Network structure and people's attitude within work from home community on Reddit

Yujing Sun

Abstract

Since the beginning of Covid-19 pandemic, working remotely is encouraged by many companies as a solution to tackle the lockdown policy and potential contagion. Not everyone can smoothly switch to this working style, leaving some problems that employees face when they try to adapt to it. But whether people's attitude could change over time? As the new norm is widely accepted for more than two years, which provides a longer timeline to study, I focus on how do people change their attitude and concerns towards work from home. Using the Reddit posts of r/workfromhome community, I analyze the network structure within the r/workfromhome community and test whether the sub-communities in Reddit are grouped by different topics. The results show the appearance of "answer-person" role in the center of sub-community. The sub-communities after community detection are not grouped by different topics. Then I perform sentiment analysis and topic modeling on the posts to extract the topics and positive/negative sentiment about work from home. First, I conduct the synchronic analysis on the posts related with work remotely to shed light on the overall attitude and concerns on this new working style. The results show that the overall attitude of WFH on Reddit is positive. The sentiment of different topics is distinct. Then I study the temporal change of the attitude and focus over the course of pandemic. The positive rate of posts is high at the beginning of the pandemic as well as during the first, second and fourth wave of the Covid-19. Also, the work remotely is an adaptive process with different attitudes in different stage. These findings could be beneficial to employers and policy makers to recognize that working from home is a complex social issue and then adjust the related policy. For further study, my study also offers a guide about the importance of the temporal analysis when analyzing telecommuting policy's influence.

Introduction

Covid-19 began to hit the whole world at the beginning of 2020 since the first case had appeared in Wuhan, China. CDC (Centers for disease control and prevention) confirmed the first case of Covid-19 in the US on January 20, 2020. After that, this contagious virus has spread widely and caused adverse effects on the economy. In March 13, 2020, President Donald J. Trump declared a nationwide emergency, followed by the social distance and lockdown policy just few days after. The companies cannot afford the long-term shutdown and stop of production caused by the government policy, and they advocate their employees to work from home to continue the business.

Since many people work remotely involuntarily, their ability and attitude toward the work from home are not the same. Those who have higher education and income, or work in information industries have higher probability to work from home (Bartik et al. 2020; Brynjolfsson et al. 2020). Also, females and males have different ability to switch to work from home (Adams et al. 2020). For those who can work from home, this lifestyle causes several impacts in different perspectives. Some of the studies show that work from home increases the productivity of the employees since they have less request to break and have better and quieter working environment (Bloom et al. 2015). However, others report a reduced efficiency due to the longer communication time and less coaching time (Gibbs, Mengel, and Siemroth 2021). The conclusions of these study are made by survey a particular industry or a company. The conflict of the conclusion reflects the variation of the productivity among distinct companies, and the survey could only present small picture of the story.

Work remotely also causes some mental health issue to workers. Feeling stressed is a common phenomenon as people find it hard to adapt to information undersupply, autonomy, unprepared and ill-equipped, potentially longer working hours, and isolation (Carillo et al. 2021; Weinert et al. 2015). However, the existing studies only record the synchronic mental effect instead of the temporal mental effect during the pandemic. When people start to work at home, the boundary between family and work becomes vague, making employees easily switch their roles, but may affecting the efficiency of work (Ashforth, Kreiner, and Fugate 2000). The studies of the co-workers relationship also have divergent results on whether employees think telecommuting hurts their relationship with their co-workers (Fay and Kline 2011; Sias et al. 2012).

All these researches shed light on the influence and people's attitude toward work from home. However, most of the studies are synchronic which cannot reflect the change of attitude and influence. The conflict of some arguments may cause from the different sample or stage of the adaptation. There is no paper focusing on the process of adaptation of work from home over different time period. People's adaptation ability can change while they get in touch with a new thing for longer time. Since Covid-19 has appeared for more than two years and many people are familiar and have experience with this new norm, I want to study how have people changed their attitude toward work from home during two years.

In order to have a bigger picture of the attitude of employees who work remotely, I scrape the data from Reddit instead of doing the survey. The advantages of using survey data are that we can get the direct answer that we care about by the interviewers. The problems are the data could only represent a small group of people and some of employees are reluctant to tell the truth as they are afraid their opinions could be supervised by their monitors when the survey is initiated by the employers (Brynjolfsson et al. 2020). Considering these potential issues, I use the social media data which has a wider user-based and is easy and cheap to collect. Some of the studies also take advantages of social media data to analyze the sentiment and topic during

work from home since pandemic (Gottipati et al. 2021; Zhang, Yu, and Marin 2021). These studies choose tweets as their data source. They only concentrate on the tweets within less than 4 months because the volume of posts on twitter is large. For my paper, I mainly focus on the change of attitude along more than two years. Collecting all the information from twitter is infeasible. Also, the half of tweets with hashtag #workfromhomme are advertisements which involve low quality data problem (Gottipati et al. 2021). This suggests that Reddit posts within the r/workfromhome community is a more suitable choice for me to detect the change.

In this paper, I have two research questions that I am interested in. First, whether the sub-communities within Reddit r/workfromhome community are grouped by different topics? Second, how have people's attitude toward work from home on Reddit changed during pandemic?

Before analyzing how have people's attitude changed toward work from home on Reddit, I also want to detect the inner network structure of the r/workfromhome community of Reddit. Compared to tweets that use a hashtag to represent the related topic, subreddits are grouped by the community. If you want to ask a question or post your opinions, you first need to join a community and then post within this community. Each community has its own rule and culture. Thus, there will be some special structures of network within the Reddit community. Some scholars find out that there is a special social role, the "answer-person" role, within Reddit community. This kind of users are those who like to respond to other person's questions but don't post a lot within the community (Buntain and Golbeck 2014). I want to test whether the "answer-person" is common within the r/workfromhome community and what are their characteristics. For the network analysis, I will only use the posts and comments within the r/workfromhome in 2021 since the network with 3 years data is too large to process. In this part, I will detect the features of the network, perform community detection, find out some patterns (e.g.: "answer-person" role) within the sub-network, and test whether the sub-community is grouped by different topics.

Beyond exploring the network within r/workfromhome community, I also conduct text analysis using sentiment analysis and topic modeling. I have two claims that I want to test in terms of the posts. First claim is related with the sentiment change during pandemic. Barrero et al. conduct survey on more than 30,000 Americans over multiple waves to know whether they think WFH will continue. The result show that people's increased support of work remotely in longer time is associated with the lingering concerns of being infectious by the diseases and the better-than-expected work from home experience (Barrero, Bloom, and Davis 2021). Thus, I hypothesize that people in the WFH Reddit community will tend to be more positive when the Covid-19 is more serious. The positive attitude in different time period is contributed by different benefits brought by work from home.

Second claim is related with change in themes that were discussed on the Work from

Home subreddit. Change to a new workplace or environment is an adaptation process (Bruque, Moyano, and Eisenberg 2008; Niessen, Swarowsky, and Leiz 2010). Since different industries have different rate of switching to work from home, and different people have different adaptation ability of change their habit (Bartik et al. 2020; Brynjolfsson et al. 2020). We know that adapt to telecommute is a process that involves different stages. I hypothesize that people's focus and concentration on work remotely along the process are varied. At first, people are intended to solve the urgent need and issues they face; while their focus change to be more comprehensive and diverse directions such as work-life balance, health issues, attitude about back to office etc.

Within this paper, the first part is network analysis, and the second part is text analysis.

Data

The data I use is the posts and comments in Reddit r/workfromhome community. Pushshift API is used to collect the data from Reddit. The data is from March 13, 2020 to April 12, 2022 which covering the whole timeline of the Covid-19 until now since former President Donald J. Trump declares a nationwide emergency in March 13, 2020. There are 6,180 subreddits from work from home community during this time period. For the first part of my study that focuses on the network of the community, I only use the data from 2021. Because I want to detect the community within the r/workfromhome community, I keep the "author name", "texts", and "topic link" of the posts, as well as "commenter name", "comments", and "topic link" of the comments. For the second part of my study, my focus is about how have people's attitude changed. I will use the "texts" and "time" of the posts.

Beyond the Reddit data, I also downloaded the Covid-19 New Cases data in the US and the global from CDC (https://covid.cdc.gov/covid-data-tracker/#trends_dailycases). Covid-19 data will be used to compared with the volume and sentiment of subreddits.

