

# Understanding Social Support Expressed in a COVID-19 Online Forum

Anietie Andy<sup>1</sup>, Brian Chu<sup>1</sup>, Ramie Fathy<sup>1</sup>,  
Barrington Bennet<sup>1</sup>, Daniel Stokes<sup>1</sup>, Sharath Chandra Guntuku<sup>2</sup>

<sup>1</sup>Penn Medicine <sup>2</sup>Computer and Information Science  
University of Pennsylvania

{anietie.andy, Brian.Chu, Ramie.Fathy,  
Barrington.Bennett, Daniel.Stokes}

@penmedicine.upenn.edu

sharathg@cis.upenn.edu

## Abstract

In online forums focused on health and well-being, individuals tend to seek and give the following social support: emotional and informational support. Understanding the expressions of these social supports in an online COVID-19 forum is important for: (a) the forum and its members to provide the right type of support to individuals and (b) determining the long term effects of the COVID-19 pandemic on the well-being of the public, thereby informing interventions. In this work, we build four machine learning models to measure the extent of the following social supports expressed in each post in a COVID-19 online forum: (a) emotional support given (b) emotional support sought (c) informational support given, and (d) informational support sought. Using these models, we aim to: (i) determine if there is a correlation between the different social supports expressed in posts e.g. when members of the forum give emotional support in posts, do they also tend to give or seek informational support in the same post? (ii) determine how these social supports sought and given changes over time in published posts. We find that (i) there is a positive correlation between the informational support given in posts and the emotional support given and emotional support sought, respectively, in these posts and (ii) over time, users tended to seek more emotional support and give less emotional support.

## 1 Introduction

Globally, millions of individuals have contracted COVID-19 and more than 2 million people have died from the pandemic as of January 2021<sup>1</sup>. Individuals are turning to online forums focused on discussions around COVID-19 to seek and give support (Stokes et al., 2020). In online health and well-being forums, individuals tend to seek and

give two forms of social support: emotional and informational support (Wang et al., 2012; Yang et al., 2017); where: (a) emotional support sought seeks understanding, affirmation and encouragement, (b) emotional support given includes providing encouragement, (c) informational support sought seeks advice or information, and (d) informational support given provides advice and information. Below are examples (rephrased) of posts that express these social supports in a COVID-19 related online forum:

- **Emotional Support Given:** "I wish everyone on this forum good health however if any of us tests positive for COVID please keep us informed on how you are doing. We as a community will offer emotional support and also learn how it feels to have the virus"
- **Emotional Support Sought:** "Will we be forced to stay home and not work for some weeks? I don't think I can afford 2 weeks without getting payed."
- **Informational Support Given:** "A 14-day quarantine makes sense. This will ensure that you are not going to become symptomatic and prevent pre-symptomatic spread."
- **Informational Support Sought:** "When a vaccine is found, would countries be required to pay for a license to produce the vaccine?"

In this work, given posts published on an online COVID-19 forum, we build 4 machine learning models to measure the extent of the following social supports: (i) emotional support given, (ii) emotional support sought, (iii) informational support given, and (iv) informational support sought.

In online forums, some posts may express multiple social supports; hence, understanding if there is a correlation between multiple social supports expressed in posts is important for better understanding the social support needs of members of the forum. We conduct an analysis to determine

<sup>1</sup><https://coronavirus.jhu.edu/map.html>

the correlation between different social supports expressed in posts published in the COVID-19 online forum.

These social supports sought or given may change over time. Understanding how these social support needs change over time in a COVID-19 online forum is important for several reasons; first, it can help the forum moderators and members provide adequate and meaningful support to users, and secondly, it can provide insights about the long term effects of the pandemic on the well-being of the public, thereby providing the necessary authorities with sufficient and timely information on the changes in social support needs of the public so that they can make resources available to address these needs. We carry out an analysis to determine the gaps in social support supply and demand in posts published on an online COVID-19 forum over time. We conduct additional analysis to identify different time periods in which there are clear differences in the types of social support sought or given and determine language feature differences in posts published in these different time periods.

