

Changes in Topics During COVID-19 Within The Korean Sociology Studies

Korea University, Department of Sociology

BACKGROUND

- Studying the study is one of the important parts of sociological research. This is because it is one way to know where sociology researchers are interested.
- According to Kang & Kwon (2021, p. 196)[1], unlike the natural sciences or the humanities, the number of research papers and collaborators did not decrease during the COVID-19 period in sociology research. However, this does not mean that sociology researchers have not been affected by COVID-19.
- In this study, we will explore how COVID-19 has affected sociological research.

DATA SOURCES

- To obtain research subject data, I downloaded the metadata of KCI-listed papers which is published between 2018 and 2021 and tagged with "Sociology" and "Social Science in General". from the Korea Citation Index (KCI) webpage. A KCI-listed paper means a paper that is listed in an academic journal with a score of 85 or higher in the KCI review. In order to analyze using the date and keyword data of these papers, I analyzed the Korean keywords registered by the researchers.
- To check the progress of COVID-19, I used country-specific COVID-19 status data provided by the World Health Organization (WHO). Since the subject of the study is Korean papers, I used data on the number of confirmed cases of COVID-19 in Korea.

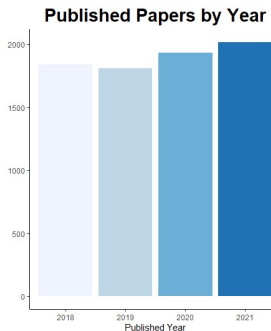


Figure 1. Published Papers by Year

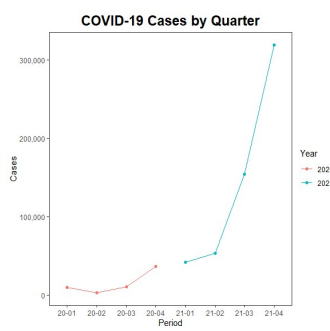


Figure 2. Confirmed Cases of COVID-19 In Korea, By Quarter

METHOD

- Data preprocessing : Although most of the keywords were in Korean, non-Korean languages were mixed, so I translated them into Korean using the Google Translation API.
- After that, using the NIA Korean Dictionary, I extracted only the nouns from the keywords.
- For the extracted words, I extracted topics using Latent Dirichlet Allocation (LDA). This methodology was inspired by Seol et al. (2018, pp. 199-200)[2].

- LDA Hyperparameters : $\alpha=.01$, $\eta=.001$, Burn-in=500, Iteration=5,000
- I set the number of topic as 7(Figure 3).

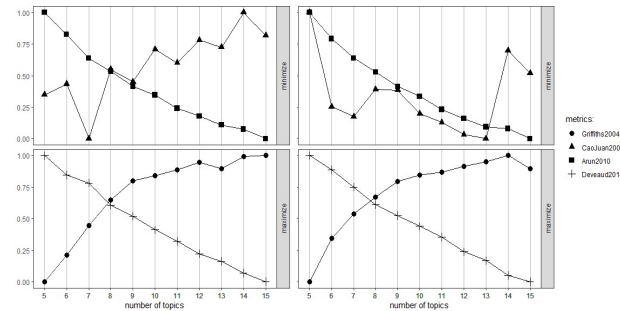


Figure 3. Optimizing Parameters of The Number of Topics

PRE-COVID & POST-COVID

- Extracted Topics of Whole Years : Life Cycle and Psychology and Education, Welfare and Policy, Minority, Industry, Business and Economics, Civil Politics and Ideology , Diplomacy and Security
- Changes Between Pre-COVID(2018-2019) and Post-COVID(2020-2021) :
 - Decreased interest in business, economy, and labor.
 - Increased interest in the individual(life cycle and psychology).
 - Decreased interest in the abstract(grand) theory : Civil politics and ideology.
 - Increased interest in "real" politics and diplomacy, national security.
 - Changed trend in education : Education is more associated with minorities than with occupations