In order to detect the communities within the r/workfromhome community, I create the undirected network within the WFH community by connecting the post's author with the comment's author by the post's link. Because some of the comments don't contain enough information, I only keep those comments that have more than 20 words. With the data from 2021, I generate a graph that contain 2745 nodes. As the graph is too large and many nodes only have few edges, I drop the nodes whose degree is smaller than 5. The final graph has 530 nodes. Further study is applied based on this graph.

Several steps are tried to preprocessing the data for text analyzing. As for the sentimental analysis, removing the stop words could affect the result of the sentiment scores, I don't modify the text. When doing topic modeling, I remove the stop words using NLTK package's list. I also manually removed some words that have less meanings. Meanwhile, punctuations, numbers, URLs are removed, followed by lemmatization using Spacy package in python. After converting the sentences into

vectors, I added bigrams and trigrams into the list. Tfidf model is also used to get rid of some low value words.

Method

In terms of community detection, I use the Fiedler vector to cluster different groups. Fiedler vector is one of the methods to detect the community within non-directed graphs. Fiedler vector is the second smallest eigenvalue of the Laplacian matrix of the graph. This method is first invited in 1973 and is widely used in many community detection paper (Fiedler 1973; Chen and Hero 2014). Compared to the edge betweenness method such as Girvan Newman algorithm, Fiedler vector is associated with the sensitivity of algebraic connectivity. In my case, when I try the Girvan Newman method on my network, the community detection performance is not ideal. Thus, I use the second eigenvalue to be the indicator. The nodes with positive and negative values are clustering into two different groups. I apply the same method on the larger sub-group from the first step. Repeating this step recursively until the larger sub-group after one splitting contains less than 50 nodes, I generate 15 small sub-communities.

After that, I perform the Word2Vec on each word of the person in the small community and calculate the average vector of each person's post. Here, I use the google Word2Vec pre-trained model (https://code.google.com/archive/p/word2vec/). The Word2Vec vectors are pre-trained on part of Google News dataset (about 100 billion words). It is used to compute the word embedding of words by an efficient implementation of the continuous bag-of-words and skip-gram architectures. With each person's post vector, I find out the correlation of posts among the people within the community. This represents the similarity of user's posts within the community. For those who have low correlation, I want to know where is their position in the graph. Beyond that, I want to know whether the sub-community is grouped by different topics.

Two other models are performed to have a better understanding of people's attitudes, feelings, and mental reaction related with work from home, which are sentiment analysis and topic modeling. Sentiment analysis aims to detect people's public sentiment on some topics and issues. For those who post on Reddit, some of the posts are talking about statements, while some others contain obvious sentimental feelings that could be positive, negative, or neutral. Grabbing the trend of the sentiment difference could give us insights about how do people's emotional attitude about telecommute change. After topic modeling, I will also analyze the sentiments among different topics to have a further understanding of people's feeling. Here, I use lexicon-based sentiment classification that has already built by researchers. VADER (Valence Aware Dictionary and sEntiment Reasoner) is one of the lexicon and rule-based sentiment analysis tools that is specifically attuned to sentiments expressed in social media. NLTK package could load this method without applying labeled data. It automatically considers punctuation, emoji, capitalization, conjunctions, tri-gram and

then assign positive, negative, neutral, compound sentiment scores to each post. The compound score is calculated by summing the valence scores of each word in the lexicon, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive) (Hutto and Gilbert 2014). Because Reddit posts are social media data, this method is suitable for my study.

Topic modeling aims to identify people's interest within their posts. Along the time period since the beginning of the pandemic, we can have a clear review of how do the topics change since people are more adapted to this new norm and the pandemic experienced several waves. For the subreddits in work from home communities, the model could show what are people talking and concerned about with this topic. To conduct this method, I use LDA (Latent Dirichlet Allocation) which is an unsupervised algorithm in python's gensim library. I calculate the coherence scores to find out the best topics number of the model.

Network analysis

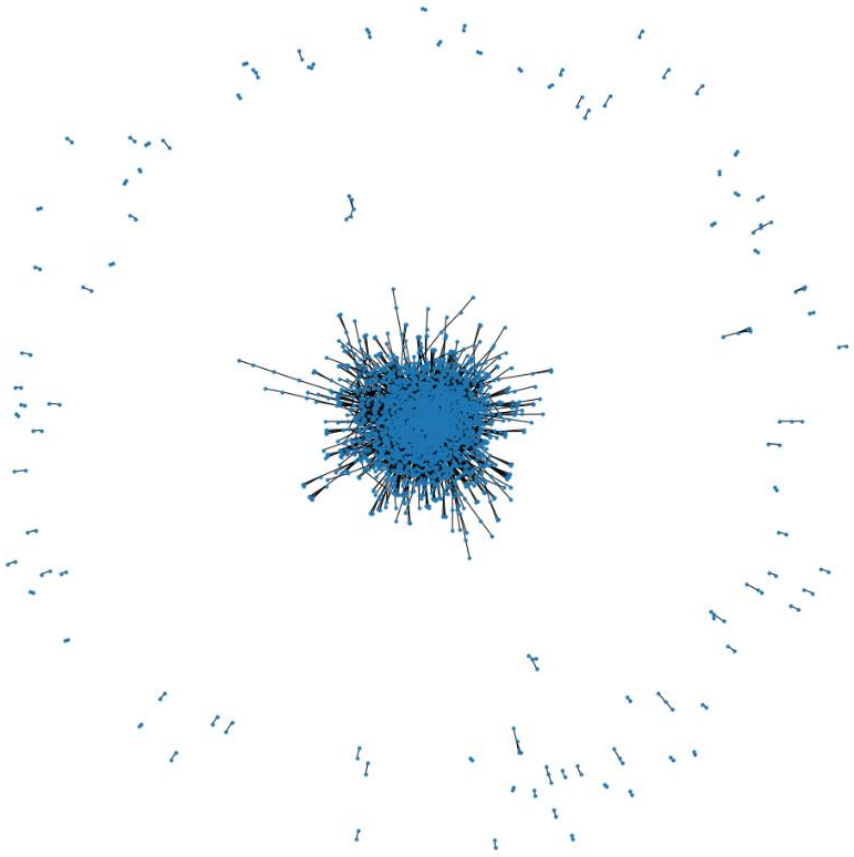


Figure 1

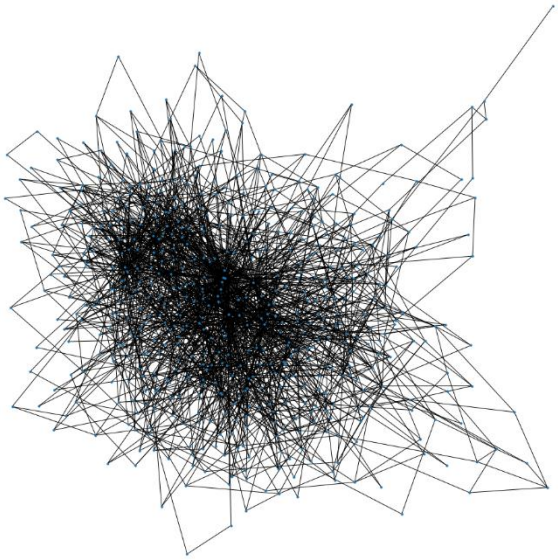


Figure 2

After connecting the post's author with the commenters under that post, I generate a network in Figure 1. We can see that most of the users within the network are connected in the central of the graph, while some of the small sub-communities are in the periphery of the graph. In this case, I drop the nodes with less than 5 edges to further simplify the graph. Figure 2 is the new graph after the selection. In this case, there are 530 users within the community.

