

Varieties of Corona News

A Cross National Study on the Foundations of Online Misinformation
Production During the Covid-19 Pandemic

Cantay Caliskan, University of Rochester

Alaz Kilicaslan, University of Wisconsin-Whitewater

Motivation

- Misinformation is all around – and especially prevalent during Covid-19
 - ***“Anybody that wants a test can get a test. That’s what the bottom line is.”*** (Donald Trump)
- There are local and global tendencies to misinformation
- Varieties of Capitalism and other ‘Varieties of...’ literature (Hall and Soskice, 2001)
- Expectation of national variations in the production of misinformation

Literature

- How rumors spread on Twitter (Zubiaga et al., 2016)
- Real-world impacts of hoaxes at Wikipedia (Kumar et al., 2016)
- How individuals consumed fake news prior to 2016 US presidential election (Allcott and Gentzkow, 2017)
- How regular people actively participate in generating disinformation in Russia (Mejias and Vokuev, 2017)

Definition

- ***Misinformation***: Incorrect or misleading information
 - Any piece of information that is partly or fully false can be labeled as misinformation
 - The definition of misinformation also contains disinformation
- We avoid the term ***fake news*** because:
 - Polarizing
 - Politically charged
 - Rather limited to forms of misinformation that are deliberately designed to mimic news (Guess and Lyons, 2020)

Misinformation and Covid-19

- **Three** main streams of research:
 - ***First stream*** tries to understand **different types** of misinformation in terms of their sources, spread patterns
 - Accurate news are less likely to be shared than inaccurate ones
 - There is more misinformation circulating on social media platforms
 - ***Second stream*** looks at **specific factors** that push individuals to believe in or share misinformation
 - Political conservatism
 - Right-leaning media consumption
 - ***Third stream*** looks at how exposure to misinformation affects **health behavior** during the pandemic
 - More misinformation -> Less willingness to take preventive measures

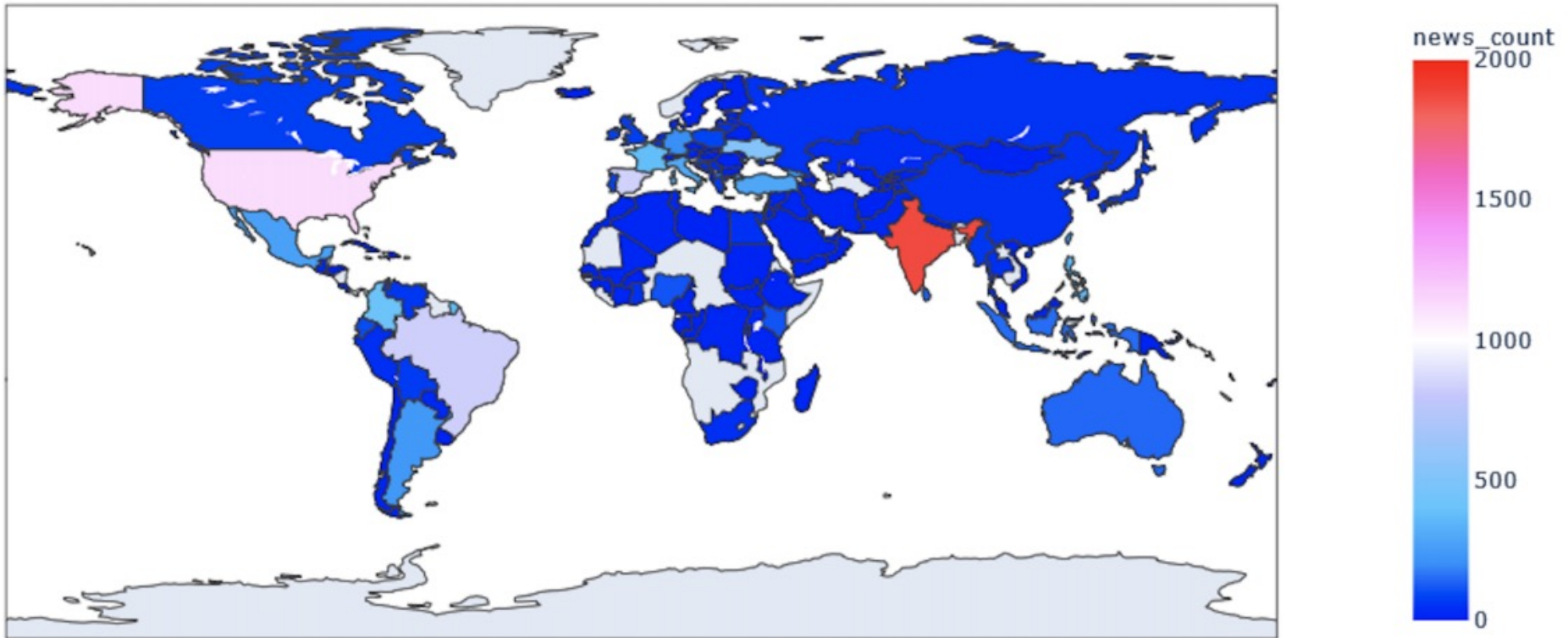
Data

Data

- **10,131** pieces of falsehood collected from the ***Poynter Institute***
- Covers the period **January 2020 – February 2021**
- Uses carefully-controlled hand-labeling of topics
 - **28 topics** identified
- Specifically: aid, animals, conspiracies, crime, cures, detection, food, governments, hospitals, individuals, insurance, laws, lockdown, medical equipment, medicine, origins, other diseases, predictions, prevention, religion, risk factors, spread, symptoms, travel, vaccines, videos, technology, and NGOs

