



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

Yujing Sun

Motivation

- About half of people work from home in May, 2020.
- Work from home: Efficiency, work-life boundary, mental health issue, co-worker relationship
- Previous study: representative problem(survey), synchronic data(short time)
- Research question:
 - **Network analysis:** Whether the sub-communities within Reddit WFH community are clustered by different topics?
 - **Text analysis:** How have people's attitude changed about WFH on Reddit over the course of pandemic?

Reference: Brynjolfsson et al. 2020; Bloom et al. 2015; Gibbs, Mengel, and Siemroth 2021; Ashforth, Kreiner, and Fugate 2000; Carillo et al. 2021; Weinert et al. 2015; Fay and Kline 2011; Sias et al. 2012

Data and Method

I scraped the data from Reddit r/workfromhome community using Pushshift API

Network analysis:

- Posts within 2021.
- Comments with more than 20 words.
- Exclude the comments by authors itself.

Method:

- Community detection: Using second eigen value (Fiedler vector)
- Post's similarity: Word2Vec

Post's link	Author's name	Posts
Post's link	Commenter's name	Comments

Text analysis:

- Posts between March 13, 2020 to April 12, 2022.
- 6,180 subreddits in r/workfromhome community on Reddit.

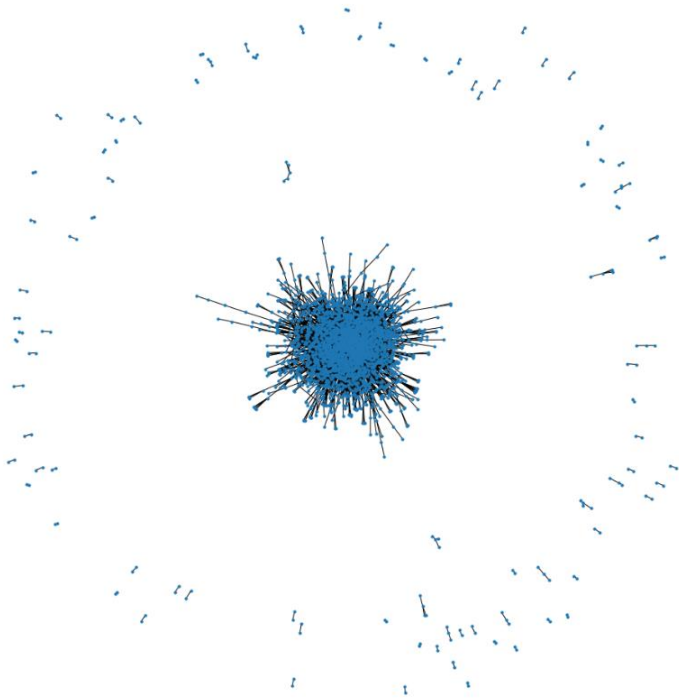
Method:

