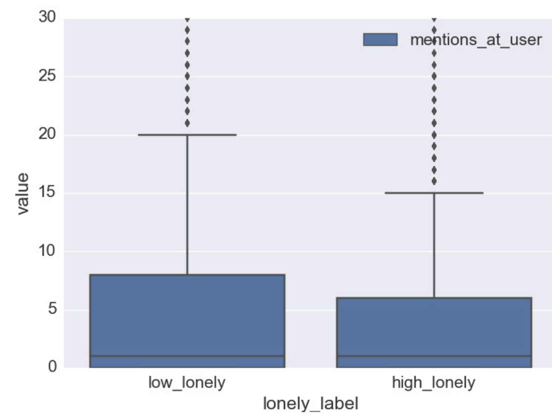

Figure 6: *Active Online Relationships - Mentions At User*: The average number of public mentions at users per user group

$$\frac{PositiveFriendTweets}{TotalTweets}$$

We classified whether or not a tweet talks about one's family member and friend by using the Language Inquiry Word Count 2015 (LIWC2015) dictionary. We particular used keywords from the Family category and the Friend category in LIWC2015.

We also classified tweets as either having positive or negative sentiment using a Linear Support Vector Machine based Classifier (refer to Figure 7). This classifier will be explained in the next subsection. We computed the ratio of positive friend tweets over total tweets and positive family tweets over total tweets of the users. The boxplots of these ratios were drawn to compare the Family-Sentiment and Friend-Sentiment of low lonely users and high lonely users (refer to Figure 8).

Sentiment Classifier for (1) and (2) We utilized emoticons to create a semi-supervised classifier for calculat-

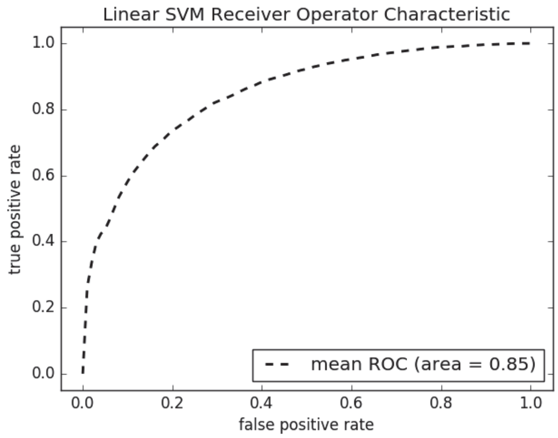

Figure 7: ROC of Linear SVM; 5 folds, 34,938 tweets

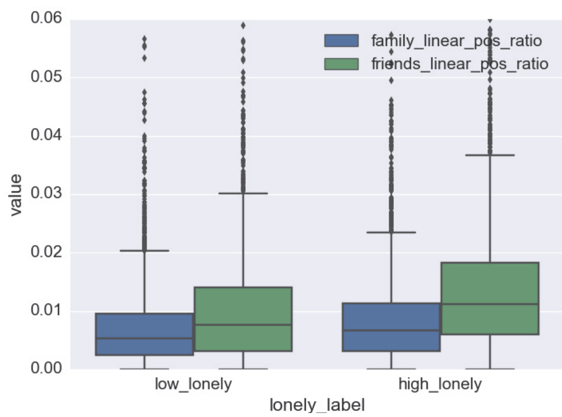

Figure 8: *Family-Sentiment Index and Friend-Sentiment Index*

ing Family-Sentiment Index and Friend-Sentiment Index. A similar method has been used before where emoticons are used in labeling data (Vosoughi, Zhou, and Roy 2015). In this study, we do this by using the ':)' and ':(' twitter query then labeled tweets coming from the query as 1 and -1, respectively. In other words, tweets from the ':)' query are considered as positive tweets, while tweets coming from the ':(' query are negative tweets. As explained earlier, the emoticons were then filtered out of the tweets and information like user names and links were masked out. The remaining information were then used to train a Linear SVM.

Family-Friend Mention Balancing Index (w_{MBI}) This index attempts to capture the balance of tweets mentioning family and friends. The underlying idea of this index is that the ingroup size of the user can be estimated through investigating how frequent a user mentions her peers. The aim is to measure whether a user has a **narrow ingroup** by talking about a small set of family and friends very often or if he has a **wide ingroup** by talking about his family and friends generally with the same frequency.