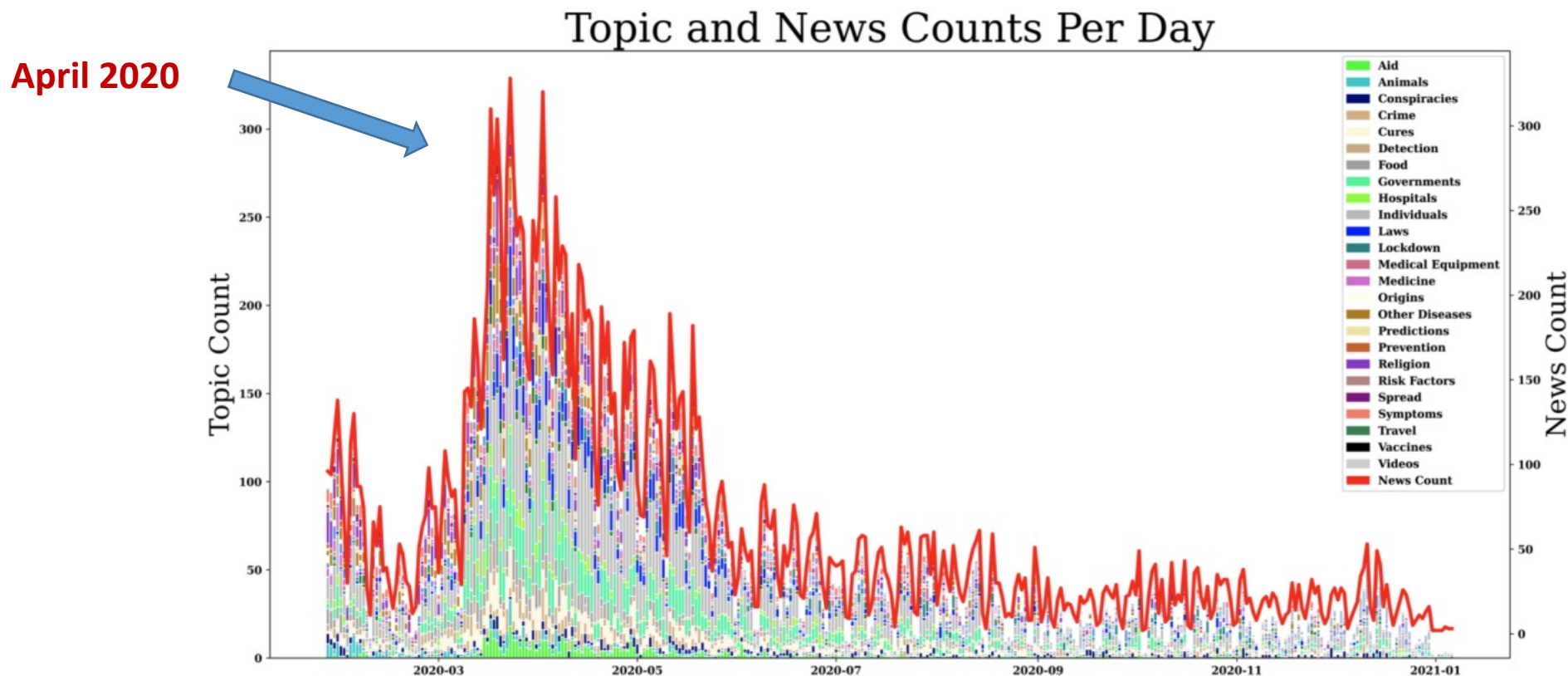
Data

- Most news are from India (1892 observations)



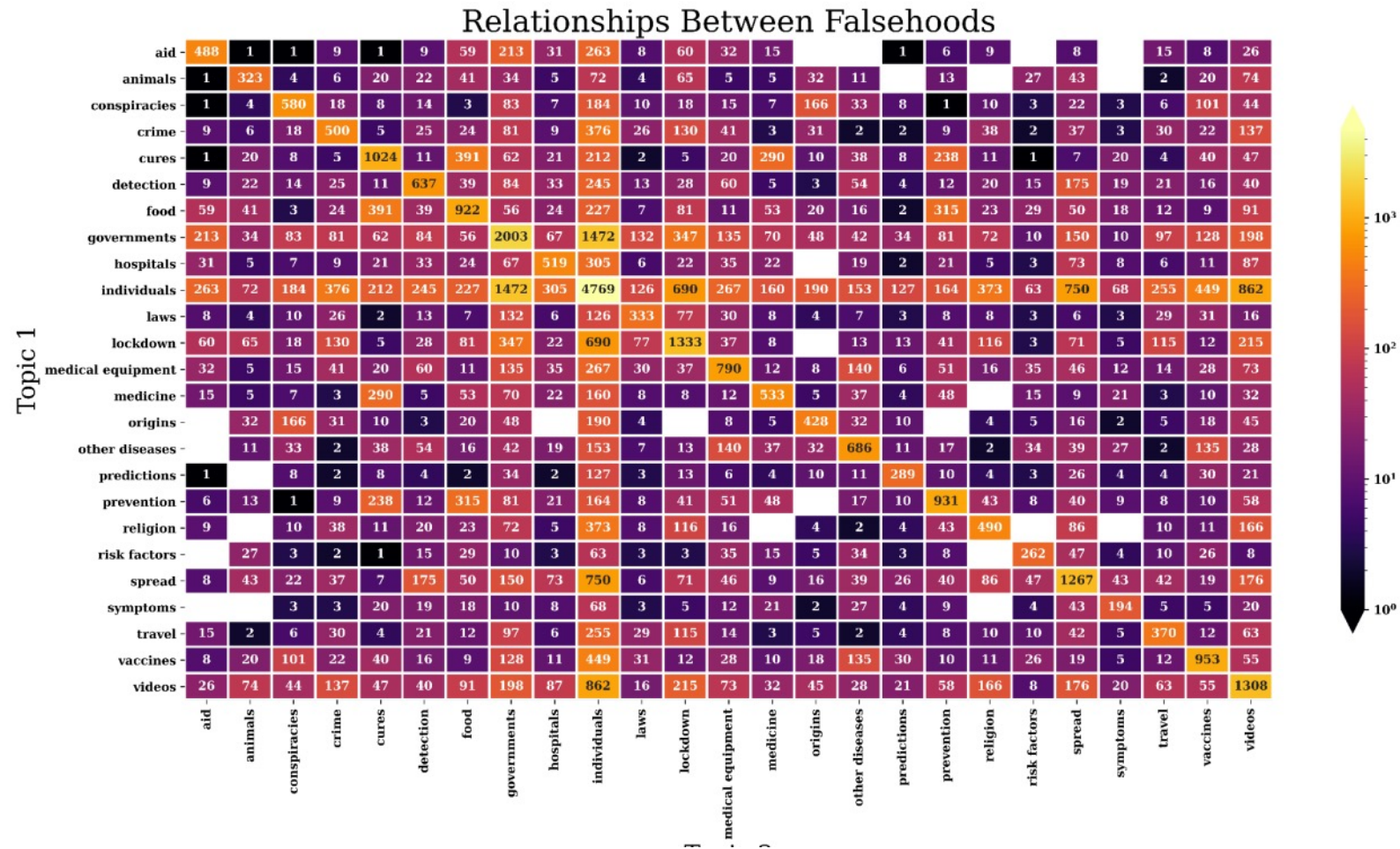
Data

- Misinformation production reached its peak in April 2020
- There are seasonal effects



Co-Occurring Topics

- Individuals, governments, spread, and videos are the most co-occurring topics



Research Questions: Topic Analysis

- **Topic Analysis**

- ***Divisive*** and ***connective*** topics

- **Q1a**: In terms of topic creation, what are the topics that two groups of countries utilize in the most comparable amount vis-à-vis each other? What are the topics that are the most ***connective***?
 - **Q1b**: What are topics that are the most ***divisive***?