## 2 Related Work

A lot of prior work has been done to gain insights and respond to the COVID-19 pandemic.

In order to identify treatments and gain insights about COVID-19, the Allen Institute for AI in partnership with the White House and other collaborators, created a resource to use various artificial intelligence and natural language techniques to access thousands of research papers and articles related to COVID-19 to retrieve pertinent information about published research work about COVID-19 (Wang et al., 2020). Using this resource, several research work has been done: (i) in (Esteva et al., 2020), an automated system was built in which given a query, it retrieves relevant publications related to the query and ranks them, (ii) in (Verspoor et al., 2020), a system was developed to help medical researchers easily find research articles relevant to COVID-19, and (iii) in (Hope et al., 2020), a system was built which identifies the relationships between researchers working on COVID-19 related projects and their networks with the aim to help researchers interested in COVID-19 related projects to easily identify potential collaborators and research ideas to investigate. In (Wei et al., 2020), a dataset of more than 1,500 questions from several sources were obtained and annotated to different categories

and a model was built to classify questions into these categories.

A resource made up of a dashboard was created by researchers at Johns Hopkins University for tracking the number of cases, information about contact tracing, testing, and information about work on the vaccine for COVID-19<sup>2</sup>. In (Guntuku et al., 2020), using Twitter data, a dashboard was created to track (over time) the changes in language use (related to mental health and COVID-19 symptoms), in different states in the United States. Similarly, the World Health Organisation created a dashboard for monitoring the number of COVID-19 cases globally<sup>3</sup>.

Some prior work used social media data to gain insights into peoples reactions to the COVID-19 pandemic. Posts from a Reddit forum focused on discussions around COVID-19 were analyzed to understand the patterns in topics discussed and how they change over time in the forum (Stokes et al., 2020). In Zhang et al. (2020), two Reddit forums focused on discussions around COVID-19 were analyzed to gain insights about peoples responses to the pandemic. In Zong et al. (2020), Twitter data related to COVID-19 events such as positive and negative COVID-19 tests were labelled and models were developed to automatically identify these events. In (Chen et al., 2020), a multilingual COVID-19 Twitter dataset was created with the aim to aid researchers better understand what people are conversing about on social media, as it relates to the pandemic and enable researchers study COVID-19 related misinformation. In (Ayers et al., 2020), internet searches during the time period of the pandemic were studied and it was determined that peoples expression of anxiety increased. In (Serrano et al., 2020), a method was created for detecting COVID-19 related misinformation in Youtube videos by using the comments to the videos. In (Aggarwal et al., 2020), the differences in language use in COVID-19 related Reddit discussions between female and male users was studied.

The main difference between our work and prior work is that in this work we build 4 machine learning models to measure different social supports expressed in posts in an online COVID-19 forum with the aim to better understand the types of social support people were expressing in these posts and

<sup>2</sup><https://coronavirus.jhu.edu/map.html>

<sup>3</sup><https://covid19.who.int/>

how they changed over time.

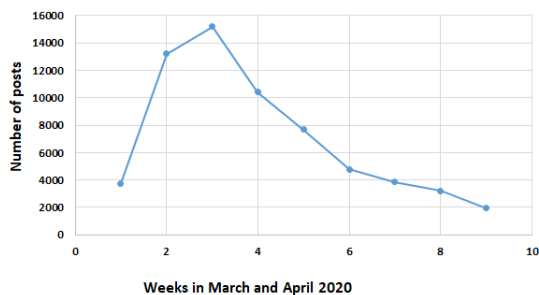
### 3 Dataset

Our dataset consists of posts published on a Reddit forum, */r/Coronavirus* - which is focused on discussions around COVID-19 and is the COVID-19 subreddit with the most number of members (i.e. 2.4 million members as of January 2021). Specifically, we collected 64,074 posts published daily in the “Daily Discussion Post” thread in the */r/Coronavirus* subreddit between March 3 and April 30 2020 (Stokes et al., 2020). Table 1 shows information about the number of posts published in March and April, respectively. Figure 1 shows information about the number of posts published weekly in our dataset; it can be observed that the number of posts published weekly peaked in March and steadily declined afterwards.