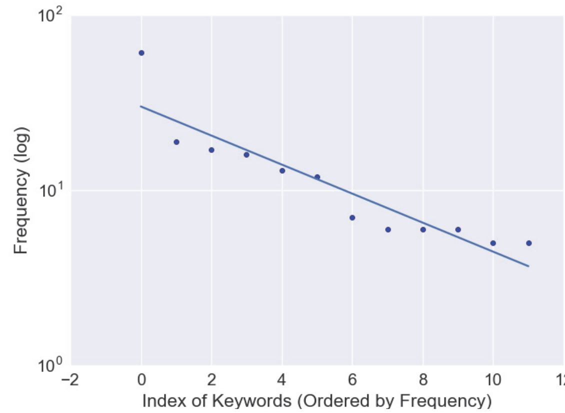

Figure 9: *Family-Friend Mention Balancing Index*: An example of log-linear regression applied to a keyword frequency array of a user. (w_{MBI}: 0.191)

For example, a user that very often tweets about their dad but rarely tweets about other friends or family members would have a narrow ingroup. On the other hand, another user would have a wider ingroup if she talks about her friend, mother, sibling, and father through roughly the same amount of tweets.

For each user, we generated an array containing the frequencies of keywords that the user used in his tweets at least once. This array was then sorted in descending order. With this data, *log-linear regression was used to determine the exponential curve that best describes the relationship between keywords and frequencies* (see Figure 9 for an example). Given the following formula:

$$\ln y = -w_{MBI}x + b$$

$$w_{MBI} = -\frac{\ln y - b}{x}$$

where y refers to the frequency of the keyword, x is the keyword and w_{MBI} is the weight for the Mention Balancing Index.

The parameter w_{MBI} obtained from the regression signifies the weight determining the drop from the first few data points. The higher the value of w_{MBI}, the steeper the line.

In other words, frequency arrays that produce a higher weight would reflect narrower ingroups while frequency arrays with lower weights would reflect wider ingroups. To measure and compare the FFMBI of low lonely users and high lonely users, boxplots of the weights were drawn (see Figure 10).

Results

The results are summarized at Figure 11. The comparisons under all indices—ranging from the active online relationship index to the Family-Friend Mention Balancing index—have a p-value lower than 0.05.

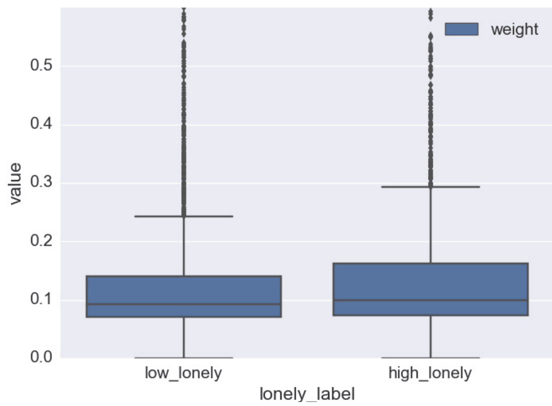

Figure 10: *Relation between Family-Friend Mention Balancing Index and Loneliness*: High lonely users have higher values compared to low lonely users.

Online Relationships

Passive Online Relationship Index As shown in Figure 2, low lonely users have a significantly higher Passive Online Relationships Index compared to high lonely users. The table at Figure 11 details that the average followers, followees, and mutuals of high lonely users (respectively: 177.5, 245.2, 82.4) are lower than the averages for low lonely users (respectively: 235.2, 279.4, and 103.4).

Active Online Relationship Index Figure 6 and Figure 5 depict that high lonely users significantly have a higher Active Online Relationship Index compared to low lonely users. Low lonely users have an average of 0.4 mention by user ratio and 9.2 mentions at user, while high lonely users have an average of 0.3 mention by user ratio and 7.8 mentions at user.

Real Life Relationships

Family-Sentiment Index For the Family-Sentiment Index, the values are significantly greater for high lonely users than for low lonely users, with low lonely users averaging at 0.007 and high lonely users with 0.009 (see Figure 8).