- Topic ***co-occurrences***:

- **Q2a**: What are the topics that ***co-occur*** the most?
 - **Q2b**: Are some ***groups of countries*** statistically significantly different from others in terms of topic co-occurrence?
 - **Q2c**: Is there a ***time frame*** in which topics were more similar to each other?

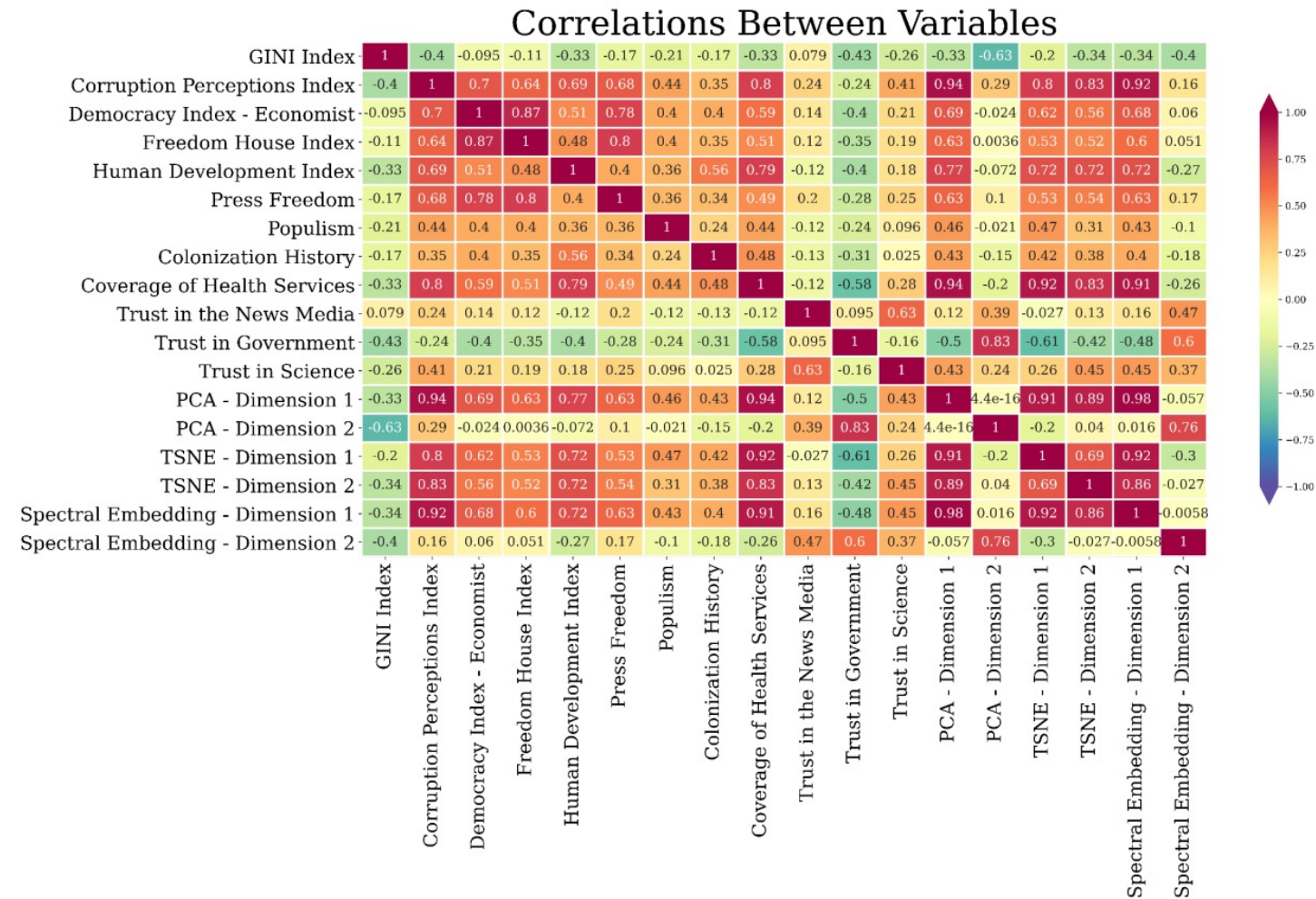
Research Questions: Content Analysis

- **Content Analysis**
 - ***Content Similarity***
 - Q3a: Are there groups of countries that produce news that are significantly more similar to each other?
 - Q3b: How does the similarity between news change over time?
 - ***Misinformation Unusualness / Creativity***
 - Q4a: Are there groups of countries that are more creative than others in content formation?
 - Q4b: How does creativity evolve over time?

'Varieties of Variables'

- Freedom in the world (Freedom House)
- Public trust in government (Edelman Data and Intelligence)
- Corruption Perceptions Index (Transparency International)
- World Press Freedom Index (Reporters Without Borders)
- Trust in Media (Reuters Institute for Study of Journalism)
- Trust in Science (Wellcome Trust)
- Populism Index (Luigi Curini – Università degli Studi di Milano)
- World Cultural Map (World Values Survey)
- Clusters of Business Systems (Witt et al.)
- Human Development Index (United Nations Development Program)
- GINI Index (World Bank)
- Coverage of Health Services Index (Global Burden of Diseases)
- Colonization History
- United Nations Geoscheme (United Nations Statistics Division)