- Sentiment analysis: VADER
- Topic modeling: LDA

Posts	Post's time
-------	-------------

Network

Original graph

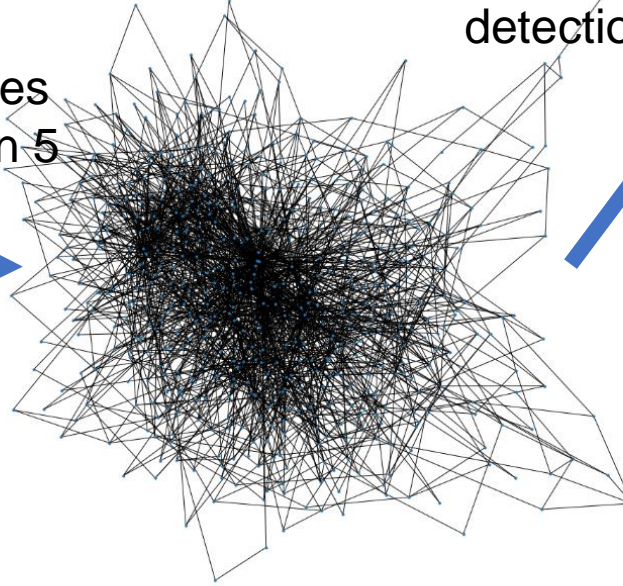


2745 nodes

drop the nodes
with less than 5
edges

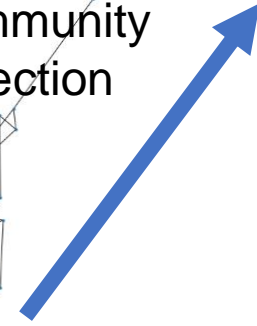


New graph

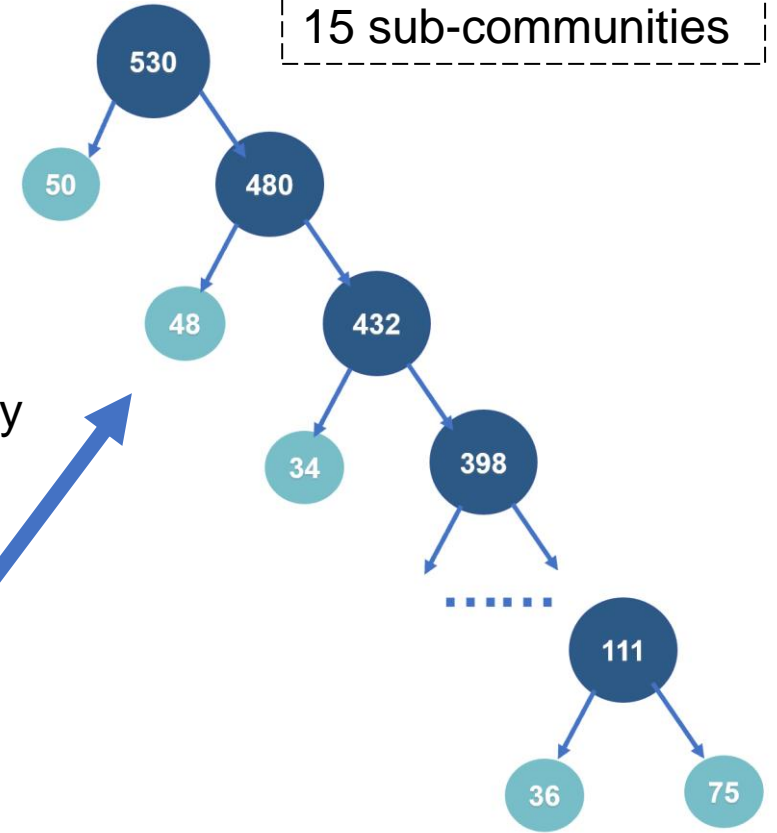


530 nodes

community
detection



15 sub-communities



Size of sub-communities:
Min: 11 nodes
Mean: 35 nodes
Max: 75 nodes

Network

Figure 1: Correlation matrix of community NO.6

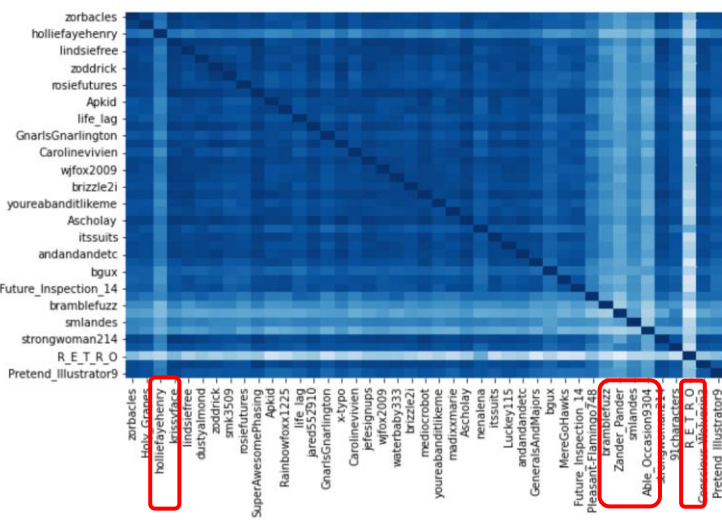


Figure 2: Correlation matrix of community NO.2

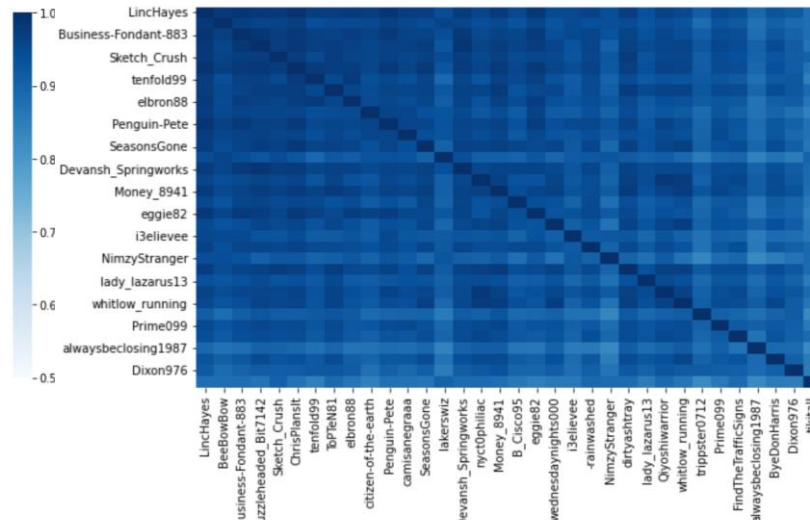


Figure 3: Correlation matrix of 30 random nodes

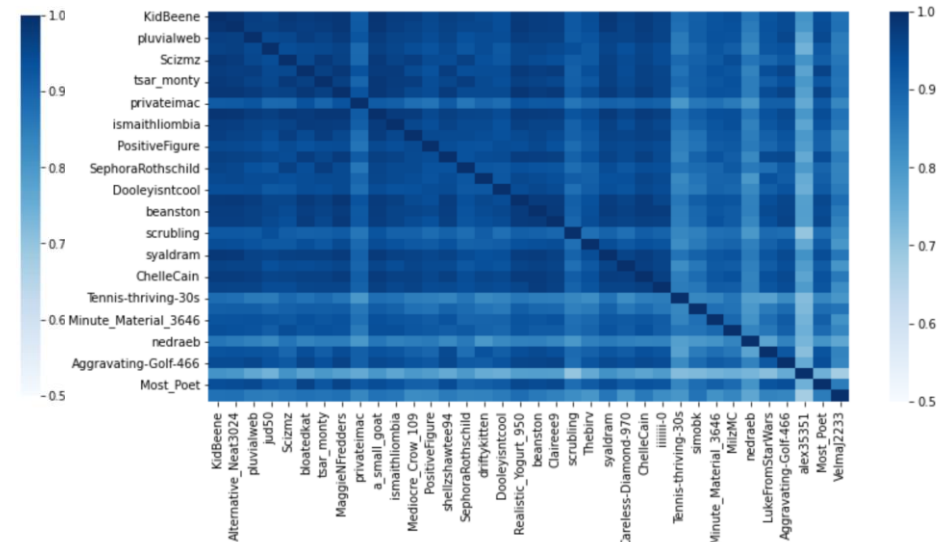
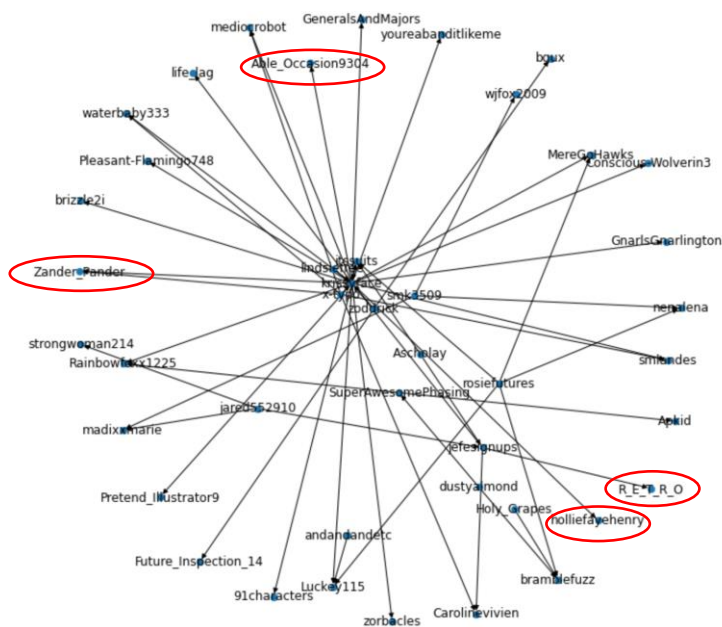