Friend-Sentiment Index Similar to the Family-Sentiment Index, the values for Friend-Sentiment Index are higher for high lonely users (ave: 0.014) compared to the values for low lonely users (ave: 0.11). This was visualized in Figure 8.

Family-Friend Mention Balancing Index The Family-Friend Mention Balancing Index of high lonely users (ave: 0.4) show significantly higher weights compared to the weight of low lonely users (ave: 0.2) (see Figure 10).

Discussion

Among the various explored indices, there are three notable findings in the results. These findings may reflect the relations of loneliness to certain social activities. In particular, high lonely users—compared to low lonely users—have low online activity, high positive expressions on real life relationships, and narrow ingroup estimates.

Loneliness and Low Online Activity The first noteworthy result concerns the online activity of the users seen through their online relationships. Figure 2, Figure 5 and Figure 6 show that high lonely users have fewer online relationships compared to low lonely users. In other words, lonelier users tend to have lower online social activity.

Two possible explanations can be drawn to support the result. The first centers on how low online activity may cause loneliness. In other words, loneliness can spring from how not all people fully utilize the connection that social media provides. In a way, this explanation supports the phenomenon of loneliness in today's connected world. Even if connectivity is available, if people do not harness its potential to improve social relations, then this may increase the likelihood of loneliness.

In contrast to the first explanation, the second possible account for the result makes loneliness the cause of low online activity. Here, perhaps a lonely user has low online activity because he only has a few people to connect with to begin with. Simply put, a user is may be lonely because he lacks significant real life relationships and that low online activity is just a consequence of this deficiency.

In either of the explanations, we can hypothesize that loneliness can be reduced through an increase of social activity. While this is obvious for the first account, more clarification is needed for the second explanation where low online connectivity is an effect rather than a cause. The underlying idea is that perhaps the lonely user can try to find meaningful relationships online. By participating more in online activity, a lonely user may be able to develop friends to help alleviate loneliness.

Loneliness and High Positive Real Life Relationships Sentiment The next striking result concerns the lonely user's sentiment on real life relationships. We expected that low lonely users would have higher ratios of positive family and friend sentiment. This was our expectation because we thought it would be intuitive if expressed positive sentiments on friends and family would hint on the user having good relationships with them. For example, a person who is not lonely may spend a lot of time with friends and family. They would then tweet about these positive experiences frequently. If this was so, users who express more positive sentiments on real life relationships should be less lonely.

However, our study surprisingly produced results opposite to our expectation—*high lonely users have higher ratios of positive sentiment regarding family and friends* compared to low lonely users (refer to Figure 8). In other words, lonelier users express more positive sentiment on real life relationships.

While we find the outcome surprising, we hypothesize that this result may be because the impact of good social events would seem greater for lonely users compared to less lonely people. This is because people likely tweet about events that they find special or rare. For example, a person who is not very lonely may find enjoying the company of one's family or friend normal. Since engaging with peers is

Social Activity	Subcategory	Source	Low Lonely Average	Low Lonely (SD)	High Lonely Average	High Lonely (SD)
Online Relationships	Passive Online Relationships	Followers	***235.2**	(138.07)	177.5	(139.31)
		Followees	***279.4**	(269.68)	245.2	(289.07)
		Mutuals	***103.4**	(85.72)	82.4	(81.40)
	Active Online Relationships	Mentions By User Ratio	***0.4**	(0.27)	0.3	(0.27)
		Mentions At User	***9.2**	(17.73)	7.8	(17.10)
Real Life Relationships	Family	Positive	0.007	(0.01)	***0.009**	(0.02)
		Negative	0.993	(0.01)	0.991	(0.02)
		All	0.033	(0.03)	***0.04**	(0.04)
	Friends	Positive	0.011	(0.02)	***0.014**	(0.02)
		Negative	0.989	(0.02)	0.986	(0.02)
		All	0.03	(0.03)	***0.041**	(0.03)
	Family-Friend Mention Balancing (MBI)	Coefficient	0.2	(0.74)	***0.4**	(1.11)

Figure 11: *Summary of results*: This table contains the results per index. (*p-value<0.05)

a regular occurrence, this will not motivate them to tweet about since it happens often. However, a lonely person may experience the company of others rarely. Hence, the impact of social activities will be greater for lonely users—making them tweet about these events and the family or friend involved in a positive way.