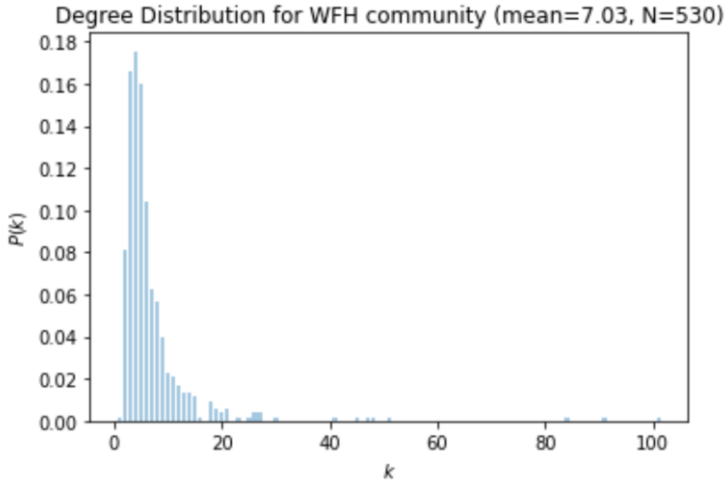


Figure 3

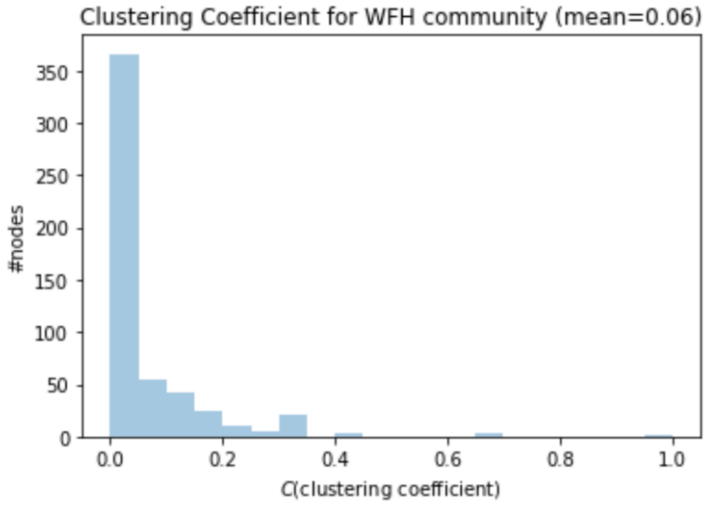


Figure 4

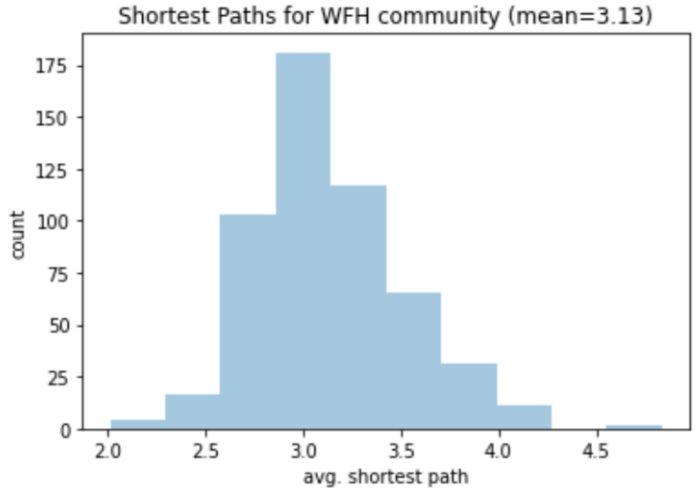


Figure 5

To detect the characteristics of this network, I plot the distribution of degree, clustering coefficient and average shortest path. The results are shown in Figure 3-5. Distribution of degree reflects the probability that a randomly chosen node has degree k. The mean is 7.03 showing the average degrees of the node. Because I only keep the nodes than have more than 5 edges, the distribution is quite reasonable. The connection within this community is not very close. The majority of the edge's number is smaller than 10. Clustering coefficient is the fraction of possible triangles through a node. The mean of clustering coefficient is very small, showing that few nodes have triangles. In the network, this means that most of the users within the community are not closely connected with a group of users. As for the average shortest path, the mean is 3.13. It presents like a normal distribution.

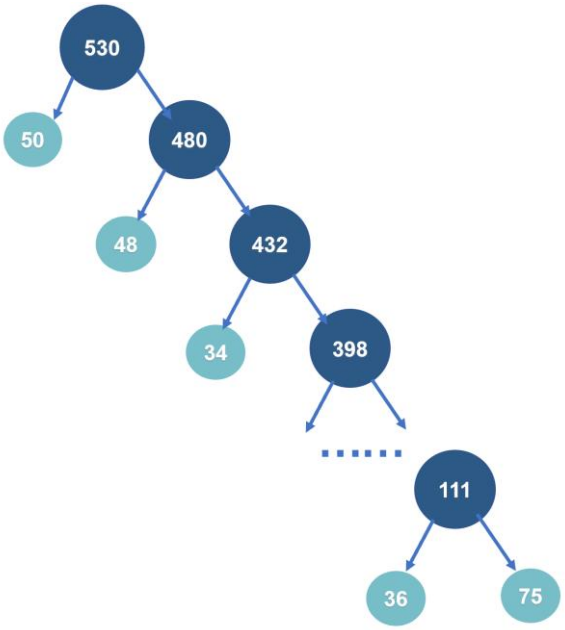


Figure 6

Since it is hard to find out the pattern of this network, I conduct the community detection to divide this network into some sub-networks. The result and step are showed in Figure 6. After applying the Fiedler vector classifier, the smallest sub-community contain 11 nodes, and the biggest one contains 75 nodes. The sub-community contains 35 nodes in average.

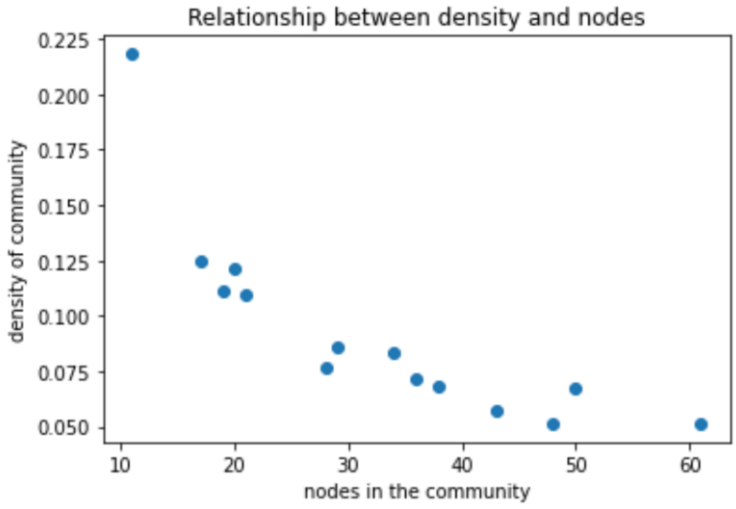
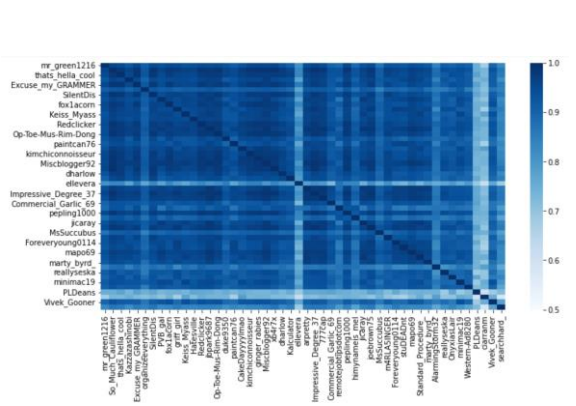


Figure 7

Since different sub-community contains different number of nodes, I want to know whether these communities have different density. In Figure 7, I plot the relationship between size of the group and density of the smaller sub-group after each splitting. The correlation of these two features is -0.82, meaning that the larger community will have lower density. This result is intuitive since within the larger community, there will be less connections among group members.