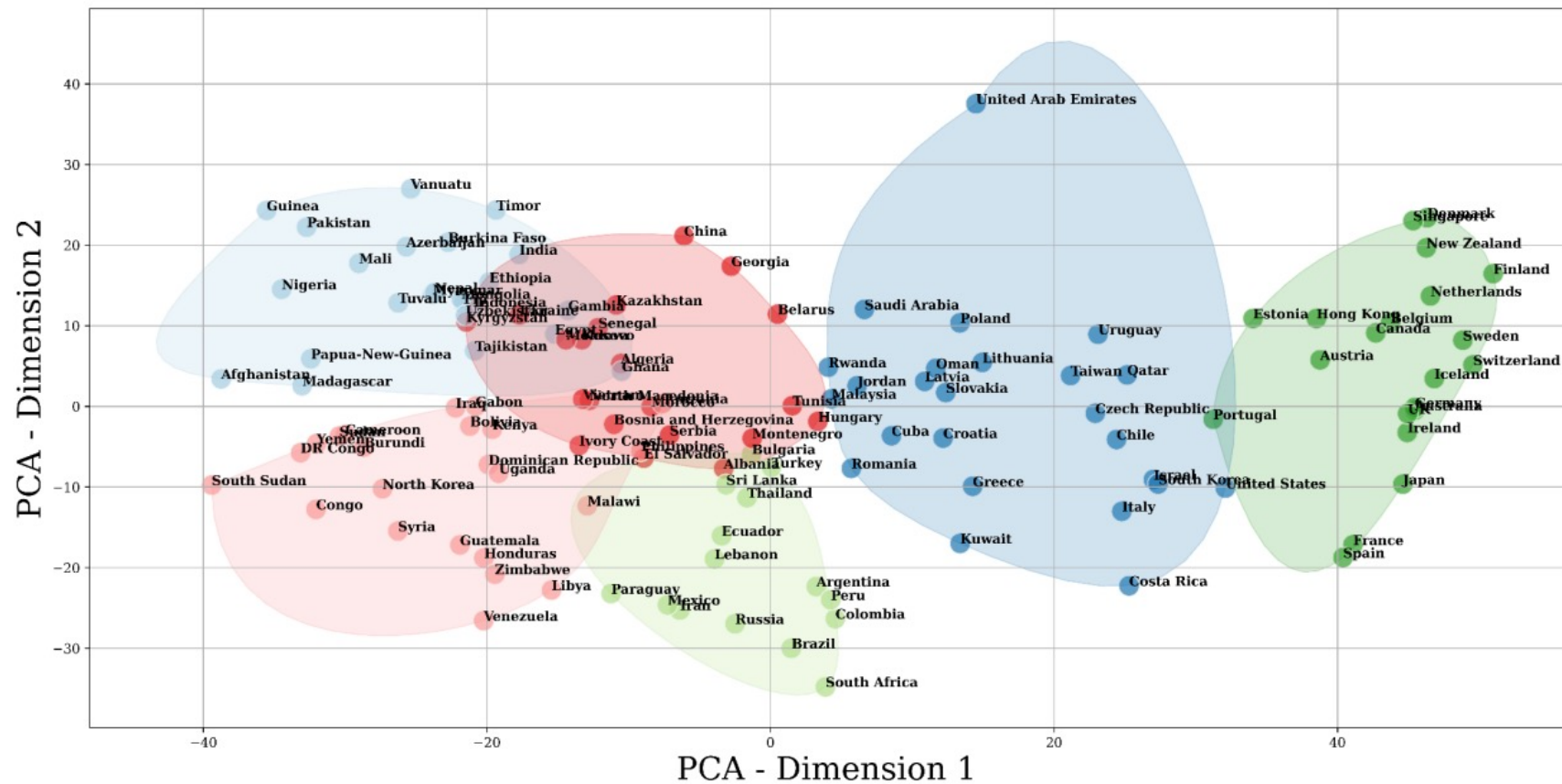
'Varieties of Variables'



It looks like: the differences between countries can be summarized by **GINI Index (i)**, **Corruption Perceptions Index (ii)**, and **Coverage of Health Services Index (iii)**

‘Varieties of Variables’

Two-Dimensional Map of Countries (PCA)



Methods

'Varieties of Variables'

- ***Descriptive analysis***
 - Kmeans++
- ***Topic analysis***
 - Entropy and GINI Index (as a measure of information gain)
 - Network similarity algorithms (Frobenius and Quantum-JSD similarity)
- ***Content analysis***
 - Cosine similarity (n-grams and TF-IDF scores)
 - t-test

Results

Topic Analysis: Divisive and Connective Topics

	Minimum Entropy / GINI Index (most connective)	Maximum Entropy / GINI Index (most divisive)
1	Conspiracies (0.071)	Animals (0.387)
2	Medical Equipment (0.058)	Predictions (0.135)
3	Risk Factors (0.053)	Symptoms (0.067)
4	Other Diseases (0.048)	Laws (0.063)
5	Vaccines (0.047)	Travel (0.046)
20	Food (0.005)	Videos (0.002)
21	Lockdown (0.004)	Vaccines (0.001)
22	Governments (0.003)	Governments (0.000)
23	Spread (0.003)	Cures (0.000)
24	Prevention (0.002)	Individuals (0.000)



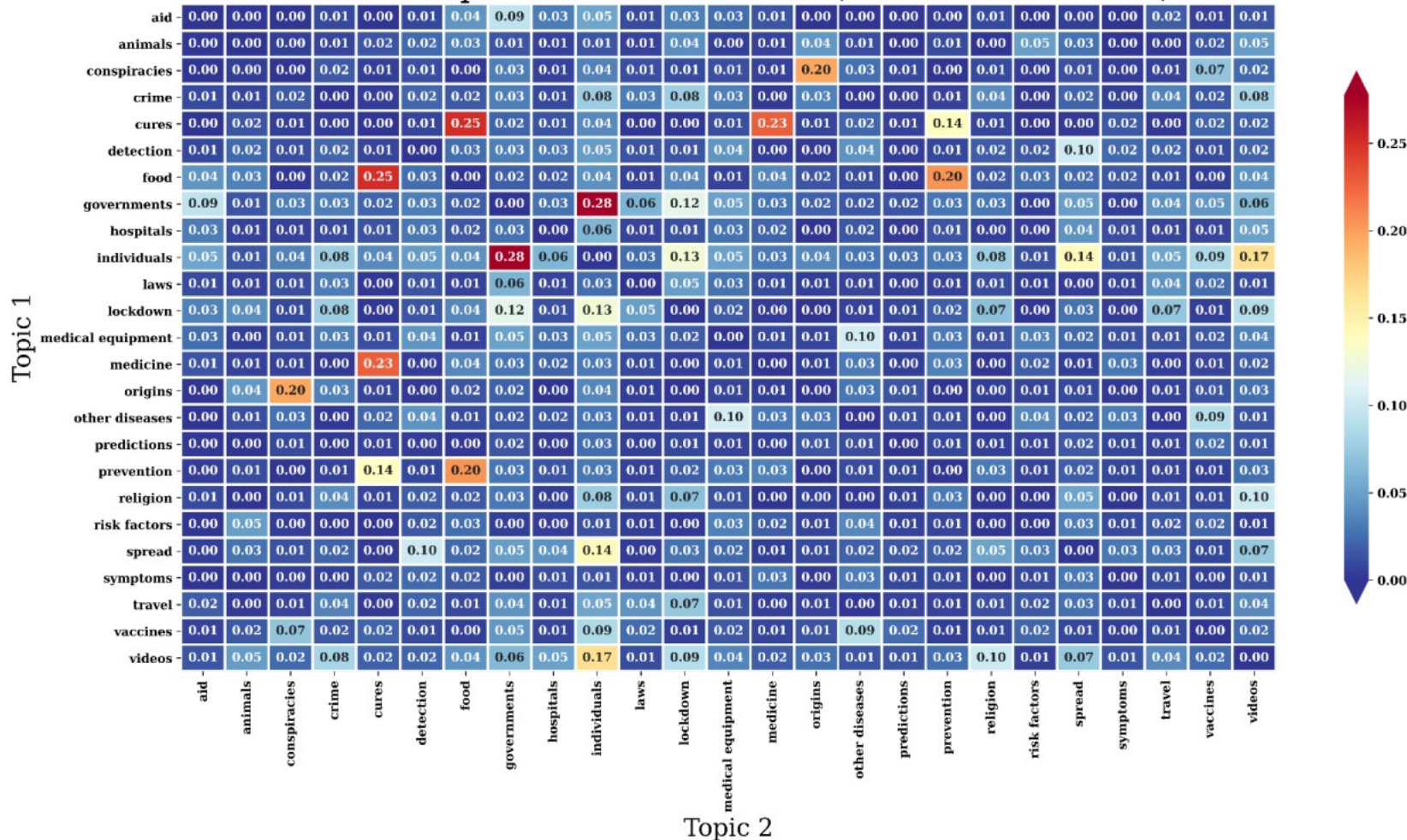
Most **connective** and most **divisive**