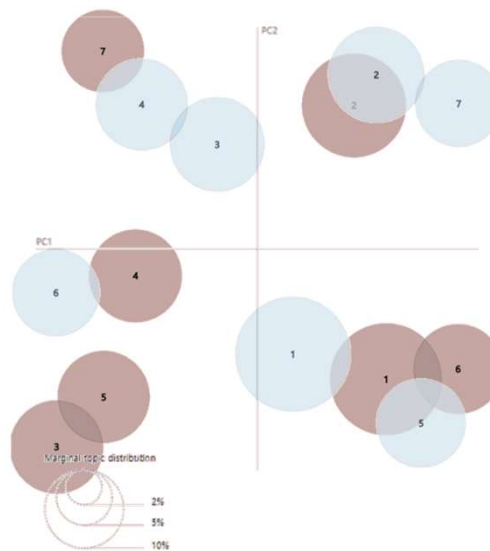


Figure 4. Topics of Pre-COVID(Red) and Post-COVID(Blue), In each year, the topic number is different.

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7
정치	운동	교육	협력	교사	소비	소득
평화	정치	이주	개발	양육	금융	노인
안보	주의	여성	분석	부모	투자	노동
테러	전환	결혼	뉴스	교육	감사	돌봄
주의	시민	다문화	디지털	자기	소비자	보험

Table 1. Topics and Main Words of 2018 to 2021

DURING COVID-19: 2021

- Since the number of confirmed cases of COVID-19 has increased exponentially, I saw a change in 2021, when the number of confirmed cases of COVID-19 surged during the study period. According to Kang & Kwon (2021, p. 196)[1], considering the publication cycle of the paper, the unit of analysis time should be 3 months. In addition, the first half of 2021 and the second half of 2021 show a sharp difference in the number of confirmed cases(Figure 2). Therefore, I analyzed based on the two quarters (6 months), which is a multiple of 3 months.
- As a result, between the first half of 2021 and the second half of 2021 during the COVID-19 period, Topic 2 and Topic 7 showed a slight increase. However, at the same time, Topic 4 also increased, which is contrast to the previous analysis. Therefore, different topics may have received attention in the outbreak and progression of COVID-19.

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7
금융	자기	건강	산업	정치	서비스	교육
협력	만족	이주	가치	시민	관리	청소년
개발	생활	여성	경제	주의	돌봄	장애인
이용	우울	장기	사회	분석	여성	복한
미디어	스트레스	의료	노동	이론	감사	공동체

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7
평가	자기	종교	소비	주의	돌봄	교육
주민	조직	주의	경제	사회	이주	정책
혁신	교육	폭력	공간	신뢰	여성	기반
정보	직무	전환	보험	운동	자녀	코로나19
패널	스트레스	여성	지역	평화	이용	거버넌스

Table 2. Topics and Main Words of 2021-1 and 2021-2. Reordered Similar Topics for Comparison.