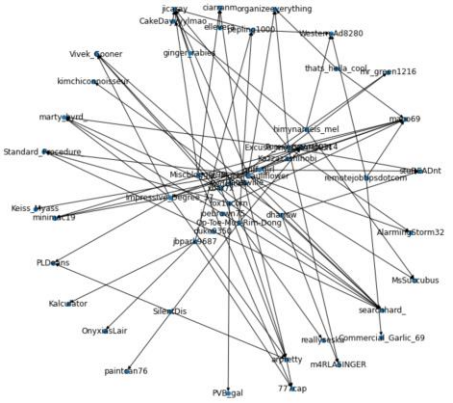
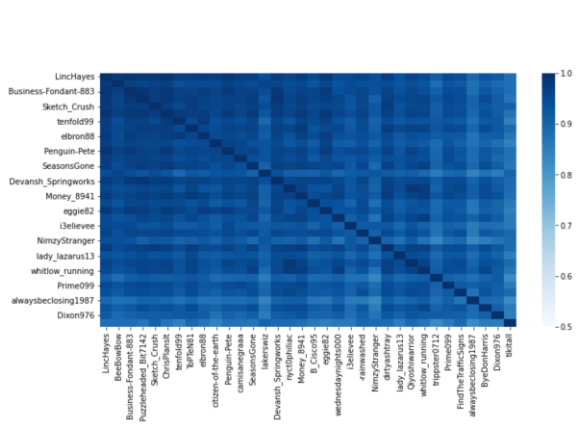


Figure 8: Correlation matrix of each member within community NO.0 (left)
Figure 9: Graph of community NO.0 (right)



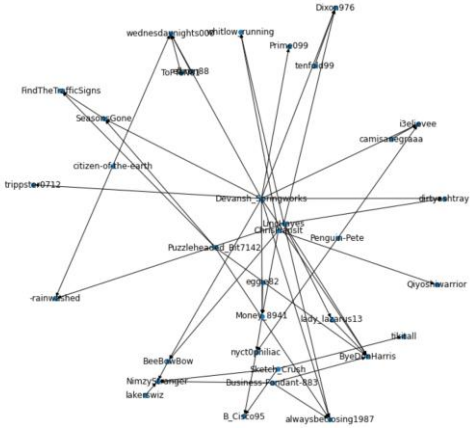
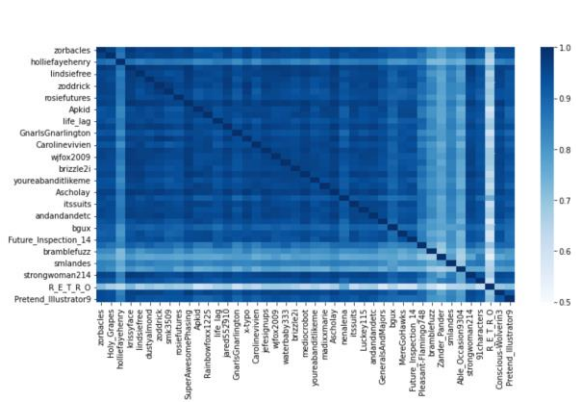


Figure 10: Correlation matrix of each member within community NO.2 (left)
Figure 11: Graph of community NO.2 (right)



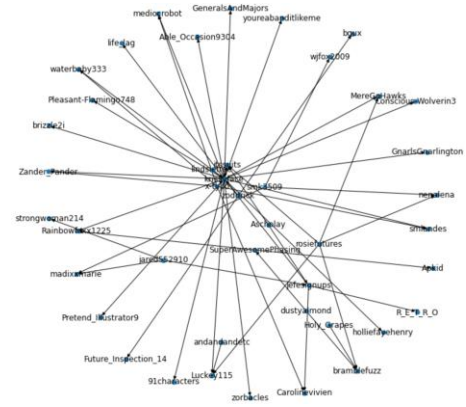


Figure 12: Correlation matrix of each member within community NO.6 (left)
Figure 13: Graph of community NO.6 (right)

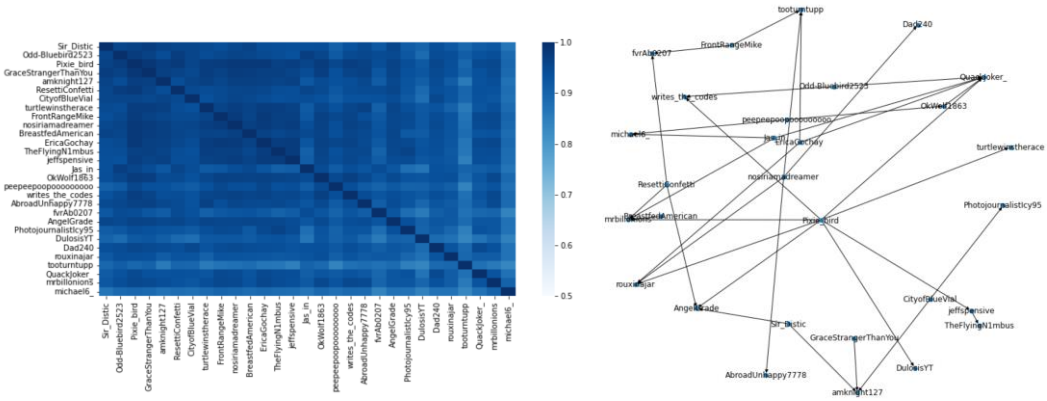


Figure 14: Correlation matrix of each member within community NO.9 (left)
Figure 15: Graph of community NO.9 (right)

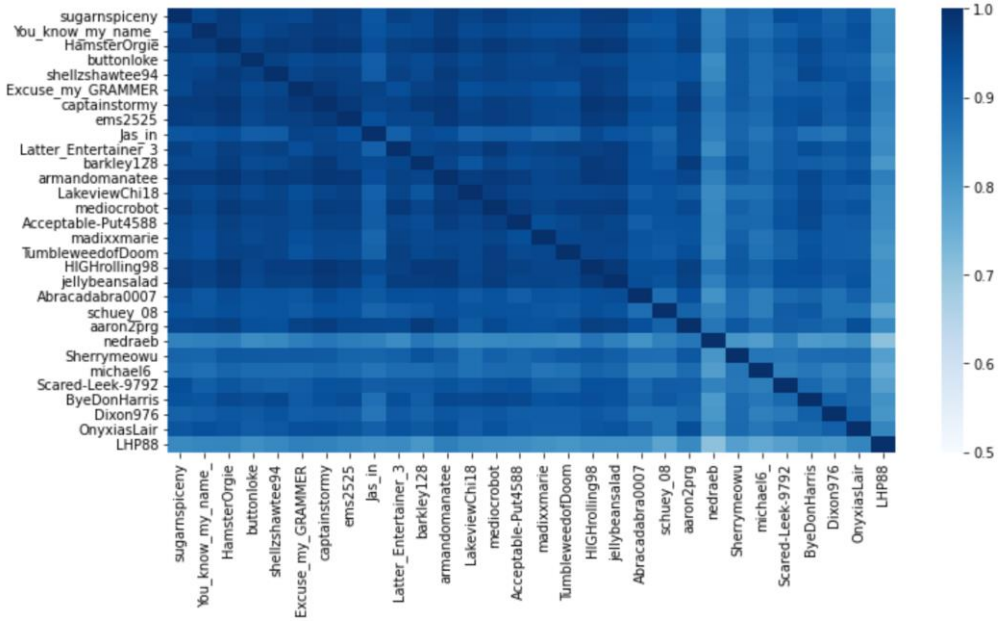


Figure 16: Correlation matrix of each member from group of random nodes

Within each sub-community, I also want to know whether they are talking about the same topic. If this is proved, then I could use the sub-community to divide the different topics. For each user, I collect all his/her posts and comments and aggregate them together. Then I calculate the average word embedding vector for each user's text within the small community. With the average word embedding vector of each person, I generate the correlation matrix based on the data. From the heatmaps, it is very clear that most of the members within the community have similar word embedding of their text. And the correlation matrix of the low-density and high-density network doesn't have significant difference. The values of correlation matrix are not related with the size of the community. For those big communities that have small density, the values within the correlation matrix of some communities are all very high (Figure 10 and Figure 14), while others have some relatively low correlations (Figure 8 and Figure 12). Most of the correlations between two persons are larger than 0.8. Small amount of the correlations is around 0.5. For those who distinctly have relatively lower correlation

among all other users within the sub-group, the position of this user is always at the periphery of the graph. In this case, I think it could be possible that the sub-network is related to the topic discussed within this group. Thus, I randomly choose two nodes from each sub-groups, and plot the correlation heat map of these 30 users' text. However, the correlations among these 30 users are still high (Figure 16). The results indicate that the sub-network is not grouped by topics. One potential explanation is that I combine both the posts and the comments together to represent a person. But this way could affect the result since it compresses many information of one person into one vector. This could lose some information. If I want to know the topics within the r/workfromhome community, topic modeling is still a useful tool to use compared to this way.

After that, I plot several sub-communities to learn the pattern within it. In this case, I want to know the post-comment relationship of the group. Thus, I choose to plot the directed network here. Interestingly, I find that the majority users in the center of the graph is the "answer-person". This kind of social role is the person who like to comment under other's post within the community. They like to give suggestions or share their stories. But they often don't post their question. These are a group of very essential users that activate and contribute to the whole small community. From the Figure 9, Figure 11, Figure 13, Figure 15, and Figure 18, we notice that there are many answer-persons within the Reddit r/workfromhome community.