Loneliness and Narrow Ingroup Estimates The third notable result is related to the Family-Friend Mention Balancing Index. We proposed this index as a novel measurement of real life relationships using online data. The focus of this index is to detect if a user mentions only a small specific set of family members or friends, or if they spread out and balance their mentions across a larger group. We think that this is a novel way to estimate the size of the ingroup of a user.

We hypothesize that lonely people have smaller ingroups compared to people who are not lonely. Having a narrower ingroup would mean that the person would have less social relations that can help attenuate loneliness. We then expect that lonely users would have a narrow in group.

The results support this expectation. The Family-Friend Mention Balancing Index show that *high lonely users tend to have narrower ingroup estimates* compared to low lonely users. Simply put, this result may indicate that lonely users have a small or narrow social group to rely on. This is in contrast to less lonely users who may refer or identify with a larger social group. Moreover, this has the potential to explain the surprising outcome of lonelier users having high positive real life relationship sentiment. We put forward the hypothesis that having a narrow ingroup is a major factor in causing loneliness. If this was true, having a high number of expressions about real life relationships but only doing so with a narrow ingroup would still leave a user feeling lonely.

In light of this, we put forward the hypothesis that loneliness can be attenuated by having a wider ingroup. This allows a person to have more people to rely on and relate to.

Future Work This paper revealed several features of loneliness: (1) low online activity, (2) positive expressions on real life relationships, and (3) narrow ingroups. These results also indicate the potential feasibility of online based intervention for easing loneliness. We seek to investigate this in the future under the manner of a more clinical study such as conducting surveys and investigating the Twitter feeds of the survey participants.

Conclusion

To understand the phenomenon of loneliness, this study investigated the Twitter data of lonely users. These users expressed loneliness at least once in their timeline. We then developed various indices that can measure social activities such as online relationships and real life relationship solely through online Twitter data. Through these indices, the possible relationship between social activity and loneliness were explored and discussed. Our study revealed that high lonely users seem to have (1) low online activity, (2) high positive expressions on real life relationships, and (3) narrow ingroups. We believe that this finding would help in the development of online based support for easing loneliness as well as exploring possibilities of online based intervention for lonely people.

Acknowledgments

This study was supported (in part) by JSPS KAKENHI Grant Number JP16H06395 & 16H06399, JST ACT-I.

References

Holt-Lunstad, J.; Smith, T. B.; Baker, M.; Harris, T.; and Stephenson, D. 2015. Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science* 10(2):227–237.

2017. Merriam-webster's dictionary.

James W. Pennebaker, Ryan L. Boyd, K. J., and Blackburn, K. 2015. The development and psychometric properties of liwc2015. Technical report, LIWC.net, Austin, TX. Available: https://liwc.wpengine.com/. Accessed 31-January-2017.

Perlman, D., and Peplau, L. A. 1981. Toward a social psychology of loneliness. *Personal relationships in disorder* 31–56.

Petrocchi, N.; Asnaani, A.; Martinez, A. P.; Nadkarni, A.; and Hofmann, S. G. 2015. Differences between people who use only facebook and those who use facebook plus twitter. *International Journal of HumanComputer Interaction* 31(2):157–165.

Pittman, M., and Reich, B. 2016. Social media and loneliness: Why an instagram picture may be worth more than a thousand twitter words. *Computers in Human Behavior* 62:155–167.

Stravynski, A., and Boyer, R. 2001. Loneliness in relation to suicide ideation and parasuicide: A population-wide study. *Suicide and Life-Threatening Behavior* 31(1):32–40.

Vosoughi, S.; Zhou, H.; and Roy, D. 2015. Enhanced twitter sentiment classification using contextual information. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 16–24.

Yang, C.-C. 2016. Instagram use, loneliness, and social comparison orientation: Interact and browse on social media, but don't compare. *Cyberpsychology, Behavior, and Social Networking* 19(12):703–708.