CONCLUSIONS

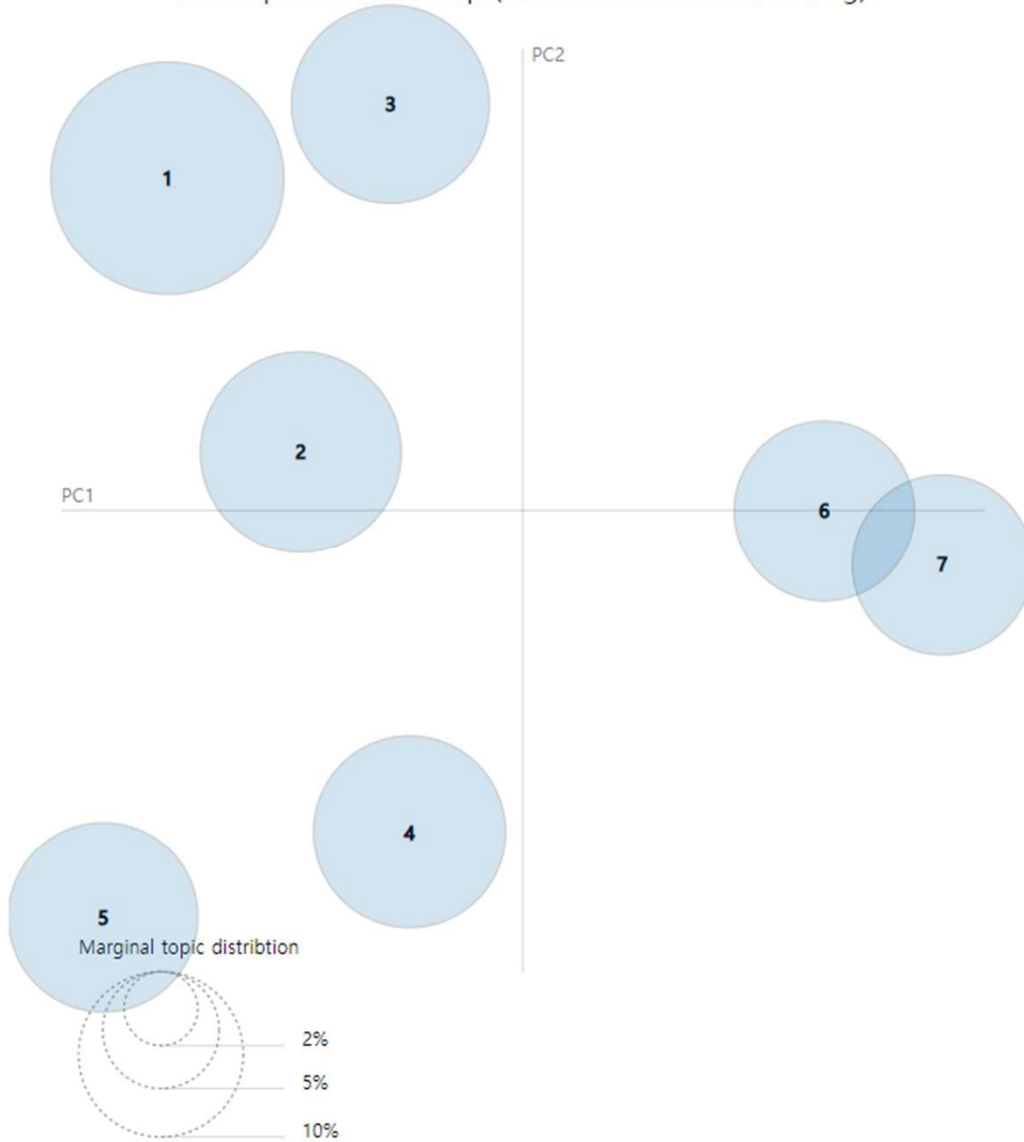
- Due to COVID-19, sociology researchers in South Korea have become more interested in personal life and psychology, real politics and security than in economics and grand theories. In summary, **after COVID-19, sociology researchers in Korea have begun to deal with more realistic issues and issues directly related to security.**
- However, different aspects may appear depending on the progress of COVID-19, which is an area that needs further research. In addition, since the change in the research topic cannot be attributed to COVID-19 alone, a longer-range study is needed.