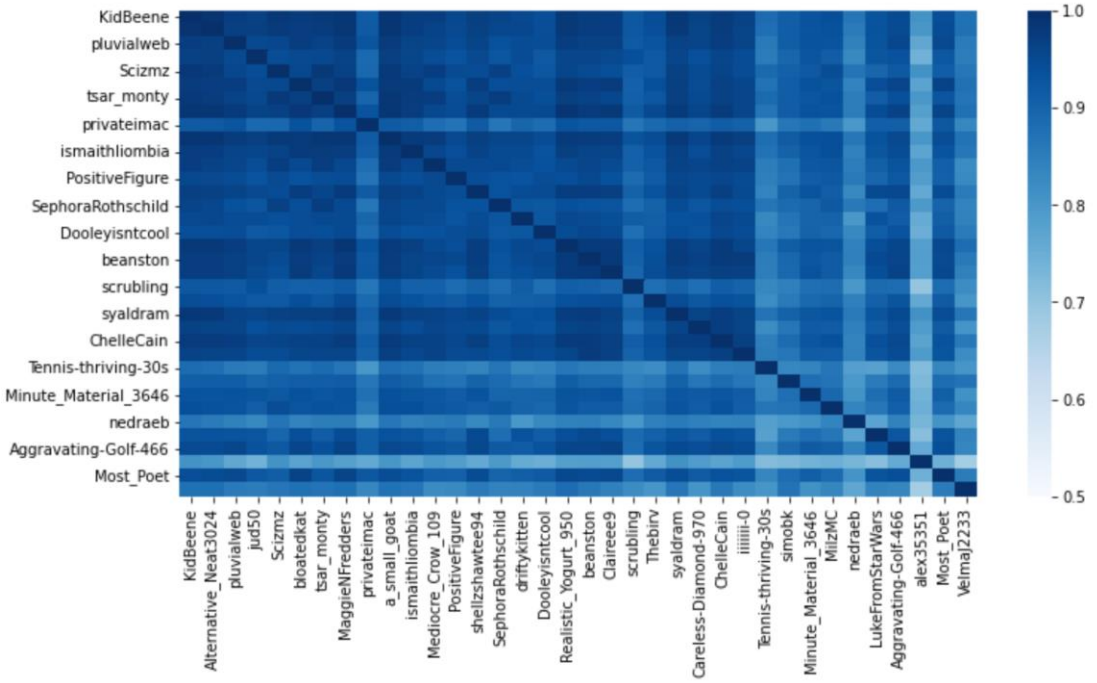


Figure 17: Correlation matrix of each member within community NO.13

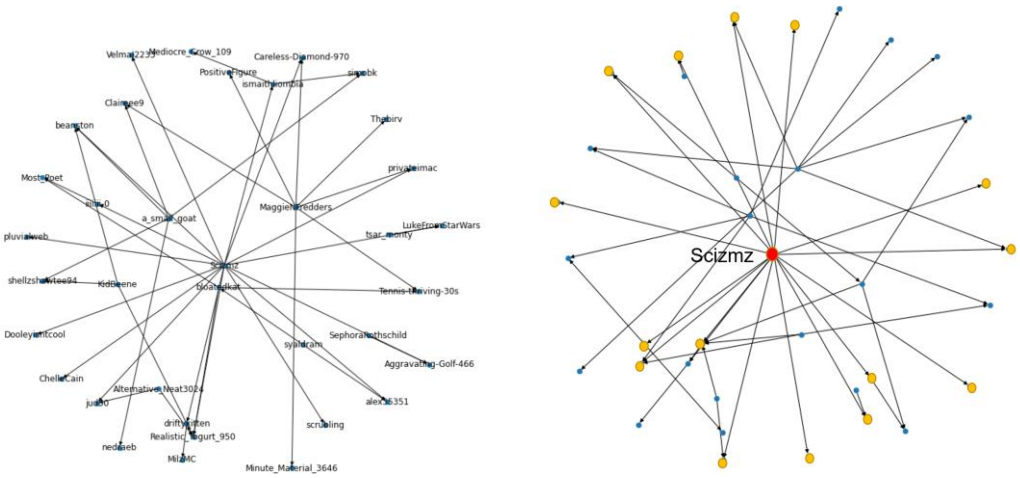


Figure 18: Graph of community NO.13

To have a closer look at the features of the answer-person, I choose one user as an example, and perform topic modeling to find out whether this person only answers one or few topics of posts. Here I choose the user “Scizmz” in the center of the NO.13 community in Figure 18. This user answers 16 questions in this community. First, I apply topic modeling on all the posts within 2021. The model presents 8 topics. Secondly, I find the most related topic of the posts returned by the model that the answer-person comments under. I count the number of posts for each topic. The results show that this answer-person comments on questions of several different topics. For example, “Scizmz” answers 5 questions related with how to increase work efficiency, 4 questions about equipment, 3 questions about work-life balance, 2 questions about team work issues, 1 question about mental health and 1 question about finding jobs. But this may not be the case for another answer-person. Further study could dive deeper into this study. From the result of the network analysis, I conclude that sub-communities within the Reddit r/workfromhome community is not grouped by topic. In order to understand the topics within the community, text analysis is a useful and helpful tool.

Text analysis

Figure 19

Before detecting the change of people's attitude, I first do a synchronic analysis on the posts. In order to know the important topics within the r/workfromhome community, I preprocess the data for the better cluster result. After lemmatizing and vectorizing the text, I add bigram and trigram into the data, as well as delete the stop words. I also create the list of my own stop words that are extracted from the result of topic modeling and substitute some words with their synonyms. From the LDA model, I get the topic and key words of the subreddits. I compute the Coherence Score of the model to find out the number of topics that is suitable for the texts. I choose 12 topics as the parameter of my final model considering it gives great coherence score. Figure 19 shows the topics and percentage of tokens for the Top-30 most relevant terms in that topic. I remove the smallest topic here because the texts within that topic don't present a distinct pattern. For the other 11 topics listed here, "find WFH jobs, interviews & experience for jobs" is the largest topic, accounting for 15% of total texts. Most of the discussion within this topic is asking help about finding remote job, self-introducing their previous work experience for advice of remote jobs, and sharing interview information. People also like to discuss the requirement of their company and their work experience during pandemic on Reddit. Asking for and giving recommendation of equipment (such as: desk, chair, headset, software, computers and so on) is another very popular topic among people. Since employees move into online, people also talk a lot about how to build a team online, how to maintain the relationship with coworkers and boss. Meanwhile, work-life balance, life style during WFH, discussion about back to office, healthy problems and time management are the other topics.

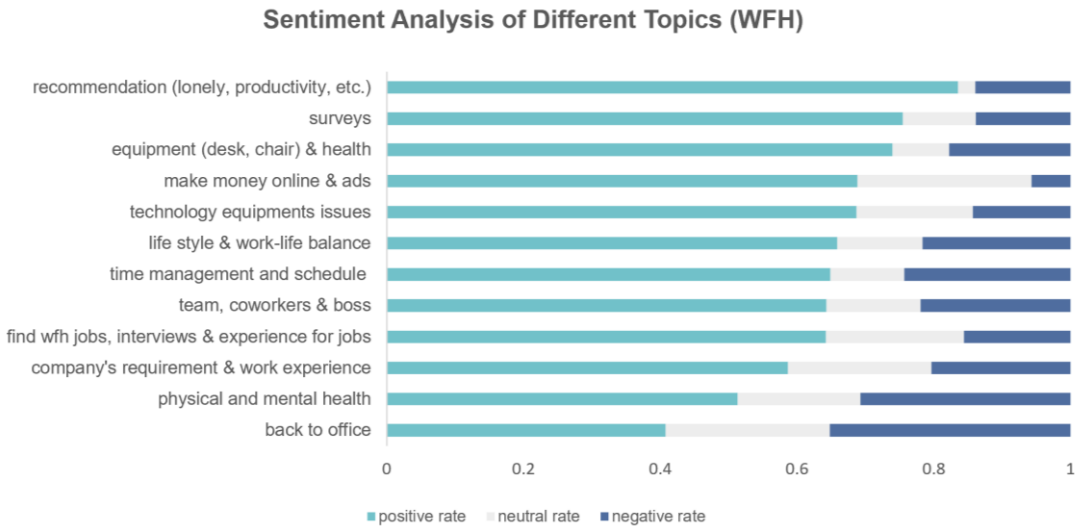


Figure 20

I perform sentiment analysis on each topic to detect people's overall feeling and attitude about different topics. Firstly, I get the contribution of each topic to every sentence. I assign the topic that has the largest contribution to that subreddit. Then, each subreddit has their domain topic and the percentage of contribution of that domain topic. I only keep those subreddits whose contribution of the domain topic is larger than 20%. I also notice that there is a topic contain content of both "back to office" and "physical and mental health" due to the common word "back". Thus, I manually separate this topic into two topics for further analysis. After that, I create a new dictionary where the keys are the topic's name, and the values are all the subreddits within that topic and whose domain topic's contribution reaches 20%.

