

MAPPING GLOBALIZATION THROUGH EVENT THEMES

Tan, Jefferson; Nebres, Elisa Jr.; Legara, Erika; Monterola, Christopher
Asian Institute of Management, Philippines

Presented on CCS 2019, NTU, Singapore; September 30, 2019

A large, semi-transparent graphic on the left side of the slide features a complex arrangement of overlapping hexagons and octagons in shades of blue and grey. A thin blue curved line starts from the top right corner and sweeps down towards the bottom left, partially enclosing the geometric shapes.

OUTLINE OF THE PRESENTATION

Introduction
Review of Related Literature
Data and Methodology
Results
Recommendations

MOTIVATION OF THE STUDY



- Politics and Campaigns
- Natural Disaster and Relief
- Technology, Industry and Trade



● ABSTRACT

We can map out symptoms of globalization from GDELT data to analyze influence and trends for the use of prioritization of political action, international diplomacy and events prediction.

REVIEW OF RELATED LITERATURE

Keertipati, S., Savarimuthu, B. T. R., Purvis, M., & Purvis, M. (2014). Multi-level Analysis of Peace and Conflict Data in GDELT

- Analyzed historical trends from Sri Lankan civil war and 2006 Fijian coup, to show that GDELT news data accurately captures global events.
- Study used change point analysis to identify significant points along the course of conflicting events in the cases. These significant points can be used as triggers for an upcoming conflict, which organizations can use for preventive and monitoring measures.

REVIEW OF RELATED LITERATURE

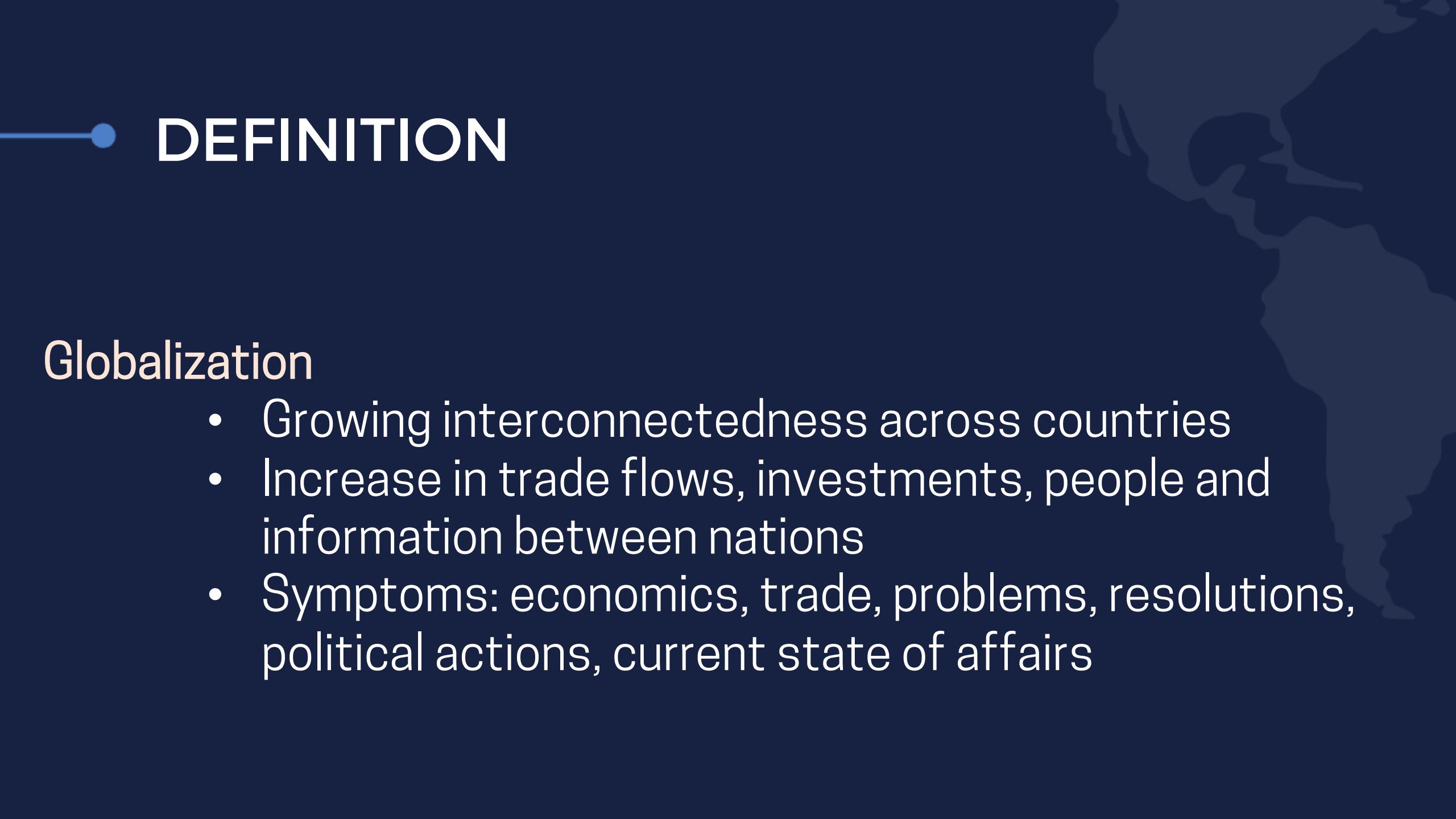
De Spiegeleire S., Solodovnik I., and Farnham N. (2017). Conflict and Cooperation The Hague Centre for Strategic Studies