Table 1: Number of posts in our dataset published in March and April 2020

Month	Number of Posts
March 2020	45,880
April 2020	18,194

Figure 1: Graph showing the number of posts published weekly in our dataset



In this work, we aim to determine the types of social support being expressed in posts published in the */r/Coronavirus* forum; while analyzing the comment responses to these posts would give insights as to how users respond to posts in this forum, our goal in this work is to better understand the social support expressed in posts in this forum and how it changes over time. Hence, in our analysis, we exclude the comments these posts received.

### 4 Social Support

The importance of social support expressed in online health forums has been demonstrated (Wang

et al., 2012; Yang et al., 2017).

Similar to prior work (Wang et al., 2012), we measure the extent of social support expressed in posts. We had 3 medical students rate a randomly selected sample of 1,000 posts from our dataset; Table 2 shows a brief description of the sample posts.

Table 2: Summary of sample dataset

Attributes	Result
Avg. number of sentences per post	3.6
Avg. number of words per sentence	52.53
Avg. standard deviation of number of sentences	2.23
Avg. standard deviation of words per sentence	39.5

Given a post, the annotators rated the posts on: (i) the extent of emotional support given, (ii) the extent of emotional support sought, (iii) the extent of informational support given, and (iv) the extent of informational support sought. Specifically, using a 7-point Likert scale, where 1 meant “a particular social support was not expressed” and 7 meant “a particular social support was expressed a lot” the annotators rated the 1,000 sample posts. For each post, the annotators ratings were aggregated by averaging their ratings and each post had a numeric score between 1 and 7 expressing the extent of emotional support given, emotional support sought, informational support given, and informational support sought, respectively in the post. Using intra-class correlation (ICC) (Bartko, 1966), we measured the reliability of the annotators; the ICC for emotional support given, emotional support sought, informational support given, and informational support sought was 0.650, 0.660, 0.778, 0.918, respectively.

We extracted the following features from each post (Wang et al., 2012):

#### 4.1 Features

We extract the following Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015) categories which are relevant to emotional and informational support: “cognitive mechanism”, “verbs”, “religion”, “negative emotion”, “positive emotion”, “biological processes”, “present focus”, “past focus”, “future focus”, “time”, “death”, impersonal pronoun”, personal pronouns (i.e. “1st person singular”, “1st person plural”, “2nd person”, “3rd person singular”, “3rd person plural”). LIWC is a psycholinguistic dictionary made up of several categories with curated words associated with each category (Wang et al., 2012).

From each post we: (1) count the number of

words and sentences, (2) count the number of sentences that contain negation words such as "not", (3) count the number sentences which are phrased as questions; these sentences were identified as follows: (i) sentences that ended with a question mark, (ii) sentences that began with a modal verb e.g. "could you", and (iii) sentences that began with an indirect question pattern e.g. "would it be possible" (Wang et al., 2012).

The number of certain parts-of-speech tags, strong subjectivity words e.g. "abuse" and weak subjectivity words e.g. "absence" (Wilson et al., 2005) were counted (Wang et al., 2012).

We identified and counted posts involving requests or advice (Wang et al., 2012) by using the following word patterns: (a) "you" and a modal verb, which counts the number of sentences that start with the pronoun "you" and then a modal verb e.g. "might" and (b) "please" and a modal verb which identify sentences that start with the word "please" and then a verb (Wang et al., 2012).