The result of the sentiment analysis for each topic in r/workfromhome community is showed in this Figure 20. From the result, "recommendation", "survey", "equipment", "make money online" and "advertisements" are relatively positive compared to other posts. In contrast, "back to office", "physical and mental health", "team and coworkers", "life style and work-life balance" and "time management" are relatively negative topics (Table 1). More specifically, the topic "back to office" receives the highest negative rate since many posts are about people's unwillingness and complain of back to office request from their boss. "Physical and mental health topic" also has some negative feedback as people experience the back pain caused by the desk or chairs or unhealthy lifestyle at home, as well as lonely feelings and anxiety. Within the "time management and schedule" topic, the negative comments come from struggle of managing the hours of work. People also have problems with work-life balance since work-life boundary is vague and not everyone get used to it. This shows that not all the people on Reddit could adapt to the work from home pattern well. In terms of cooperation with other coworkers and boss, some employees think telecommute helps them to escape from the meaningless social life, while some think that working online decreases the efficiency of communicate and it is hard to connect with boss. For those positive posts, most of them are some objective contents such as

recommendation and advertisements. The result from the sentiment analysis reflects how do people enjoy the work from home and what are the issues caused from that.

Topics	Posts
Back to office	"Calling back to office: So they created a policy during the pandemic called choice and responsibility which allows people to exercise their choice in where they want to work as long as they deliver. The program has worked exceptionally well the last two years but now with the pandemic behind us they are asking folks to come to the office a few times a week. They are saying the policy above has two faces where we can exercise our choice but also have to be responsible and they find that many teams fo very well in a face to face setting. This is very disappointing. How do you deal with this?" "Back to the office 😒today: At least it's only 2 days per week. Had a good thing going there for a while. Saved so much money by not driving and going out to lunch. My Covid pup (20 months old) attacked me with kisses when I came home…like I was gone for months." "Been back to a hybrid home/office situation for 3 weeks now and it's just SILLY to have to be back there!....."
Time management	"Any tips for time management when starting to WFH? I find myself going hours without taking a break and then I'm burnt out more than when I worked in the field!" "WFH parents-- any advice on corralling a toddler during work hours: Hello! I am about to start a new job that is totally remote. I have a 1 year old son who is very high energy who will be home with me most days. I'm wondering if anybody has tips or advice on ways to optimize my workspace, keep my son entertained, schedule my day (emails, phonecards, meetings, etc), or just generally set myself up for success in this new role. I'm sure I'm not the only parent figuring this stuff out, so I'd love to know your secrets! TIA!" "If you worked 20 hours a week how would you split the time?"
Mental and physical health	"Back pain from sedentary life style (any tips?): About me: I'm 26 years old and I started two businesses and love my work. However, since I've been working from home for several years, I'm on my computer sitting for ~12-15 hours per day. I've invested into a great work station with great monitors, a herman miller chair, and a nice Uplift desk……Does anyone have any other tips/advice/tricks that they use to be a little more healthy? I get that I work a ton but I really am enjoying what I'm doing." "Single guy… WFH is killing my mental health: I can't do this anymore. I don't even know where to begin looking for a non-remote job in accounting. I can't sit here in my house all day long anymore. Work is really the only place I have ever been able to make friends. I'm in a new city and haven't gotten close to anyone in 1.5 years. I'm fucking over it. I feel like I'm stuck between a rock and a hard place." "How to stop having a full panic attack every time your manager wants to have a meeting?"
Work-life balance	"WORK/LIFE BALANCE: I have been working from home for 5 months now. Although I love all the pros of WFH, I am finding it much harder to find a good work/life balance. I've noticed that I rarely leave the house. I feel trapped and actually a little depressed. I TRY to keep with a routine - walks, home work outs, etc - but find that after work my brain is completely spent and I have no energy to do anything other than sit on my couch and watch mindless TV. What are some things that have worked for you all?" "Moms are finding it tough to manage work-life balance amid Covid-19 crisis: ……." "Trouble with Work Life Balance: So I've been working (9 hours a day) remotely for a year and am having trouble sleeping. I try going to sleep around 11, however, I end up starting work at midnight because I feel like it's better to work on the next day's timecard instead of just lying awake in bed. Does anyone else have this problem? If so, what do you do?"
Team and company	"Team building while working from home.: Looking for advice on how to team build while working from home. We used to be in the office working together (team of about 10) and have lost the team feel since we all are now wfh. We work in healthcare, so a very stressed out group too. Thanks in advance for any advice." "WFH Tip: How to always appear active on Microsoft Teams: ……" "Struggling with the lack of supervision: Although I did some WFH over the last 2 years, I've just moved into a remote position and I'm struggling to adjust. It's a bit of a double whammy, it's my first full time WFH position, and also I'm basically a stand alone member of staff. I over-see a small team who are all fairly self sufficient, part time and WFH as well …... I find I'm getting really easily distracted because I don't have any deadlines or managers following up on me. This is half a rant/half looking for suggestions."

Table 1

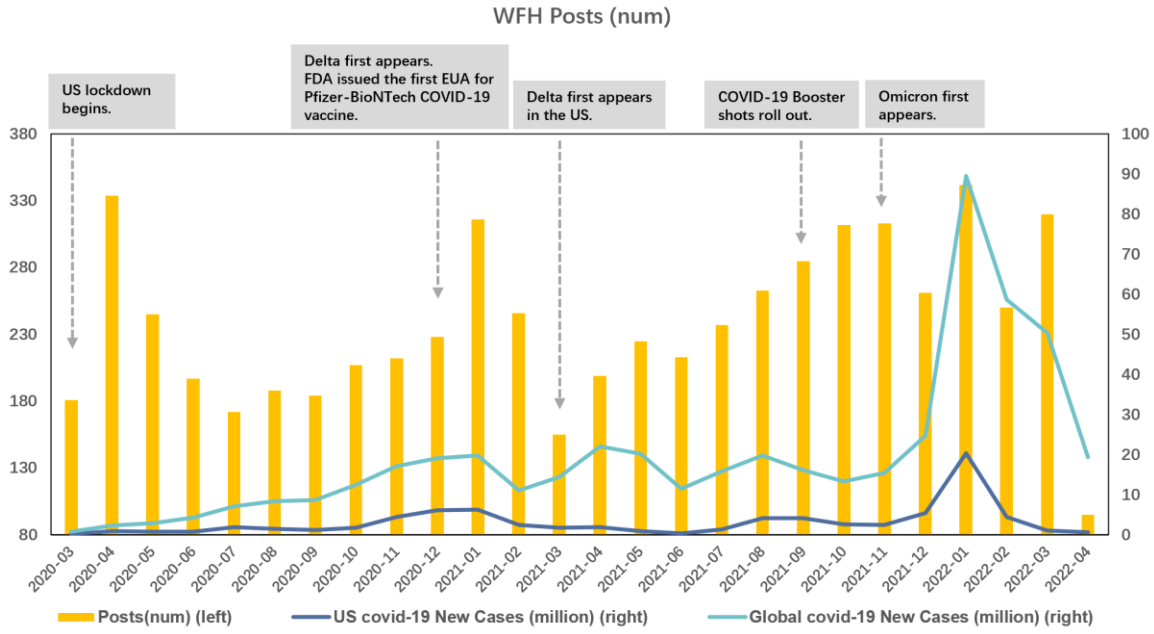


Figure 21

In this part, I focus on the temporal analysis. Figure 21 shows the monthly number of posts in Reddit's r/workfromhome community and the number of Global and the US Covid-19 cases from March 13, 2020 to April 12, 2022. As we can see from the graph, the number of posts follow some similar pattern with the number of Covid-19 new cases in the world. The correlation between Global Covid-19 new cases and the volume of posts is 0.41 and the correlation between percentage change of Global Covid-19 new cases and percentage change of the volume of posts is 0.51. April 2020, January 2021, January 2022 and March 2022 have the top 4 highest number of posts. Since March 2020, many companies try to shift into work from home policy as the pandemic became a nationwide emergency in the US and lockdown policy began in countries. The first peak explains people's discussion for firstly adapt to this new working style. The next two peaks happened when Covid-19 new cases reached the peak until that time period. In January 2021, the number of cases reached the first turning point since the vaccine have come out to the world and were massively promoted to public. There are lots of debate about working remotely topic at that time. In January 2022, Covid-19 new cases dramatically jumped to nearly 90 million as omicron widely spread. The number of posts in r/workfromhome community also reached the top. The recent peak is in March 2022, when the omicron got control and many states relieved their virus control policy such as wearing masks and no dine-in allowed.