References

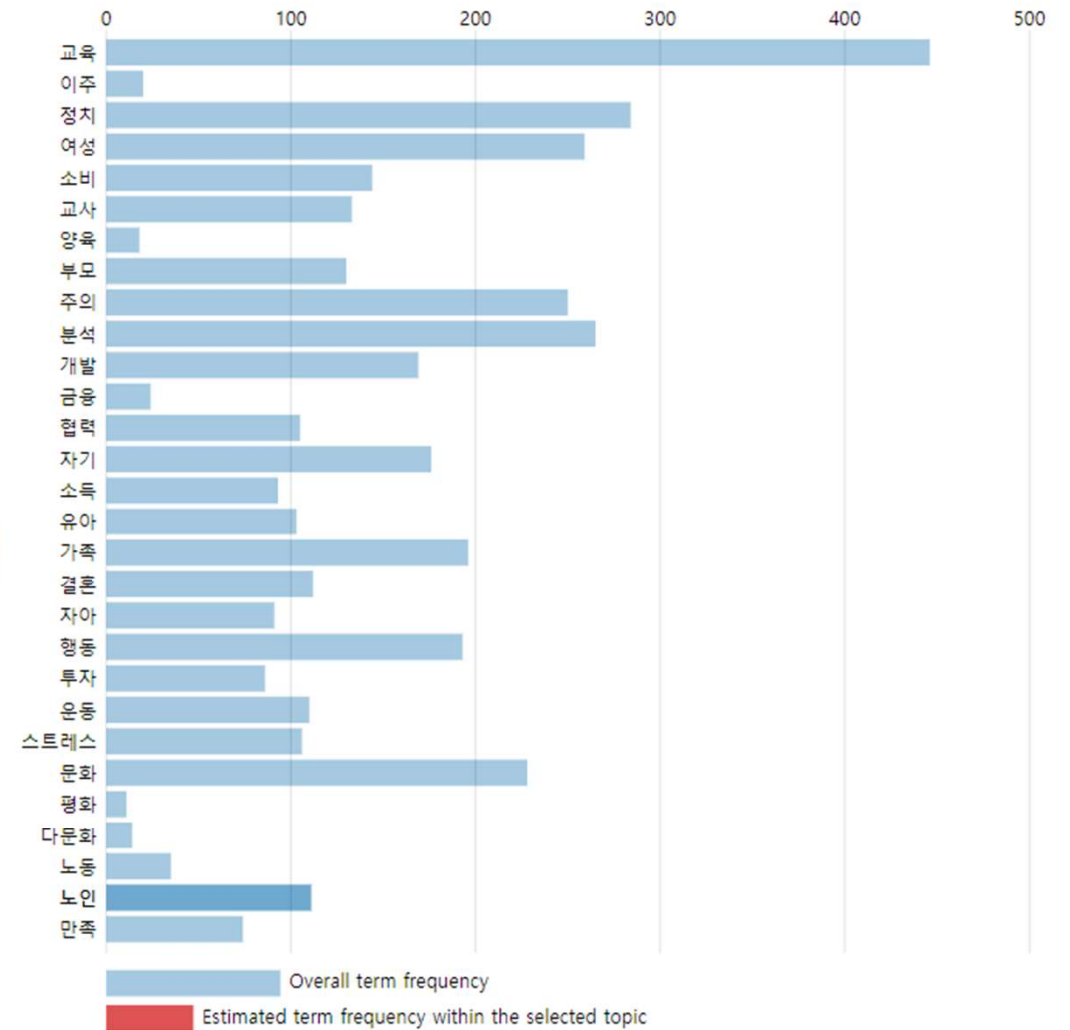
- [1]Kang, J. H. & Kwon, E. R. (2021). The Effect of COVID-19 Pandemic on Research Productivity in South Korea: A Comparative Analysis of Korean Journal Articles across Academic Fields. Korean Journal of Sociology, 55(1), 179-199.
- [2]Seol, D. H., Ko, J. H. & Yoo, S. H. (2018). Korean Sociological Association and Sociological Research: Changes in the Areas of Sociology in Korea 1964-2017. Korean Journal of Sociology, 52(1), 153-213.

LDA Result of Whole Year, $\lambda=.6$

Intertopic Distance Map (via multidimensional scaling)