Using the Mallet (McCallum, 2002) implementation of Latent Dirichlet Allocation (LDA) (Blei et al., 2003), we generated 20 topics from posts in our dataset (section 3) (we varied the number of topics and 20 topics gave the most coherent topic themes when examined by one of the co-authors) (Wang et al., 2012); Table 3 shows the top 5 words associated with the LDA topics.

## 4.2 Social Support Prediction Model

Given the language features extracted from posts in section 4.1, using Scikit learn (Pedregosa et al., 2011), we built four machine learning Random Forest regression models. For each post, these models produced numerical values (between 1 and 7) indicating the extent of: (i) emotional support given, (ii) emotional support sought, (iii) informational support given, and (iv) informational support sought. The 1,000 annotated posts from section 4 were randomly partitioned into training set (80%), validation (10%), and testing set (10%). The validation set was used to evaluate the models performance and when a satisfactory performance was obtained, the models were then evaluated on the testing set. To measure the models performance, we used Pearson's correlation (Wang et al., 2012) to correlate the models outputs with the average annotator labels for each post.

Table 3: Top 5 words associated with LDA topics

Highly correlated words associated with LDA topics
work, home, job, company, pay symptom, fever, cough, sick, flu people, think, saying, right, pandemic good, essential, thanks, great, hope lockdown, China, virus, million, infected care, hospital, medical, health, workers family, quarantine, self, friend, parents people, risk, high, immune, old virus, hands, spread, wash, infected days, weeks, today, ago, yesterday positive, test, tested, gloves, negative question, corona, ask, answer, curve coronavirus, covid, flu, vaccine, sars state, people, city, died, county death, rate, social, distancing, recovered home, food, store, going, grocery open, closed, online, school, America mask, use, wear, face, protect cases, deaths, numbers, testing, Italy months, long, end, states, travel

## 4.3 Result from Social Support Prediction Model

On the annotated dataset (section 4), these models correlated with the average annotator ratings as shown in Table 4. The models were then applied to all the posts in our dataset (section 3). Each post ended up having scores for the emotional support given, emotional support sought, informational support given, and informational support sought, respectively. Table 5 and Table 6 shows the 5 most important features (these features are ranked by the Random Forest models).

Table 4: Results from the analysis to determine the correlation between the social support models and the annotator ratings: all the p-values were less than 0.001

Social Support	Pearson's Correlation
Emotional Support Given	0.400
Emotional Support Sought	0.468
Informational Support Given	0.503
Informational Support Sought	0.547

Table 5: Top 5 most important features as ranked by the Random forest model for emotional support given and emotional support sought

Emo. Support Given	Emo. Support Sought
Number of words (0.188)	Number of words (0.213)
Strong Subjectivity (0.165)	Noun (0.171)
Noun (0.145)	Strong subjectivity (0.147)
Adjective (0.126)	Adjective (0.111)
Weak subjectivity (0.109)	Sentence length (0.101)

Table 6: Top 5 most important features as ranked by the Random forest model for informational support given and informational support sought

Info. Support Given	Info. Support Sought
Number of words (0.196)	Question (0.199)
Noun (0.173)	Number of words (0.169)
Strong subjectivity (0.115)	Noun (0.136)
Sentence length (0.091)	Strong subjectivity (0.112)
Question (0.072)	Sentence length (0.098)

## 5 Correlate the different types of Social Supports expressed in posts

In this section, we aim to determine if there are correlations between the types of social supports expressed in posts in our dataset. This is important for better understanding the support needs of the individuals publishing posts on the forum, and can guide the moderators of the forum and users responding to these posts to tailor their responses accordingly to match the support needs expressed in the post.

For each post in our dataset (section 3), given features that represent the predicted scores for the emotional support given, emotional support sought, informational support given, and informational support sought, respectively, we use the Pearson’s correlation to correlate the feature that represents the emotional support given with the features that represent the informational support given and informational support sought, respectively. Similarly, we correlate the feature that represents the emotional support sought with the features that represents the informational support given and the informational support sought, respectively. Tables 7 and 8 show these correlations.