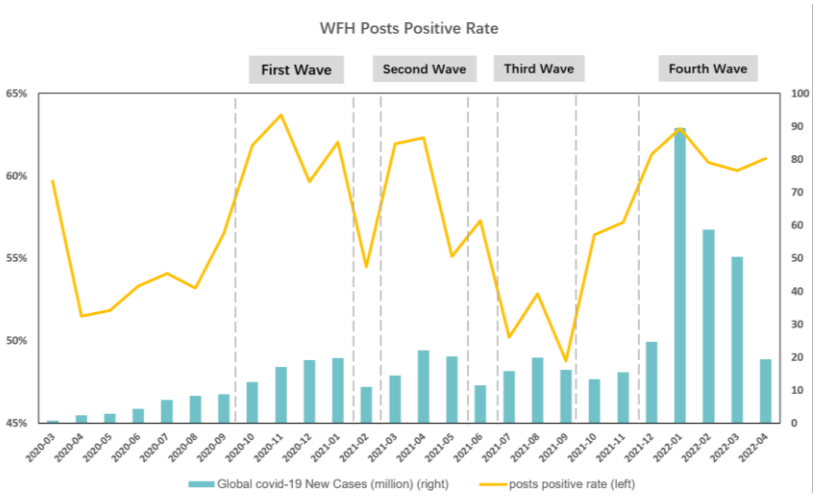


Figure 22

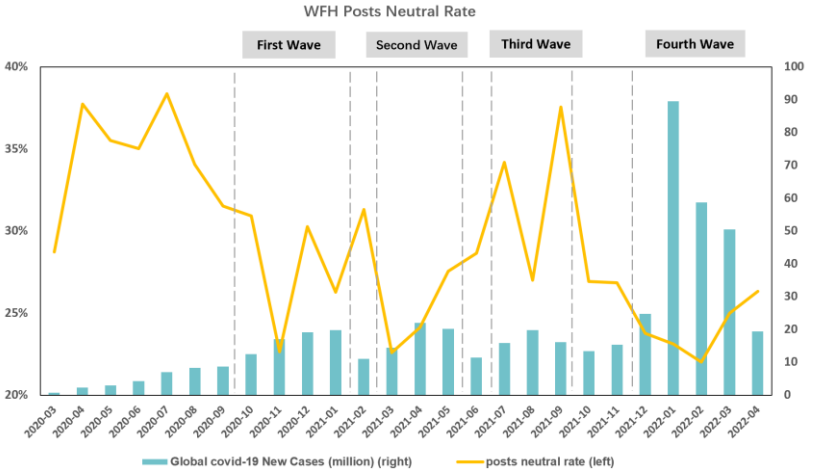


Figure 23

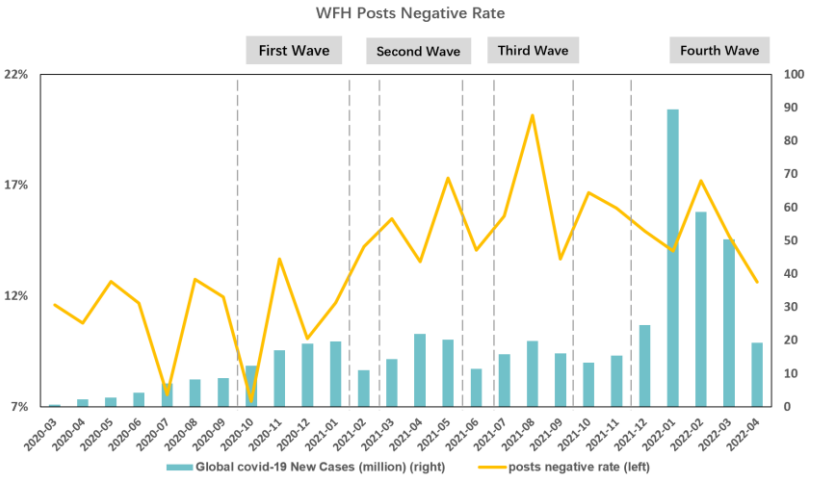


Figure 24

	Correlation (with Global Covid-19 new cases)	p-value
Number of posts	0.413	0.0349
Number of posts (% change)	0.507	0.0208
Posts positive rate	0.454	0.0139
Posts positive rate (% change)	0.149	0.4747
Posts neutral rate	-0.574	0.0009
Posts neutral rate (% change)	-0.095	0.6547
Posts negative rate	0.089	0.1235
Posts negative rate (% change)	-0.041	0.8507

Table 2

Then I do the sentiment analysis using the VADER classifier to find out the sentiment of working from home. For VADER classifier, I set the boundary of positive, neutral, and negative posts by the compound score. When the compound score is larger than 0.35, then the post is classified into positive post. If the score is between -0.1 and 0.35, the subreddit is treated as neutral post. In this case, the classification accuracy is relatively high compared with other boundaries I tried. The overall posts positive rate in the r/workfromhome community is 58%, while the neutral rate and negative rate are 14% and 28%. The correlation is showed in Figure 22-24 and Table 2. From the result of Pearson correlation coefficient, there is no correlation between sentiment of posts and the global Covid-19 new cases. Even the correlation between posts positive rate and global Covid-19 new cases is 45% and the p-value from permutation test is 0.03, but when applying the first order difference to these two data and calculating the correlation, the value is only 0.15 with 0.47 p-value. This means that the change of posts positive rate doesn't have a close relationship with the pattern of global Covid-19 new cases.

Using the Covid-19 cases as a benchmark may not very meaningful because though the Covid-19 cases increased dramatically since January 2022, most of the cases are omicron which causes more mild disease compared to prior variants. Thus, I mark the four waves of Covid-19 in the world by the peak of the confirmed new cases using the monthly data to reflect the timeline of it. The positive rate is high at the beginning of the pandemic, then the rate falls until the first wave of Covid-19. The posts positive rate is relatively high in the first, second and fourth waves of Covid-19. For these four periods when post positive rate is relatively high, I also perform the topic modeling to see whether the positive attitude within these different periods is credited to the same topic. The results show that the positive posts in the first period are mostly related with the excitement and freshness of this new working style, and some tips from those who have experience of telecommuting. The positive posts in the second period are more about covid-19, work-life balance, family members, and equipment. For the third and fourth period, the positive posts also include some team and company

topics. From the change of the topics during these four different periods, we know that the high posts positive rates are brought by different reasons.

The posts neutral rate, which has the similar issue with the posts positive rate, also cannot show the relationship with Covid-19 new cases. The data shows that at the beginning of the 2020, people are more neutral to telecommute compared with the following years. During first, second and fourth wave, the neutral rate drops to bottom. For the negative attitude toward working from home, posts negative rate is high in the third wave of the Covid-19. Further study could be focusing on why the third wave experiences the highest negative attitude about telecommuting.