Figure 4: Graph of community NO.6



Method:

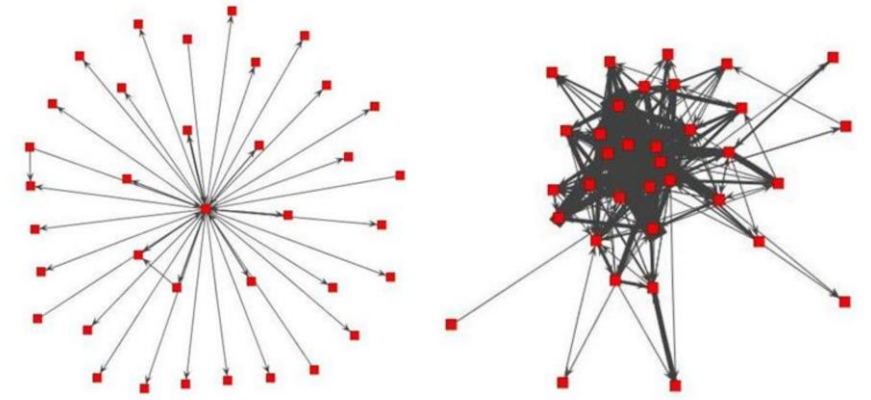
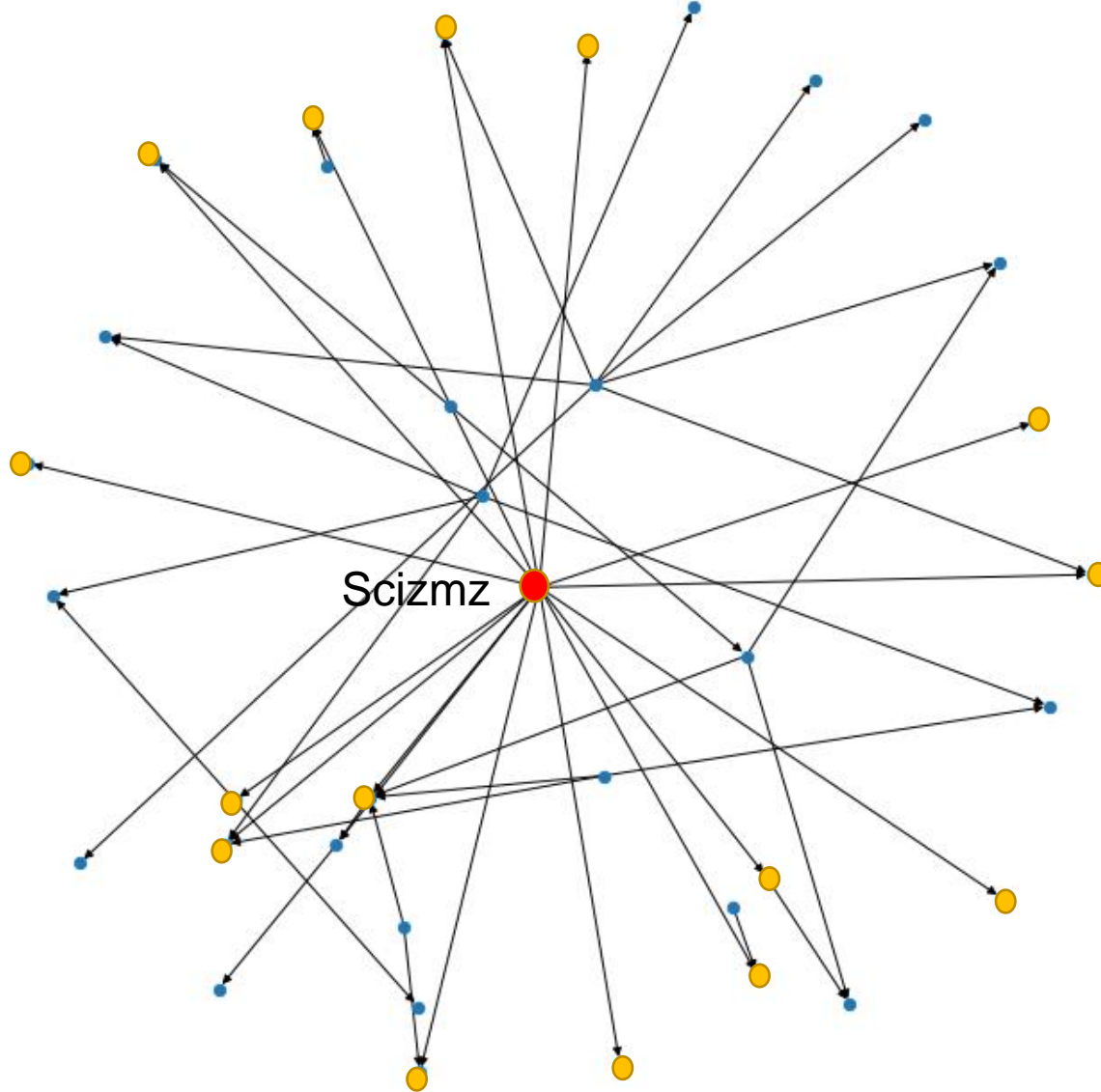
- For **each user**, I collect all his/her posts and comments and aggregate them together.
- I calculate the **correlation matrix** of average word embedding vector for each user's text within the small community.