- Tracked the overall trends in global cooperation and conflict as contribution to the Dutch government's Strategic Monitor using GDELT Data
- Aimed to formulate conceptual foundation to help decide matters regarding security, especially its armed forces
- Their results show that there is a downward trend in cooperation and an upward trend in conflict since the early 2000s, and also a downward trend in the Goldstein Scale

REVIEW OF RELATED LITERATURE

Cingolani, S. and Panzarasa, P. (2017). Countries' Positions in the International Global Value Networks: Centrality and Economic Performance

- Adopted a network-based perspective and proposed a three-faceted measure of centrality
- Captures countries' distinct roles at the upstream, midstream, and downstream stages of the international production process
- Rankings of countries based on our measures of centrality vary across production stages.
- Emerging and developing countries tend to secure central positions at upstream and midstream production stages
- High-income countries tend to exert prevailing roles at downstream stages. The network-based rankings differ from traditional measures of market power (aggregate trade values), especially within the more traditional industries



DEFINITION

Globalization

- Growing interconnectedness across countries
- Increase in trade flows, investments, people and information between nations
- Symptoms: economics, trade, problems, resolutions, political actions, current state of affairs

DATA SET

GDELT 2.0

GLOBAL DATABASE OF EVENTS,
LANGUAGE AND TONE

Global Knowledge Graph

Broadcast, print, and web news across
different countries in 100 languages



Themes
List of themes found in the
document or news article

Locations
List of locations found in the text
extracted through the Leetaru Algorithm
(2012)

DATA PREPARATION AND METHODOLOGY

1 Extract Raw Data from GDELT

2 Prepare Theme List

3 Aggregate the Records

4 Construct Network

PREPARE AND PROCESS THEME LIST

GOV_REPATRIATION
Repatriation

GRIEVANCES
Mention of the term "grievances" as explicit discussion of societal unhappiness.

HARASSMENT
Explicit mention of the word "harassment"

HEALTH_VACCINATION
Mentions of vaccination, immunization

HUMAN_TRAFFICKING
Human trafficking, sexual slavery, prostitution, human smuggling,



Original GDELT Theme
HEALTH_VACCINATION
GENERAL_HEALTH
MEDICAL
SMUGGLING
TREASON
RAPE
BULLYING
DISCRIMINATION
GENDER_VIOLENCE

New General Theme
Health
Health
Health
Crime
Crime
Crime
Social Issues
Social Issues
Social Issues

PREPARE AND PROCESS THEME LIST

21 Themes (inspired by World Bank Topical Taxonomy)

Law, Food, Crime, Media, Crisis, Health, Unrest, Military,
Religion, Violence, Education, Corruption, Leadership,
Agriculture, Social Issues, Infrastructure, Economic Policy,
Government and Politics, Urban and Rural Development,
Social Development and Protection,
Environment and Natural Resource Management