Table 7: Correlation between the emotional support given and the informational support given and sought, respectively. The p-values are less than 0.05

	Emo. Support Given
Informational Support Given	Pearson $r = 0.721$
Informational Support Sought	Pearson $r = -0.069$

Table 8: Correlation between the emotional support sought and the informational support given and sought, respectively. The p-values are less than 0.05

	Emo. Support Sought
Informational Support Given	Pearson $r = 0.381$
Informational Support Sought	Pearson $r = 0.019$

## 6 Changes in Social Support expressed in posts Over Time

In this section, we aim to determine how social supports expressed in posts in our dataset changes

over time. Given the posts in our dataset and features that represent their corresponding scores for emotional support given, emotional support sought, informational support given, and informational support sought, respectively, we aim to determine the statistical differences between these features and then determine how these social supports change over time.

First, for all the posts in our dataset, we use t-test to determine the difference between the features that represent the emotional support given and emotional support sought, respectively; also, we use t-test to determine the difference between the features that represents the informational support given and informational support sought, respectively, in posts in our dataset. The results from the t-tests are shown in Table 9.

Table 9: The results of the t-tests. The p-values were less than 0.05. *Emo. Given* indicates the emotional support given, *Emo. Sought* indicates the emotional support sought, *Info. Given* indicates the informational support given, and *Info. Sought* indicates the informational support sought

	T-value
Emo. Given and Emo. Sought	- 51.56
Info. Given and Info. Sought	176.97

Second, we aim to determine how these social supports changed over the time period in which we collected our dataset i.e March and April 2020. For posts published in the same week, we averaged their scores for emotional support given, emotional support sought, informational support given, and informational support sought, respectively. As shown in Figure 2, we observed that emotional support sought increased over time and the emotional support given increased in the beginning of March and reduced by April. Also, we observed that informational support given and sought followed a similar pattern, however, users published more posts in which they gave informational support, as shown in Figure 3.

## 7 Language features associated with time periods with clear differences in the types of social supports expressed in posts

From section 6, it was observed that the emotional support sought increased over time and the emotional support given increased in March and then decreased over time.

Figure 2: Graph showing the change in the average weekly emotional support sought and emotional support given in posts from our /r/Coronavirus dataset

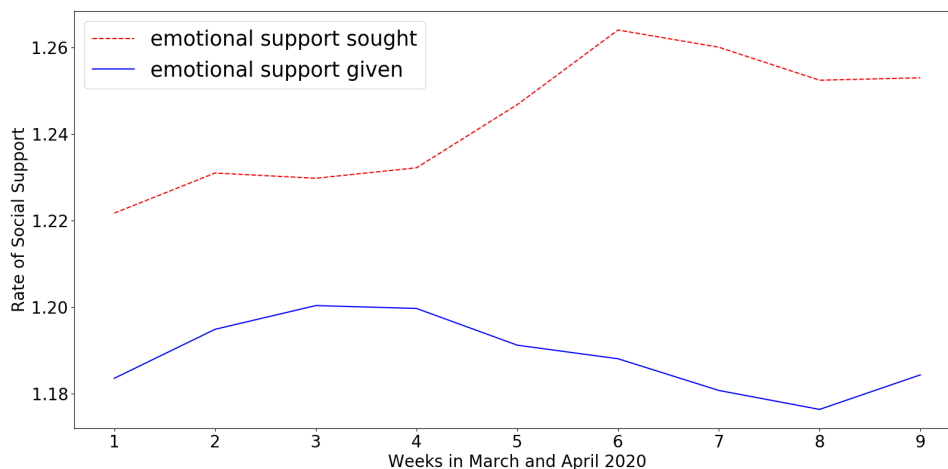


Table 10: Topics associated with posts from our dataset published in March

LDA Topics and highly correlated words	Cohen's D
travel, school, week, trip, cancel, flight, break, online, university, college	0.183
home, family, live, stay, parents, house, mom, sick, dad, friend	0.109
symptoms, fever, cough, days, throat, sick, flu, feel, cold, coughing	0.102