Findings:

- The values of correlation matrix are not related with the size of the community. (Fig 1, 2)
- For those who distinctly have lower correlation among all other users within the sub-group, the position of these users are always at the periphery of the graph. (Fig 1, 4)
- Sub-network is not grouped by topics. (Fig 3)

Network

Graph of community NO.13



(a) Answer Person

(b) Discussion Person

Buntain, Cody, and Jennifer Golbeck 2014.

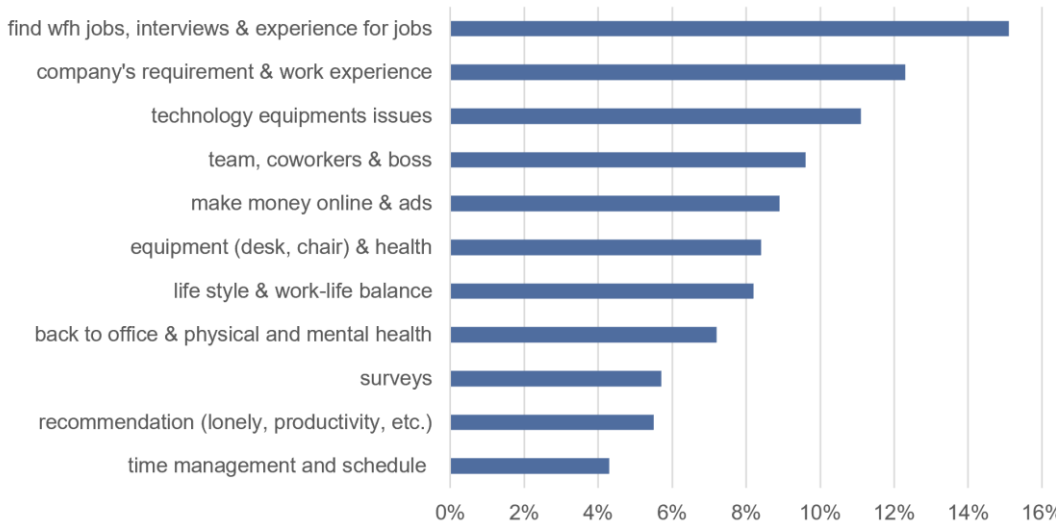
Findings:

- The majority users in the center of the graph is the “answer-person”.
- User “Scizmz” answers 16 questions from 5 different topics. (topic modeling/ manually)
- The correlation of word embedding for the 16 questions are very high.
- Again, word embedding of sub-network is hard to distinguish the different topics.

Text analysis

Synchronic analysis

Top 11 topics (% of tokens)



Sentiment Analysis of Different Topics (WFH)



- Objective topics such as “recommendation”, “survey”, “equipment”, “make money online” and “advertisements” are relatively more 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.

Text analysis

Temporal analysis -- change of post's num and sentiment

Figure 1 WFH Posts (num)

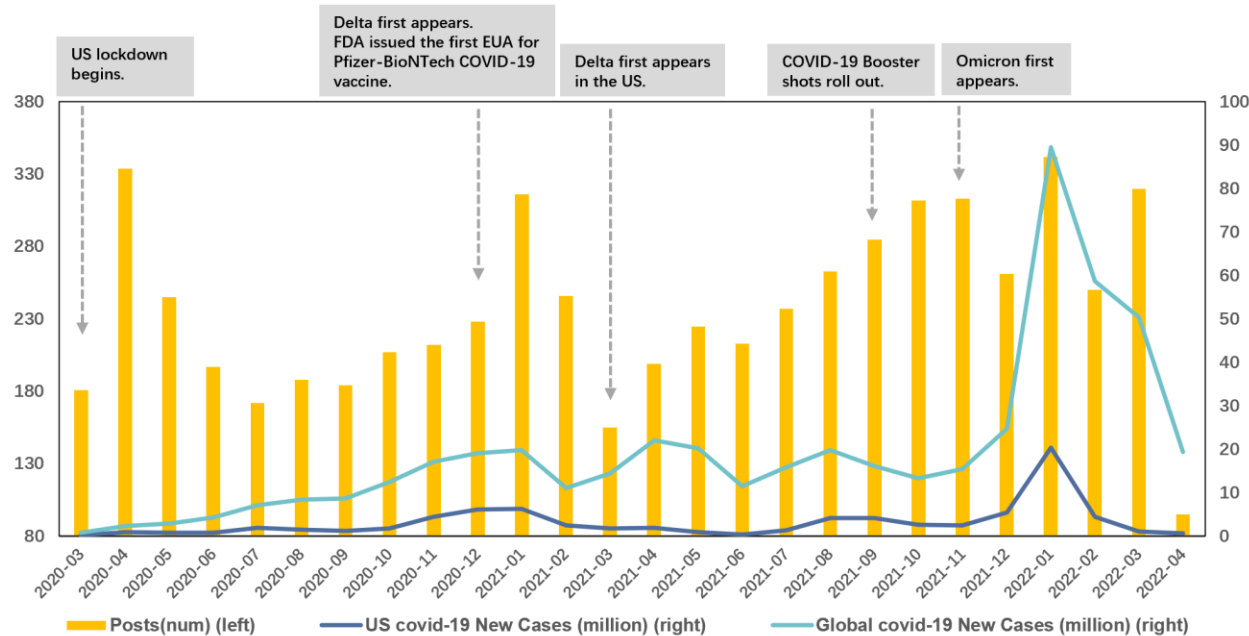


Figure 2 WFH Posts Positive Rate

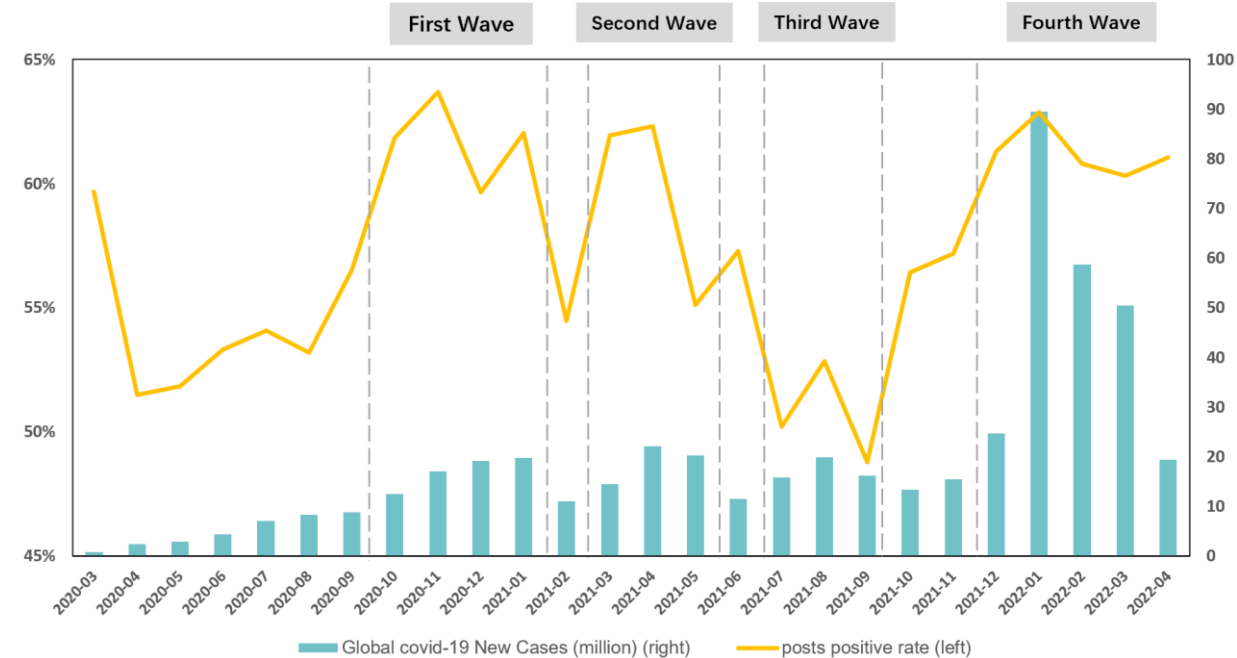


Table 1

	Correlation (with Global Covid-19 new cases)	p-value
Number of posts	0.413	0.0349
Number of posts (% change)	0.507	0.0208

Finding:

- The number of posts follows some similar patterns with the number of Covid-19 new cases in the world. (Fig 1, Table 1)
- The change of sentiment doesn't have a close relationship with the pattern of new covid-19 cases. (In my paper)
- The posts positive rate is high at the beginning of pandemic, and in the first, second and fourth waves of Covid-19. (Fig 2)

Text analysis

Temporal analysis – change of topics



- 2020: finding jobs, time management, and equipment.
- 2021: office and team, work-life balance, meeting, and interview.
- 2022: back to office, equipment, health issue.

Conclusion

- **Network analysis:**

- Sub-communities within Reddit WFH community are **not** clustered by different topics.
- Sub-communities within Reddit WFH community have “**answer-person**” **structure**.

- **Text analysis:**

- The difference of topics reflects people’s different reaction and focus as they get in touch with and get used to this new norm.
- Analyzing employee’s attitude toward WFH is **a dynamic process**.

- **Implication:**

- For future study, only focusing on the synchronic influence may not be that persuasive.
- The influences of WFH are conditional on the quality of WFH policy, and how long people have adopted it.