INFORMS Annual Meeting 2020

Understanding the Research Collaborations During COVID-19 Pandemic

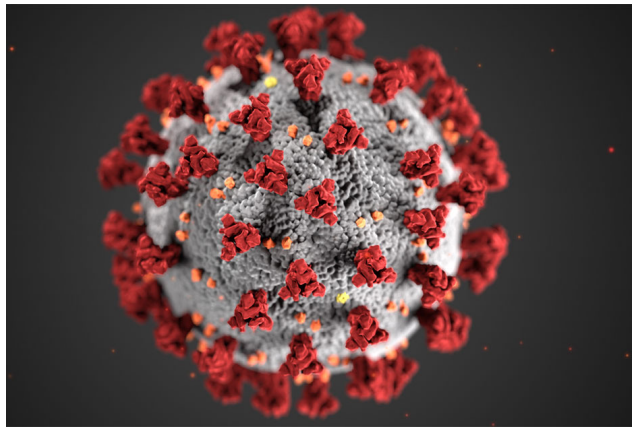
Sulyun Lee, Kang Zhao, Ning Li

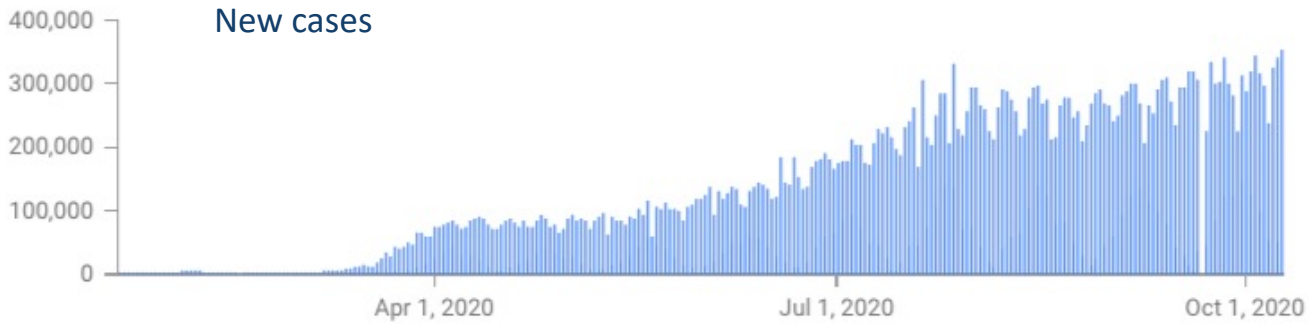
Presenter: Sulyun Lee

Introduction

COVID-19 Pandemic

- COVID-19 was first identified in Wuhan, China in December, 2019.
- 214 countries and territories around the world have reported cases.
- Total of 37,964,503 confirmed cases around the world (up to date).
- Still affecting the world.





Introduction

Global research on COVID-19

- COVID-19 has attracted global researchers' attentions.
- Many articles are released around the world from different fields.
- Example: medication, biology, psychiatric, social, economic, etc.

Introduction

Research goals

- Goal: Observe how researchers collaborate each other during COVID-19 pandemic.

- Research Questions:
  1. What factors contribute to the academic recognition of COVID-19 research papers?
  2. What factors contribute to the public recognition of COVID-19 research papers?
  3. What diversity can we observe from the collaboration of COVID-19 research papers?

Introduction

Research goals

- How can we measure the academic and public recognition of research papers?

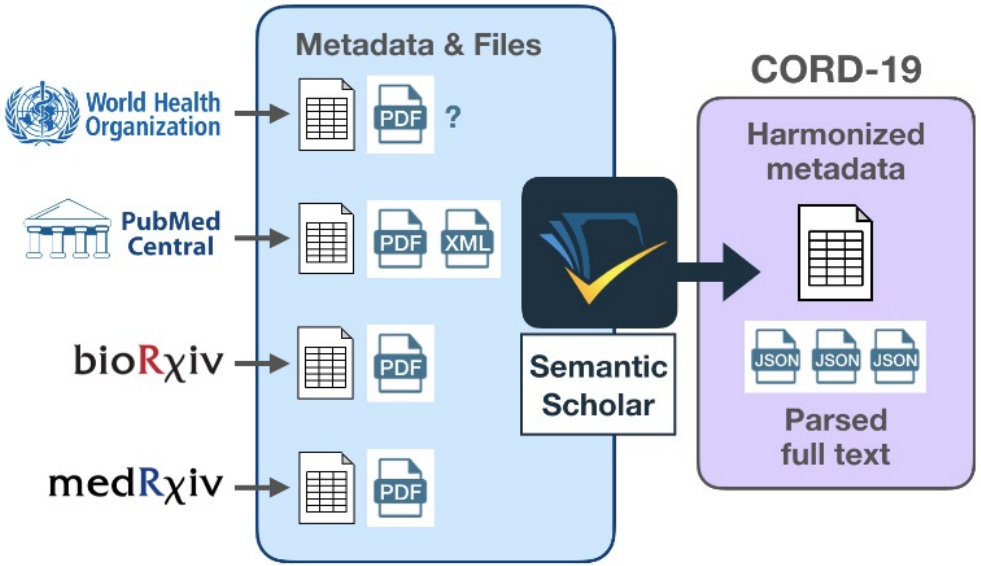
- Academic recognition: measured by citation counts