In this section, we aim to determine the language features associated with the time periods in which there was a clear change in the emotional support sought and emotional support given. Specifically, from figure 2, we determine week 4 (which corresponds to the end of March 2020) as the start of the ascent of the emotional support sought and the descent of the emotional support given, hence using the following language features: LDA (Blei et al., 2003) and LIWC (Pennebaker et al., 2015), we aim to determine the LDA topics and LIWC categories associated with posts published in these different time periods (i.e. March 03 - March 31 and April 01 - April 30 2020). Potentially, the LDA topics and LIWC categories associated with these different time periods should give some insights as to why there was a change in the emotional support given and emotional support sought.

## 8 LDA

We identified topics associated with posts published between March 03 2020 - March 31 2020 and April 01 - April 30 2020 in our dataset by generating 20 LDA topics (Blei et al., 2003) from posts in our dataset; we tried different numbers of top-

ics and 20 topics obtained the most coherent topic themes when examined by one of the co-authors. Tables 10 and 11 show the LDA topics associated with posts published in March and April, respectively.

## 9 LIWC

In this section, using LIWC (Pennebaker et al., 2015), we identify the LIWC categories associated with posts published in March compared to those published in April in our dataset. Tables 12 and 13 show the LIWC categories associated with posts published in March and April.

## 10 Discussion and Future Work

This work has several findings; in this section we summarize the results from our analysis.

We find that in posts from our dataset, the emotional support given has a positive correlation with the informational support given; the same applies to the emotional support sought. This implies that when expressing emotional support (giving or seeking), users tend to give informational support as well. Knowing this is important in that it can help the forum moderators and members of the forum

Figure 3: Graph showing the change in the average weekly informational support sought and informational support given in posts from our /r/Coronavirus dataset

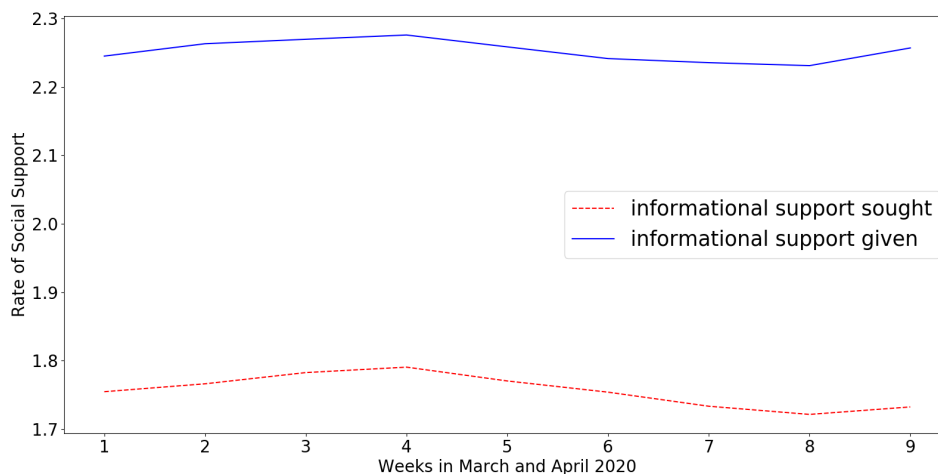


Table 11: Topics associated with posts from our dataset published in April

LDA Topics and highly correlated words	Cohen's D
cases, deaths, rates, number, infected, population, million, total, mortality, Italy	0.181
news, post, find, information, coronavirus, article, link, comments, read, source	0.153
months, normal, lockdown, point, start, vaccine, end, curve, quarantine, life	0.143

Table 12: Results from LIWC categories associated with posts from our dataset published in March

LIWC category	Cohen's D
First person singular pronoun	0.145
Health	0.133
Biological process	0.116
Anxiety	0.115

Table 13: Results from LIWC categories associated with posts from our dataset published in April

LIWC category	Cohen's D
Death	0.155
Anger	0.100

responding to posts to better understand the support needs expressed in these posts, thereby helping inform them on how to best respond to these posts.