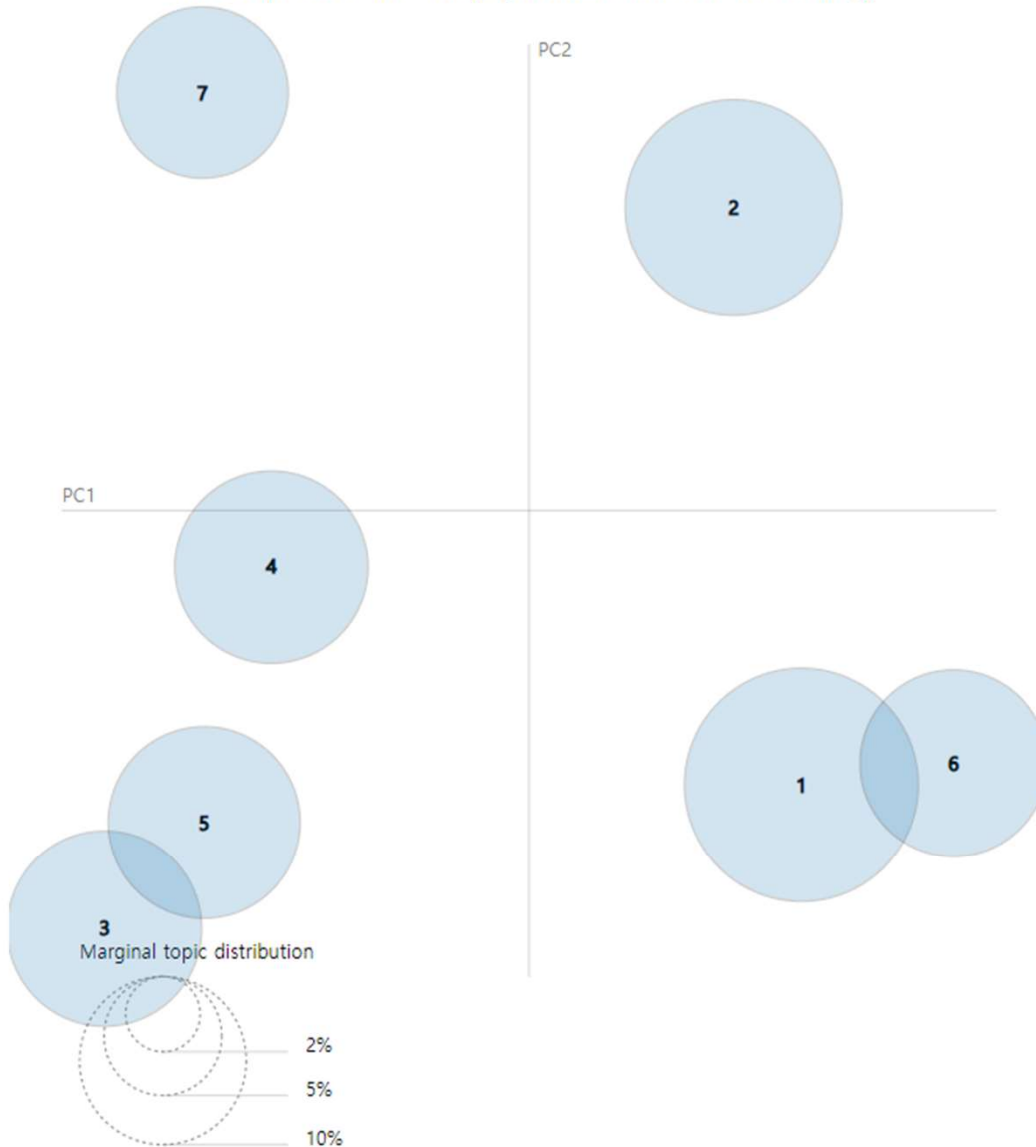
Top-30 Most Salient Terms¹



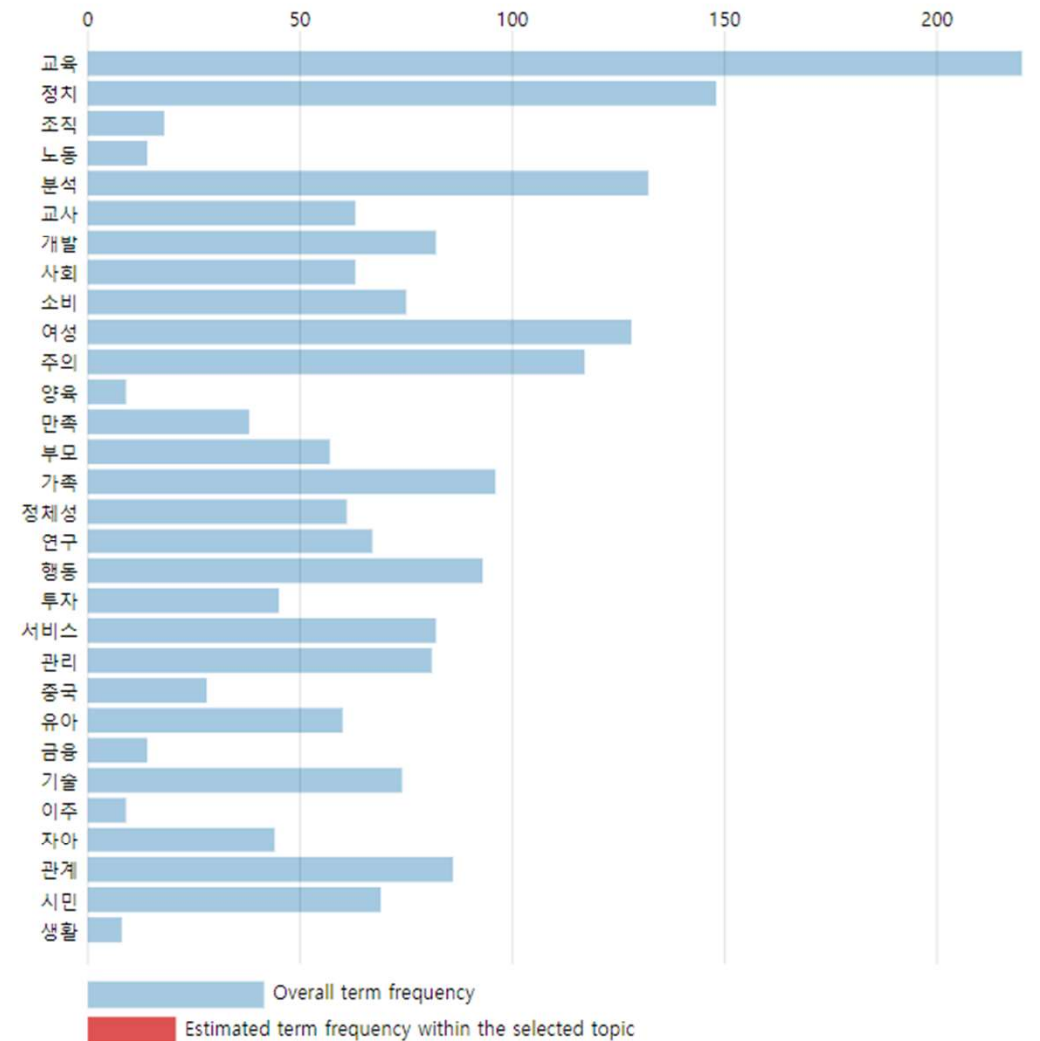
1. saliency(term w) = frequency(w) * $[\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al (2012)
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