- Public recognition: measured by PlumX metrics
  - PlumX metrics are the metrics that measure the scholarly research output with 5 categories.
  - “Mentions” category measures how people are engaged with the research through news articles or blog posts.
  - Example:
    - Number of blog posts written about the article
    - Number of comments made in Reddit, Youtube, etc. about the article
    - Number of gists in the GitHup repository
    - Number of news articles about the article
    - Number of references in Wikipedia

Data and Setup

Data Source: CORD-19

- COVID-19 Open Research Dataset Challenge (CORD-19)
- Provided by the White House and several leading research groups.
- Contains articles about COVID-19 and related historical coronaviruses from different sources (PMC, WHO, bioRxiv, medRxiv).



Source: Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., ... & Mooney, P. (2020). CORD-19: The Covid-19 Open Research Dataset. *ArXiv*.

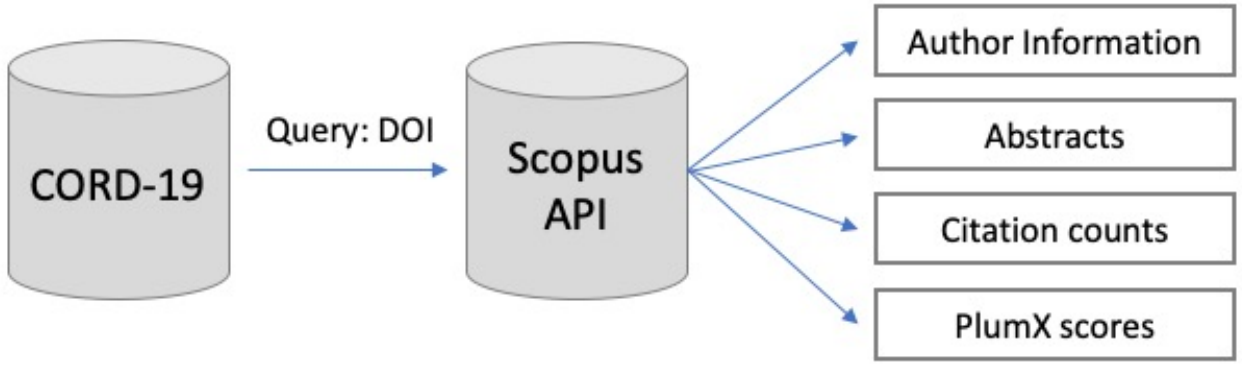
Data and Setup

Data Source: Scopus

- Scopus: Elsevier's large database of scholarly content

Scopus®

- Retrieved the author information, abstracts, citation counts, and PlumX scores



Data and Setup

Setup

- Used articles published between Jan. 01, 2020 and Sep. 03, 2020.
- Excluded articles with zero or one author.
- Excluded preprints and news/commentary articles.
- Gathered 32,281 COVID-19 related papers for analysis.

Methods

Modeling design

- Independent Variables
  - Diversity of authors' previous research topics
  - Rate of new coauthorships
  - Diversity of authors' h-index
  - Cultural diversity of authors