Given the scores for emotional support given, emotional support sought, informational support given, and informational support sought for each post published daily in our dataset, using t-test, we determine that the features representing the emotional support given and sought, respectively are

statistically different; we do the same for features representing the informational support given and sought, respectively and also determine that these features are statistically different. After determining the statistical difference between these social supports, we then conducted an analysis to determine how the social support expressed in posts in our dataset changes over time i.e. between March and April 2020. We observed that: (i) over time, users sought more emotional support and gave less emotional support and (ii) users gave more informational support compared to informational support sought. These changes in social support needs can provide insights as to the well-being and needs of the public, thereby informing the forum moderators/members and the necessary authorities on the kinds of interventions to provide. For example, given that overtime, users in the forum sought more emotional support, online psychiatry services may be provided. Another finding from this work is that users in the forum tended to give more informational support. We observed that a significant number of posts were relaying information reported by or obtained from other sources such as news articles or websites. In some cases, the veracity of the

information given are not determined, hence the potential for the spread of misinformation about the pandemic; this highlights the importance of having experts engage with these online forums in order to provide and verify information.

Given the changes in the expression of social supports (emotional and informational) in our dataset, we identified time periods in which there was a clear distinction in the types of social support expressed in posts. We observed that emotional support sought in posts increased by the beginning of April while the emotional support given decreased. In order to better understand why there was a change in the emotional support sought and given between March and April, using LDA and LIWC, we compared the posts published in this time periods. The intuition for this comparison being that the language use differences in posts published in these time periods may provide insights as to why there was a change in these expressed social supports. We observed that in March, people posted about topics related to traveling and COVID-19 symptoms as shown in Table 10 and used more words associated with the LIWC categories on health and anxiety (Table 12). Also, we observed that in April, people tended to post more about topics related to the numbers of deaths and cases and the lockdown, vaccine, and quarantine (Table 11) and used more words associated with the LIWC categories on death and anger (Table 13). From this language analysis, a potential reason for why the emotional support sought increased in April and the emotional support given decreased is that with the increase in the number of cases and deaths, the lockdown in various places, and the uncertainty of when the pandemic will end and things will go back to normal, people were worried and concerned, hence they sought more emotional support and gave less emotional support. We selected 100 posts with the highest emotional support sought in March and April, respectively, and one of the co-authors went through these posts and it was found that these post were reflective of the LDA and LIWC features that delineate posts published in March and April.

In this work, we tracked social support changes in a 2-months time period; in the future, we aim to track these changes over a longer time period and also track these changes in other Reddit forums focused on discussions around COVID-19. Also, in the future, we aim to analyze COVID-19 related

posts in other online platforms such as Twitter, to study the social supports expressed on these platforms and how they vary or are similar from one platform to the other.

**Limitations:** In this work, we analyzed posts published in the Reddit forum */r/Coronavirus*. Prior work determined that different forums, despite their similarity in topics discussed, attract users with different interests (Tran and Ostendorf, 2016), therefore, the support sought and language used in other COVID-19 related online forums may differ from that of */r/Coronavirus*. Hence the results from our analysis may vary in other COVID-19 online forums or other online platforms.

**Ethics and Privacy** For all the analysis in this work, we did not use Reddit user names and no Reddit user was contacted. We had regular weekly meetings with the annotators (who are co-authors of this paper) to address any concerns they might have and to ensure that the content of the data they were labeling did not negatively impact them and their well-being.