LDA Result of Pre-COVID, $\lambda=.6$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Salient Terms¹

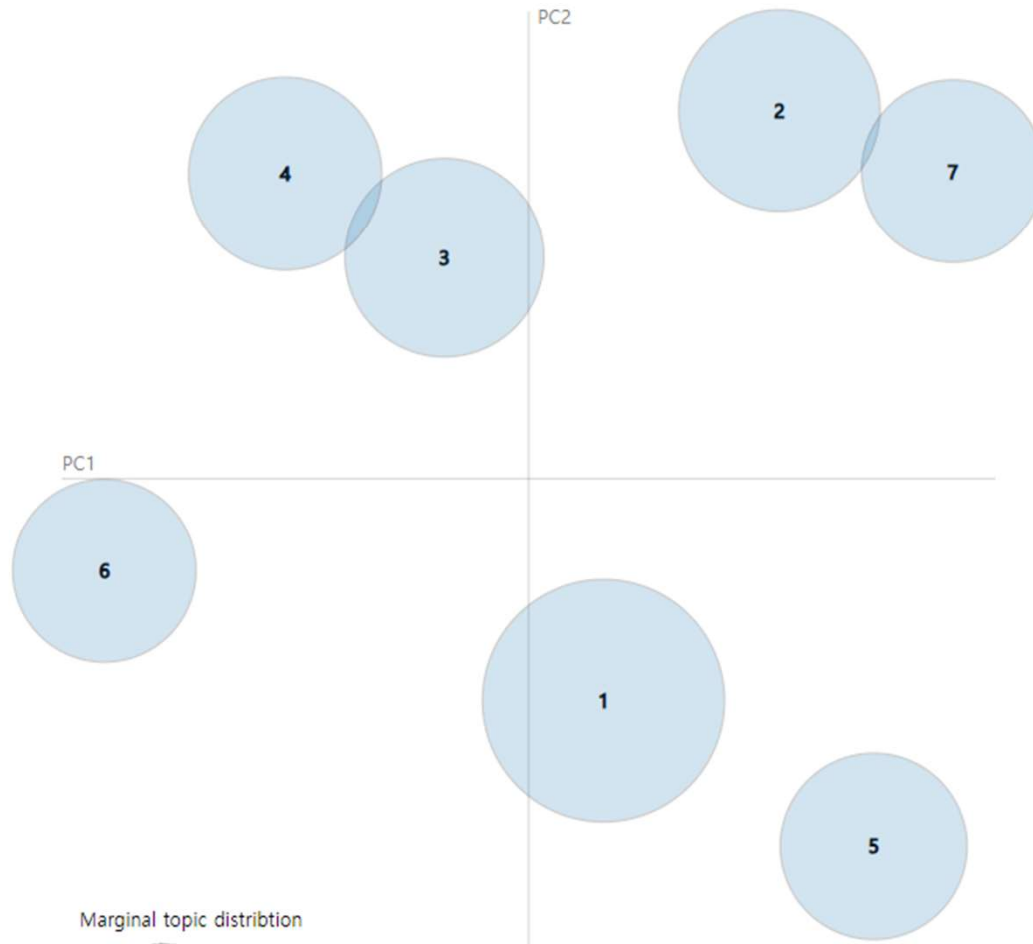


1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

LDA Result of Post-COVID, $\lambda=.6$

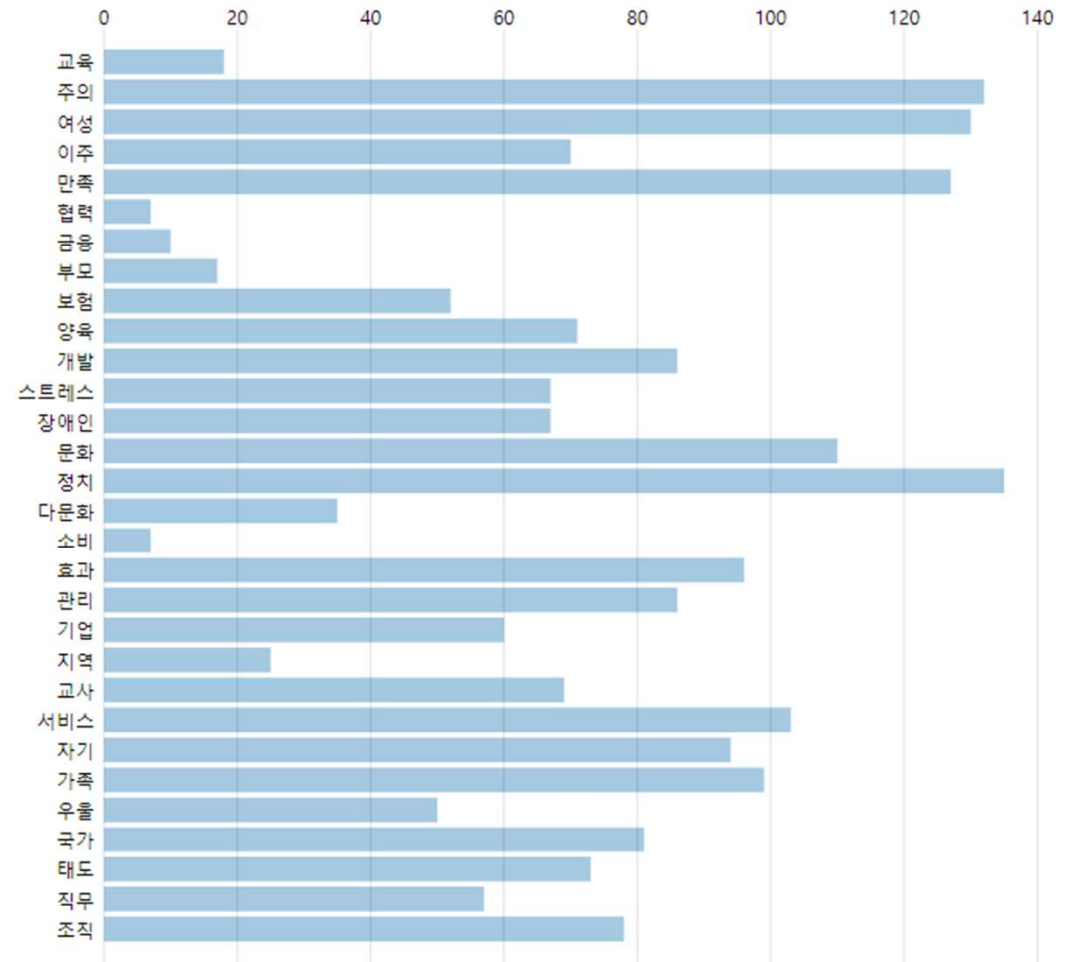
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Salient Terms¹



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t ; see Chuang et. al (2012)

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)