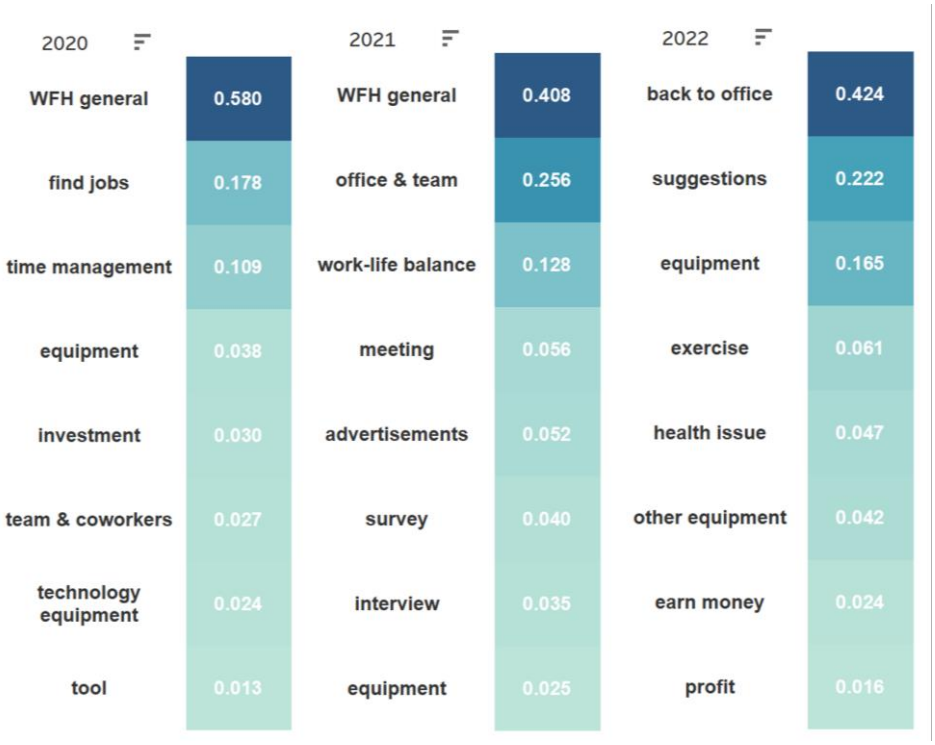


Figure 25

When evaluating the overall composition of topics in posts across time, the pattern is mostly the same with some distinct patterns. There are 2137 posts from March 13, 2020 to December 31, 2020, 3016 posts in 2021, and 1007 posts from January 1, 2022 to April 12, 2022. As the result shows in Figure 25, when I use this model on the posts of separate years, it is hard to detect all the possible topics since the top 1 topic contains about half of the tokens. Even though I change the number of topics, this situation doesn't get better. For the topics beyond the WFH general topic, people in 2020 talk more about finding jobs, time management, and equipment. In 2021, subreddits contain more about office and team, work-life balance, meeting, and interview. During the first four months of 2022, back to office becomes the major topic. Health problems and excises are also commonly discussed by people in this period. Technology and working equipment are the topics that is common during three years. The change of topics reflects the phenomenon that at first, people like to

discuss some urgent issue in terms of working from home. As people have some experience of the new working style, their focus is more diverse and personal.

Conclusion

In this paper, I first analysis the network structure of the r/workfromhome community on Reddit. The results show that the density of the sub-graph is negative correlated with the size of the sub-graph. Beyond that, the correlation matrix of the low-density and high-density network doesn't have significant difference. The values of correlation matrix are not related with the size of the community. For most of the sub-communities, the role of "answer-person" appears in the center of the graph. They are very important members to increase the activity and interaction in the Reddit community. From the word embedding of the user's posts and comments, I find out the "answer-person" is not restricted to reply to the posts from one or two topics. Also, the sub-community is not grouped by the topics. Using the word embedding of the sub-graph is not a way to understand and distinguish the different topics.

Then I perform the text analysis to test how have people's attitude changed over the course of pandemic. When conducting the synchronic analysis using three years' data together, the results show that the objective topics are relatively positive compared with these subjective topics. This implies that people on Reddit have some problems with those subjective topics such as work-life balance, physical and mental health, team and company, time management and company requirement. Then I study the temporal change of the attitude and focus over the course of pandemic. To prove my first claim that people in the WFH Reddit community will tend to be more positive when the Covid-19 is more serious, I apply sentiment analysis on the monthly posts and Covid-19 data. The results show that the positive rate of posts is high at the beginning of the pandemic as well as during the first, second and fourth wave of the Covid-19. The positive attitude in different time period is contributed by different benefits brought by working from home. To test my second claim that people's focus and concentration on work remotely along the process are varied, I perform topic modeling on the posts separately by year. The result is consistent with my hypothesis. At the beginning, people are intended to solve the urgent need and issues they face; as they gradually adapt to it, the topics are becoming more comprehensive and diverse such as talking more about work-life balance, health issues, attitude about back to office etc.

These findings could be beneficial to employers and policy makers to recognize that working from home is a complex social issue and then adjust the related policy. For further study, my study also offers a guide about the importance of the temporal analysis when analyzing telecommuting policy's influence.

Reference

Ashforth, Blake E., Glen E. Kreiner, and Mel Fugate. 2000. “All in a Day’S Work: Boundaries and Micro Role Transitions.” *Academy of Management Review* 25 (3): 472–91. https://doi.org/10.5465/amr.2000.3363315.

Barrero, Jose Maria, Nicholas Bloom, and Steven Davis. 2021. “Why Working from Home Will Stick.” w28731. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w28731.

Bartik, Alexander, Zoe Cullen, Edward Glaeser, Michael Luca, and Christopher Stanton. 2020. “What Jobs Are Being Done at Home During the Covid-19 Crisis? Evidence from Firm-Level Surveys.” w27422. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w27422.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. “Does Working from Home Work? Evidence from a Chinese Experiment*.” *The Quarterly Journal of Economics* 130 (1): 165–218. https://doi.org/10.1093/qje/qju032.

Bruque, Sebastián, José Moyano, and Jacob Eisenberg. 2008. “Individual Adaptation to IT-Induced Change: The Role of Social Networks.” *Journal of Management Information Systems* 25 (3): 177–206. https://doi.org/10.2753/MIS0742-1222250305.

Brynjolfsson, Erik, John Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe. 2020. “COVID-19 and Remote Work: An Early Look at US Data.” w27344. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w27344.

Buntain, Cody, and Jennifer Golbeck. 2014. “Identifying Social Roles in Reddit Using Network Structure.” In *Proceedings of the 23rd International Conference on World Wide Web*, 615–20. Seoul Korea: ACM. https://doi.org/10.1145/2567948.2579231.

Carillo, Kevin, Gaëlle Cachat-Rosset, Josianne Marsan, Tania Saba, and Alain Klarsfeld. 2021. “Adjusting to Epidemic-Induced Telework: Empirical Insights from Teleworkers in France.” *European Journal of Information Systems* 30 (1): 69–88. https://doi.org/10.1080/0960085X.2020.1829512.

Chen, Pin-Yu, and Alfred O. Hero. 2014. “Local Fiedler Vector Centrality for Detection of Deep and Overlapping Communities in Networks.” In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1120–24. Florence, Italy: IEEE. https://doi.org/10.1109/ICASSP.2014.6853771.

Fay, Martha J., and Susan L. Kline. 2011. “Coworker Relationships and Informal Communication in High-Intensity Telecommuting.” *Journal of Applied Communication Research* 39 (2): 144–63. https://doi.org/10.1080/00909882.2011.556136.

Fiedler, Miroslav. 1973. “Algebraic Connectivity of Graphs.” *Czechoslovak Mathematical Journal* 23 (2): 298–305. https://doi.org/10.21136/CMJ.1973.101168.

Gibbs, Michael, Friederike Mengel, and Christoph Siemroth. 2021. “Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3843197.

Gottipati, Swapna, Kyong Jin Shim, Hui Hian Teo, Karthik Nityanand, and Shreyansh Shivam. 2021. “Analyzing Tweets on New Norm: Work from Home during COVID-19 Outbreak.” In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, 0500–0507. NV, USA: IEEE. https://doi.org/10.1109/CCWC51732.2021.9375936.

Niessen, Cornelia, Christine Swarowsky, and Markus Leiz. 2010. “Age and Adaptation to Changes in the Workplace.” *Journal of Managerial Psychology* 25 (4): 356–83. https://doi.org/10.1108/02683941011035287.

Sias, Patricia M., Hannah Pedersen, Erin B. Gallagher, and Irina Kopaneva. 2012. “Workplace Friendship in the Electronically Connected Organization.” *Human Communication Research* 38 (3): 253–79. https://doi.org/10.1111/j.1468-2958.2012.01428.x.

Zhang, Charlene, Martin C. Yu, and Sebastian Marin. 2021. “Exploring Public Sentiment on Enforced Remote Work during COVID-19.” *Journal of Applied Psychology* 106 (6): 797–810. https://doi.org/10.1037/apl0000933.