Least **connective** and least **divisive**

Topic Analysis: Topic Co-Occurrences

Relationships Between Falsehoods (Relative Similarities)

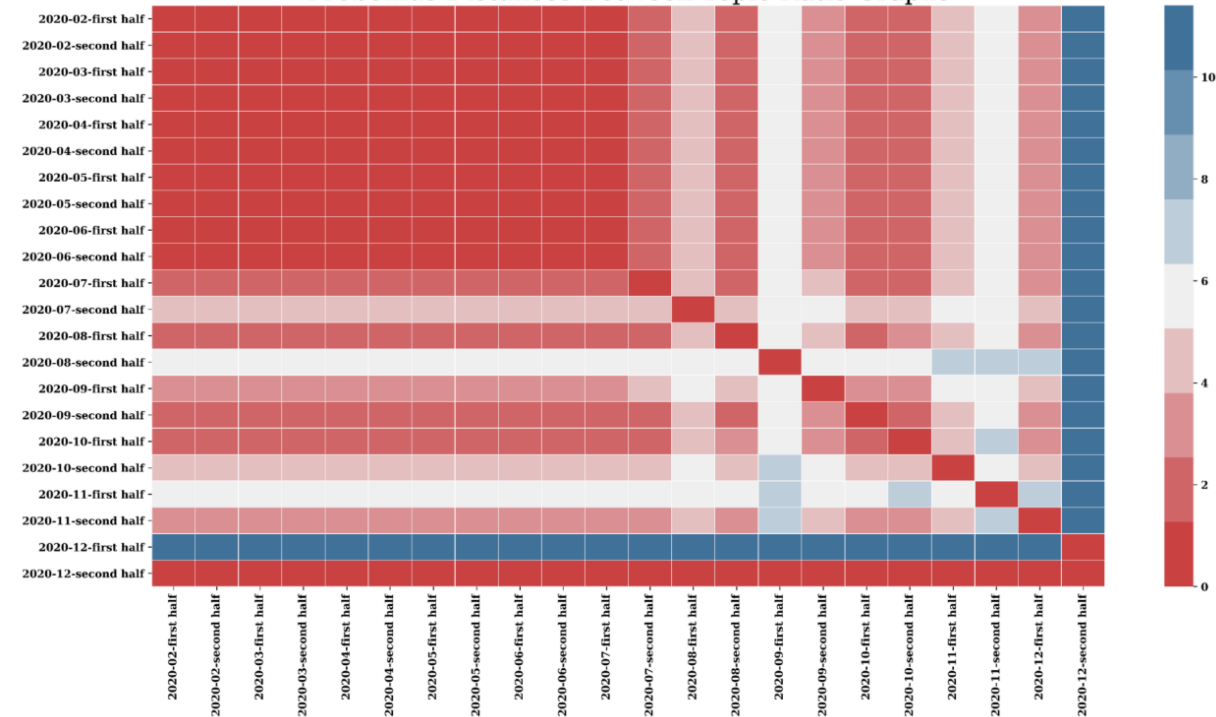


High similarity co-occurring topics:

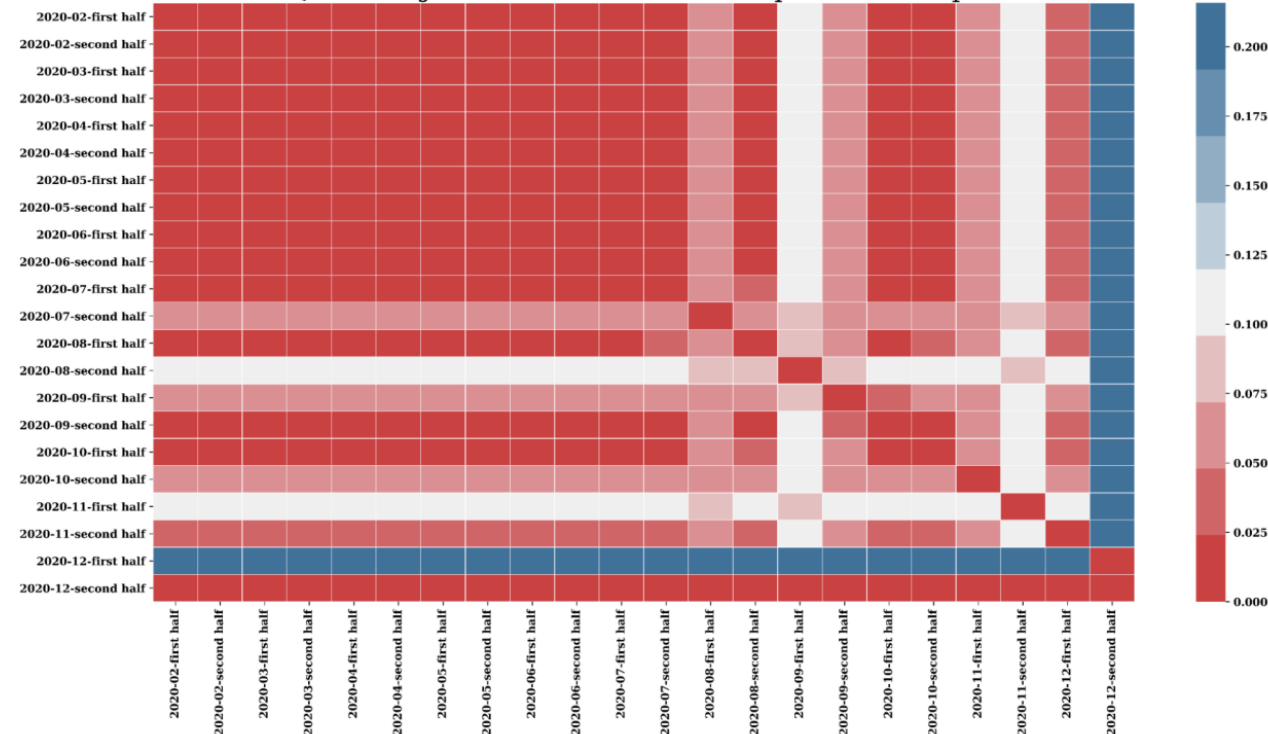
Food-cures
 Individuals-governments
 Lockdown-governments
 Lockdown-individuals
 Medicine-cures
 Origins-conspiracies
 Other diseases-medical equipment
 Prevention-cures
 Prevention-food
 Spread-detection
 Spread-individuals
 Videos-individuals
 Videos-religion