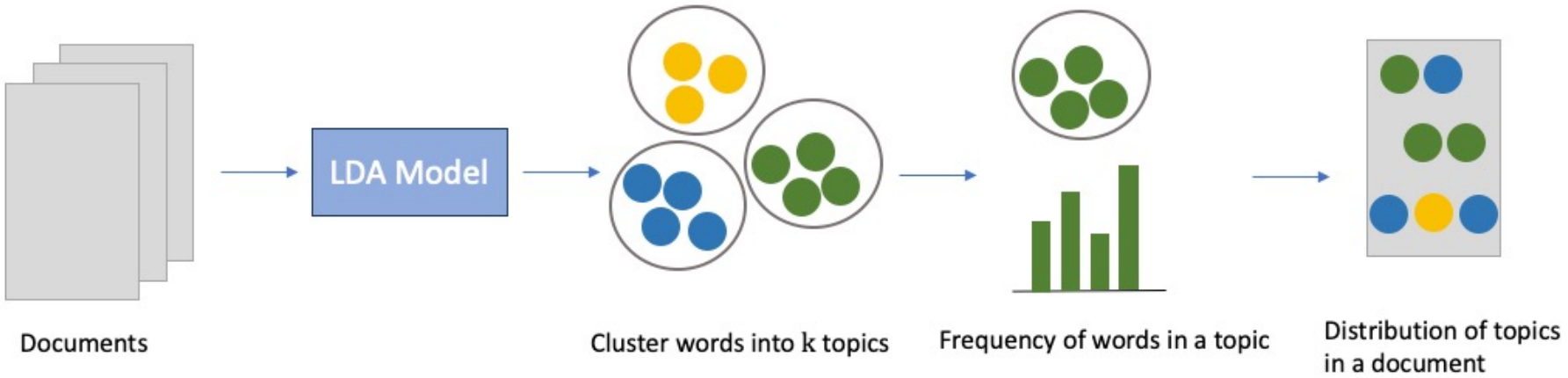
- Control Variables
  - Maximum h-index of authors
  - Research topics of COVID-19 papers
  - Time since the publication dates
  - Rate of coauthors in the practical affiliations, such as hospitals and clinics
  - Team size

- Dependent Variables
  - Citation counts
  - Number of mentions in PlumX metrics

Methods

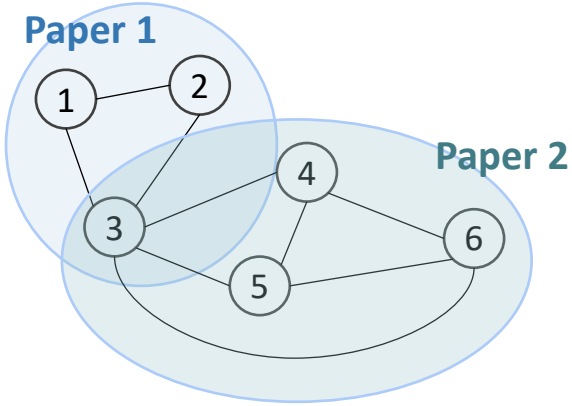
Independent variables

- Diversity of authors' previous research topics
  - Discover how authors from different fields form research collaborations.
  - Used LDA topic modeling on titles and abstracts of authors' previous publications.
  - 30 topics with the highest coherence score.



Methods

Independent variables



- Rate of new coauthorships
  - Discover if new collaborations were made in the multi-author papers.
  - Considered the collaborations as the fully connected sub-network.
  - Computed the rate of newly formed edges in the COVID-19 collaboration network.

- Diversity of authors' h-index
  - Discover the disparity of authors' h-index in a paper.
  - Used gini coefficients to estimate the disparity.

- Cultural diversity of authors
  - Discover the diversity of authors' cultural background.
  - Based on the authors' affiliation countries.
  - Computed the similarity of authors' affiliation countries using Hofstede's culture scores.

Source: Hofstede, G., & Bond, M. H. (1984). Hofstede's culture dimensions: An independent validation using Rokeach's value survey. *Journal of cross-cultural psychology*, *15*(4), 417-433.

Methods

Control variables

- Maximum h-index of authors
  - Control the effects of a prestiged author with highesth-index
- Research topics of COVID-19 papers
  - Control the effects of specific research fields gaining more attentions.
  - Used the LDA topic modeling on COVID-19 papers' titles and abstracts
  - 20 topics with the highest coherence score.
- Time since the publication dates
  - Papers with longer publication time might have gained more academic and public recognitions.
  - Computed the number of days.

Methods

Control variables

- Rate of coauthors in the practical affiliations
  - Control the effects of authors' affiliation types.
  - Computed the rate of authors from hospitals and clinics.

- Team size
  - Control the effects of collaborator size.
  - Computed the total number of authors in a paper.

Source: Hofstede, G., & Bond, M. H. (1984). Hofstede's culture dimensions: An independent validation using Rokeach's value survey. *Journal of cross-cultural psychology*, *15*(4), 417-433.

Methods

Modeling

- Used linear regression model.
- Log transformed the highly skewed variables
  - Maximum h-index and team size
- Removed the variables with high variance inflation factor (VIF) for multi-collinearity.
- Two separate models for citation counts and PlumX mentions.