## 11 Conclusion

In this work, we build four machine learning models to measure the extent of the emotional support given, emotional support sought, informational support given, and informational support sought, in posts published on an online forum focused on discussions around COVID-19. We carried out analyses to: (i) determine the correlations between these social supports expressed in posts, (2) determine how these social supports change over time, and (3) using language features, we conducted an analysis to gain insights as to why there was a change in the social supports expressed. The findings from this work show that data from social media can help gauge how people are feeling and responding to the COVID-19 pandemic and track and understand the social support needs of people as it relates to the pandemic. Potentially, these findings can help public health agencies and other agencies design and implement online interventions around COVID-19.

## References

Jai Aggarwal, Ella Rabinovich, and Suzanne Stevenson. 2020. Exploration of gender differences in covid-19 discourse on reddit. *arXiv preprint arXiv:2008.05713*.



- John W Ayers, Eric C Leas, Derek C Johnson, Adam Poliak, Benjamin M Althouse, Mark Dredze, and Alicia L Nobles. 2020. Internet searches for acute anxiety during the early stages of the covid-19 pandemic. *JAMA Internal Medicine*, 180(12):1706–1707.
- John J Bartko. 1966. The intraclass correlation coefficient as a measure of reliability. *Psychological reports*, 19(1):3–11.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273.
- Andre Esteva, Anuprit Kale, Romain Paulus, Kazuma Hashimoto, Wenpeng Yin, Dragomir Radev, and Richard Socher. 2020. Co-search: Covid-19 information retrieval with semantic search, question answering, and abstractive summarization. *arXiv preprint arXiv:2006.09595*.
- Sharath Chandra Guntuku, Garrick Sherman, Daniel C Stokes, Anish K Agarwal, Emily Seltzer, Raina M Merchant, and Lyle H Ungar. 2020. Tracking mental health and symptom mentions on twitter during covid-19. *Journal of general internal medicine*, pages 1–3.
- Tom Hope, Jason Portenoy, Kishore Vasani, Jonathan Borchardt, Eric Horvitz, Daniel S Weld, Marti A Hearst, and Jevin West. 2020. Scisight: Combining faceted navigation and research group detection for covid-19 exploratory scientific search. *arXiv preprint arXiv:2005.12668*.
- Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.
- James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.
- Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. Nlp-based feature extraction for the detection of covid-19 misinformation videos on youtube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.
- Daniel C Stokes, Anietie Andy, Sharath Chandra Guntuku, Lyle H Ungar, and Raina M Merchant. 2020. Public priorities and concerns regarding covid-19 in an online discussion forum: Longitudinal topic modeling. *Journal of general internal medicine*, page 1.
- Trang Tran and Mari Ostendorf. 2016. Characterizing the language of online communities and its relation to community reception. *arXiv preprint arXiv:1609.04779*.
- Karin Verspoor, Simon Šuster, Yulia Otmakhova, Shevon Mendis, Zenan Zhai, Biaoyan Fang, Jey Han Lau, Timothy Baldwin, Antonio Jimeno Yepes, and David Martinez. 2020. Covid-see: Scientific evidence explorer for covid-19 related research. *arXiv preprint arXiv:2008.07880*.
- Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, et al. 2020. Cord-19: The covid-19 open research dataset. *ArXiv*.
- Yi-Chia Wang, Robert Kraut, and John M Levine. 2012. To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 833–842.
- Jerry Wei, Chengyu Huang, Soroush Vosoughi, and Jason Wei. 2020. What are people asking about covid-19? a question classification dataset. *arXiv preprint arXiv:2005.12522*.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 347–354.
- Diyi Yang, Robert Kraut, and John M Levine. 2017. Commitment of newcomers and old-timers to online health support communities. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 6363–6375.
- Jason Shuo Zhang, Brian C Keegan, Qin Lv, and Chenhao Tan. 2020. A tale of two communities: Characterizing reddit response to covid-19 through/t/china\_flu and/t/coronavirus. *arXiv preprint arXiv:2006.04816*.
- Shi Zong, Ashutosh Baheti, Wei Xu, and Alan Ritter. 2020. Extracting covid-19 events from twitter. *arXiv preprint arXiv:2006.02567*.