Topic Analysis: Topic Co-Occurrence

Frobenius Distances Between Topic Ratio Graphs

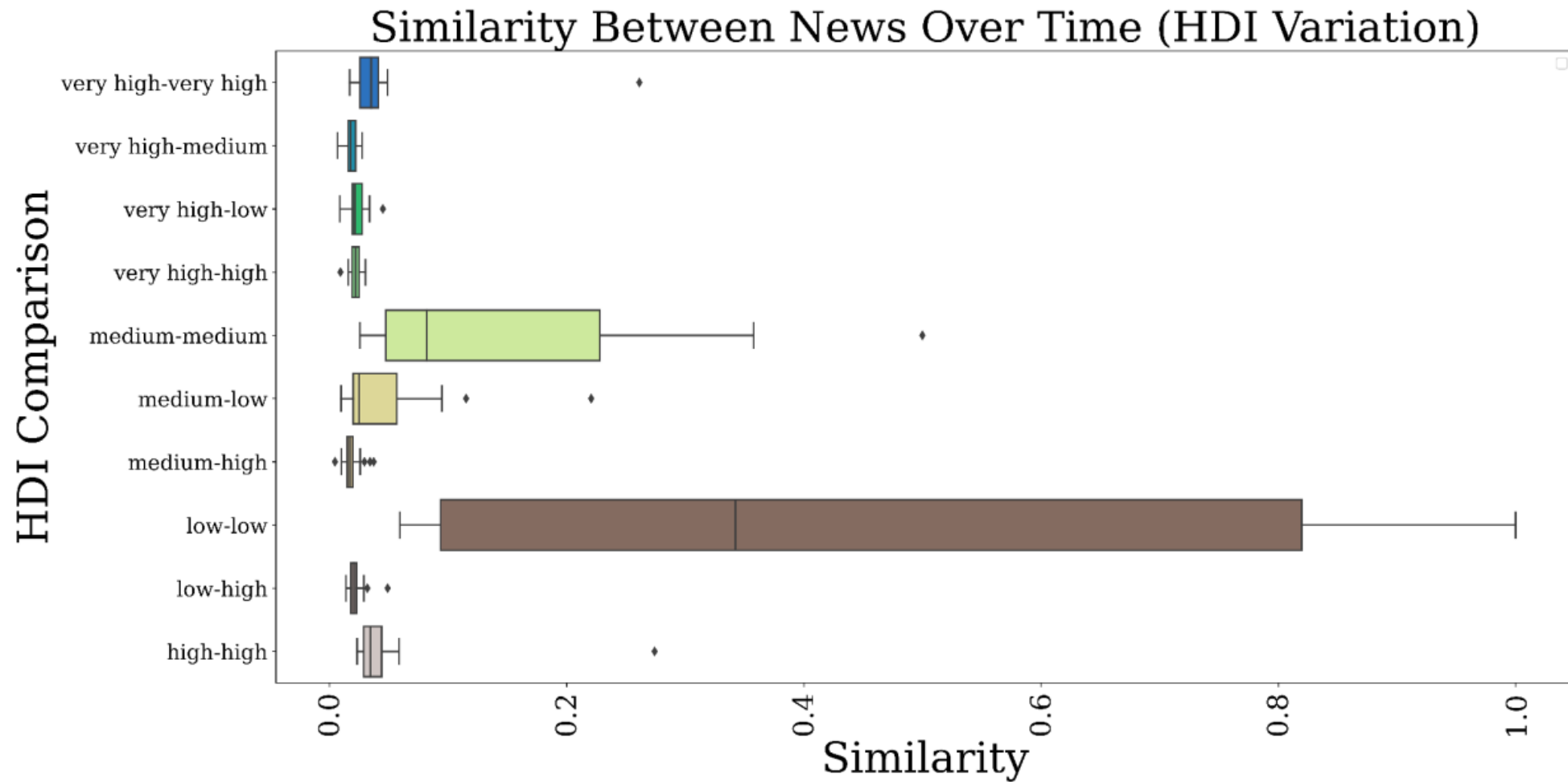


Quantum-JSD Distances Between Topic Ratio Graphs

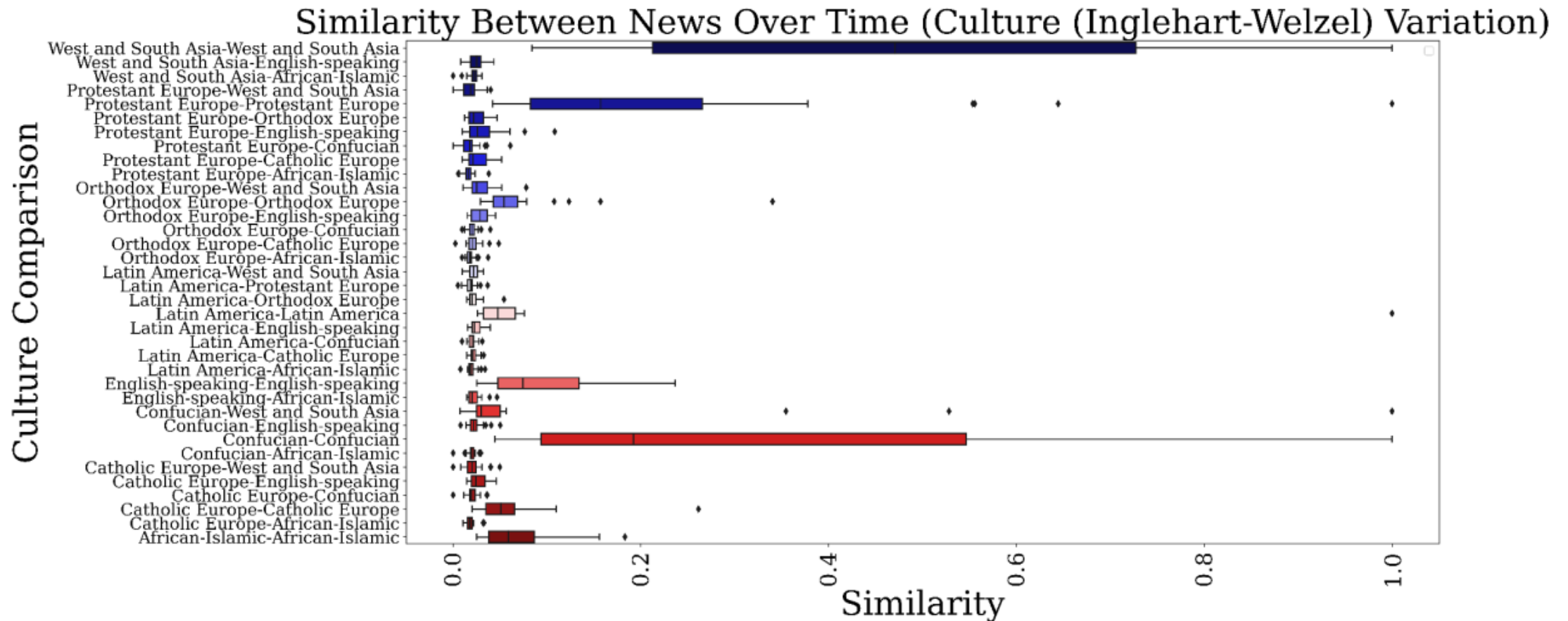


Result: More similarity in the first half of 2020.