Results

Regression coefficients

1) Citation counts

	Coefficient	t -statistic	P-value
Intercept	**0.6522**	**120.875**	**< 0.001**
Similarity of authors' previous publications	**-0.089**	**-14.449**	**< 0.001**
Rate of new coauthorships	**0.085**	**12.107**	**< 0.001**
Diversity of authors' h-index	**-0.082**	**-11.677**	**< 0.001**
Cultural similarity of authors	-0.011	-1.919	0.055
Maximum h-index of authors	**0.129**	**21.143**	**< 0.001**
Time since the publication dates	**0.437**	**77.73**	**< 0.001**
Team size	**0.099**	**14.578**	**< 0.001**
Rate of coauthors in practical affiliations	**0.041**	**7.43**	**< 0.001**
Topic distribution-1	-0.004	-0.556	0.579
Topic distribution-2	-0.008	-1.305	0.192
Topic distribution-3	-0.012	-2.06	0.039
Topic distribution-4	-0.010	-1.624	0.104
Topic distribution-5	-0.005	-0.861	0.389
Topic distribution-6	**-0.018**	**-3.308**	**0.001**
Topic distribution-7	**0.027**	**4.207**	**< 0.001**
Topic distribution-8	**-0.067**	**-10.344**	**< 0.001**
Topic distribution-9	-0.006	-0.995	0.32
Topic distribution-10	0.006	1.098	0.272
Topic distribution-11	0.008	1.451	0.147
Topic distribution-12	**0.024**	**3.385**	**0.001**
Topic distribution-13	**-0.041**	**-6.601**	**< 0.001**
Topic distribution-14	**-0.041**	**-7.289**	**< 0.001**
Topic distribution-15	-0.008	-1.366	0.172
Topic distribution-16	**-0.021**	**-3.639**	**< 0.001**
Topic distribution-17	**-0.036**	**-6.024**	**< 0.001**
Topic distribution-18	0.008	1.29	0.197

Topic-13
Model, predict, network

Topic-14
Express, level, mother, infant

Topic-16
Drug, treatment, cancer, patient, antivirus

Topic-17
Food, water, method, image, technology

Topic-6
Transplant, medicine, effect, kidney, disease

Topic-7
Patient, infect, respiratory, treatment, syndrome

Topic-8
Research, data, technology, data

Topic-12
Outbreak, data, infect, number, epidemic, spread

Results

Regression coefficients

2) PlumX Mention counts

	Coefficient	t -statistic	P-value
Intercept	**0.226**	**56.631**	**< 0.001**
Similarity of authors' previous publications	**-0.023**	**-5.101**	**< 0.001**
Rate of new coauthorships	**0.047**	**9.024**	**< 0.001**
Diversity of authors' h-index	**-0.032**	**-6.257**	**< 0.001**
Cultural similarity of authors	0.005	1.279	0.201
Maximum h-index of authors	**0.103**	**23.041**	**< 0.001**
Time since the publication dates	**0.056**	**13.718**	**< 0.001**
Team size	**0.091**	**18.092**	**< 0.001**
Rate of coauthors in practical affiliations	**-0.028**	**-6.767**	**< 0.001**
Topic distribution-1	0.001	0.27	0.787
Topic distribution-2	0.0004	0.096	0.923
Topic distribution-3	0.004	0.858	0.391
Topic distribution-4	0.004	0.85	0.395
Topic distribution-5	0.001	0.23	0.818
Topic distribution-6	-0.004	-1.056	0.291
Topic distribution-7	**0.013**	**2.733**	**0.006**
Topic distribution-8	**-0.016**	**-3.323**	**0.001**
Topic distribution-9	-0.0002	-0.04	0.968
Topic distribution-10	0.0005	0.126	0.9
Topic distribution-11	0.005	1.102	0.271
Topic distribution-12	0.009	1.696	0.09
Topic distribution-13	-0.009	-1.886	0.059
Topic distribution-14	-0.004	-1.012	0.312
Topic distribution-15	0.002	0.46	0.646
Topic distribution-16	-0.004	-0.922	0.356
Topic distribution-17	-0.004	-0.883	0.377
Topic distribution-18	0.001	0.109	0.913
Topic distribution-19	-0.008	-1.816	0.069

Topic-7
Patient, infect, respiratory, treatment, syndrome

Topic-8
Research, data, technology

Future works

- Add interaction terms to see any interacting effects between variables.
- Prediction task of how collaborations will end up, given author diversity.
- Find out any network structural properties that are found in author collaborations.

Discussion & Conclusions

- Author collaborations from different research areas decrease both academic and public attentions.
- Introduction to new collaborations increase both the academic and public attentions.
- Disparity of authors' prestige decrease both academic and public attentions.