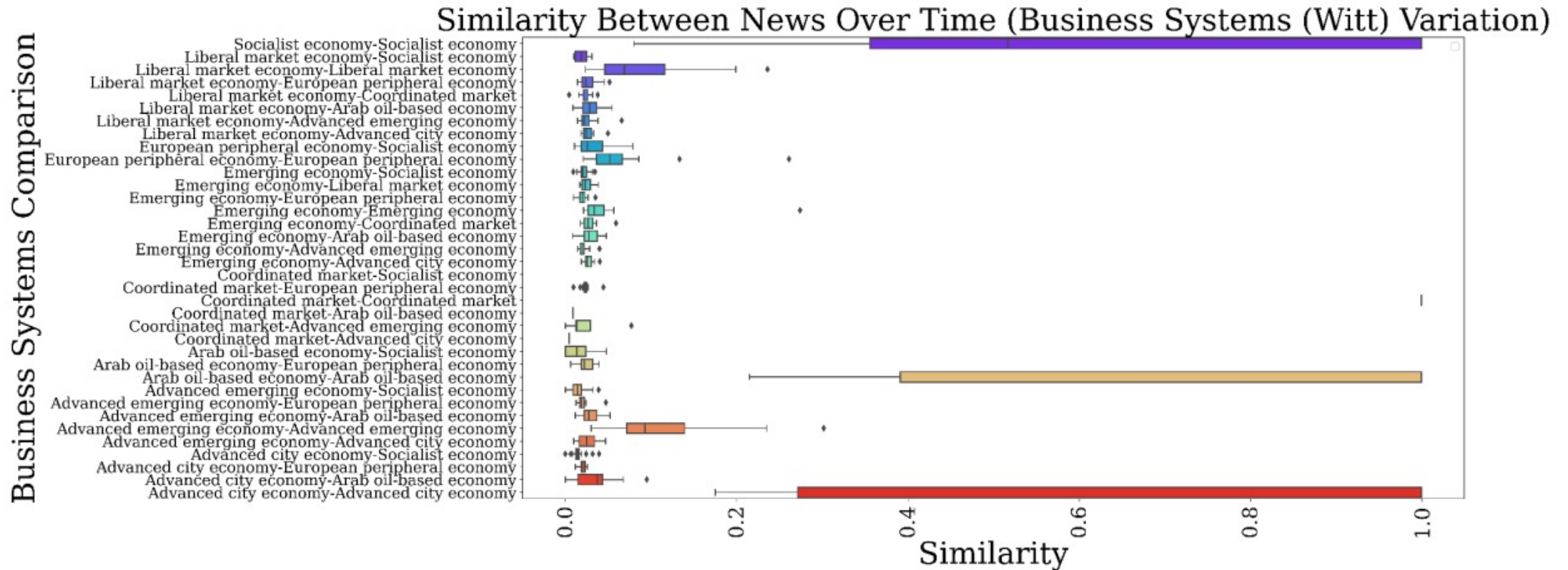
Content Analysis: Content Similarity



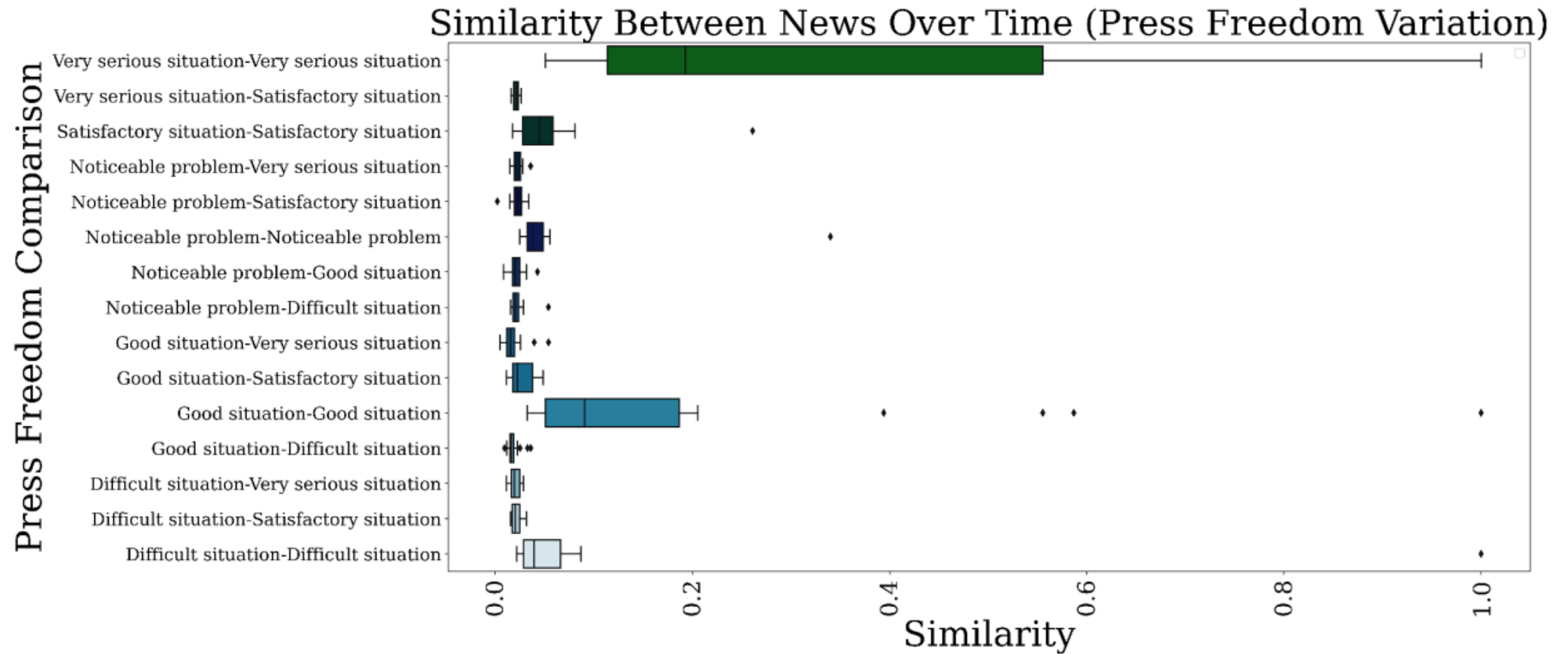
Content Analysis: Content Similarity



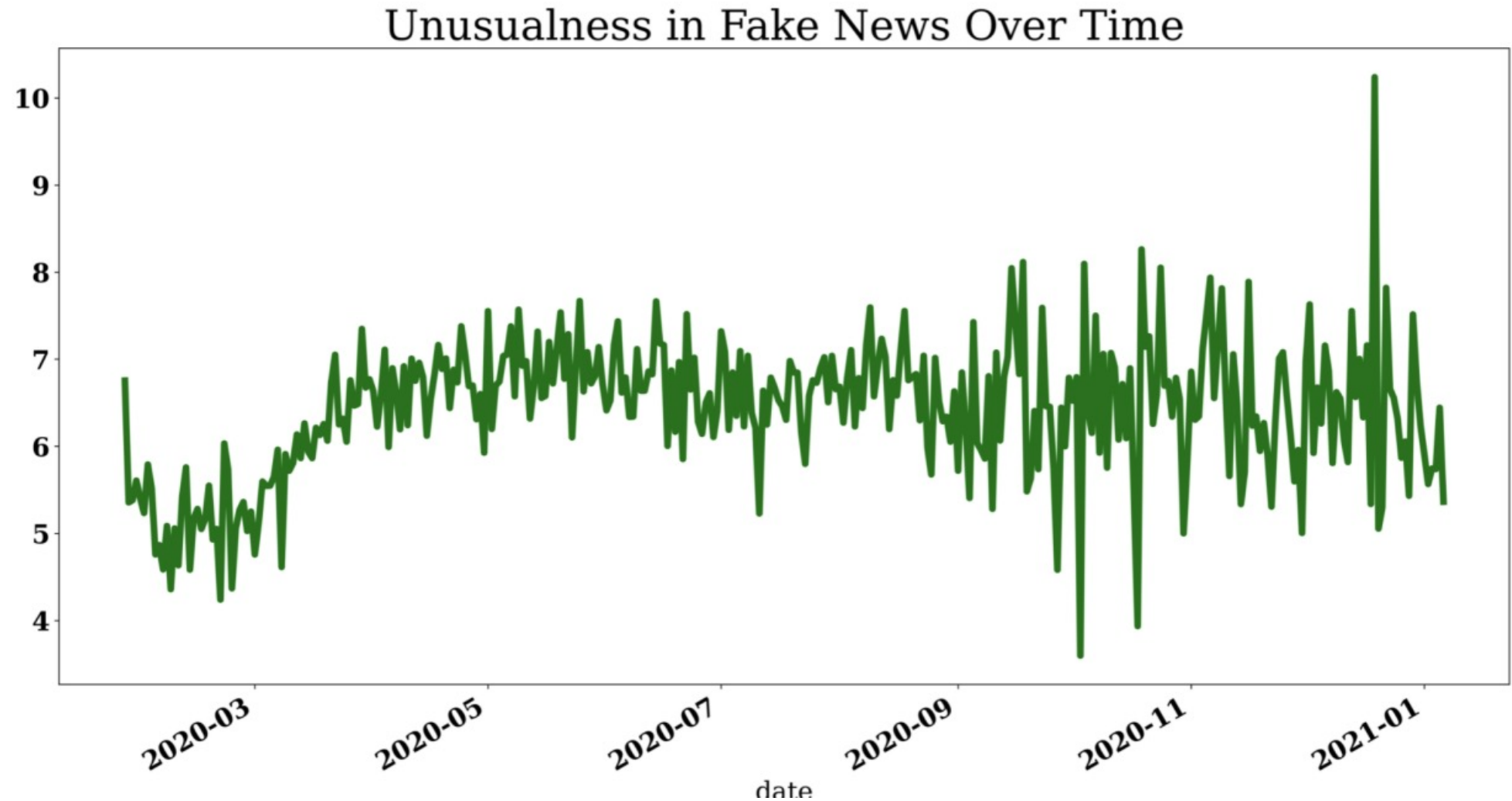
Content Analysis: Content Similarity



Content Analysis: Content Similarity



Content Analysis: Misinformation Creativity



Conclusion

Key Takeaways

- 1) Serious press freedom problems and low HDI result in similar content creation.
- 2) Intensity of discussion on animals, predictions, and symptoms is the biggest differentiator between countries.
- 3) Countries with low HDI produce misinformation related to unproven local remedies and those stemming from certain religious beliefs as well as from distrust of international organizations and Western medical practices.
- 4) Falsehoods were more common in the first months of the pandemic and then declined.
- 5) Most prominent misinformation topics vary across countries; but, the word-groups used in misinforming news stories are remarkably similar.