CONSTRUCTING THE NETWORK

Nodes: Countries and Themes

Edges: A significant number of events connecting country and theme

Weight: Standardized count of articles mentioning both country and theme

*Separated by Year



Compute for Centrality
Measures



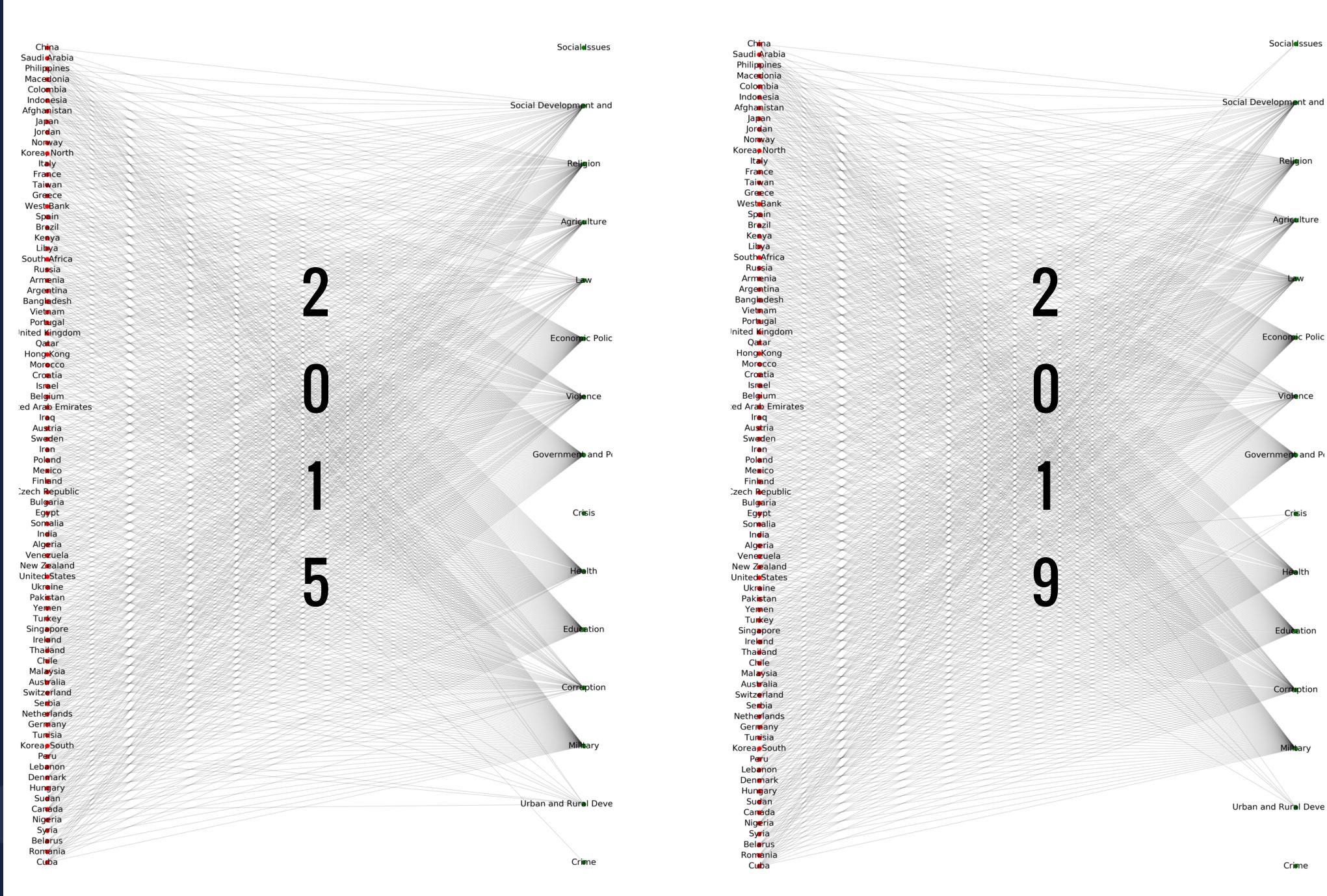
Identify Communities



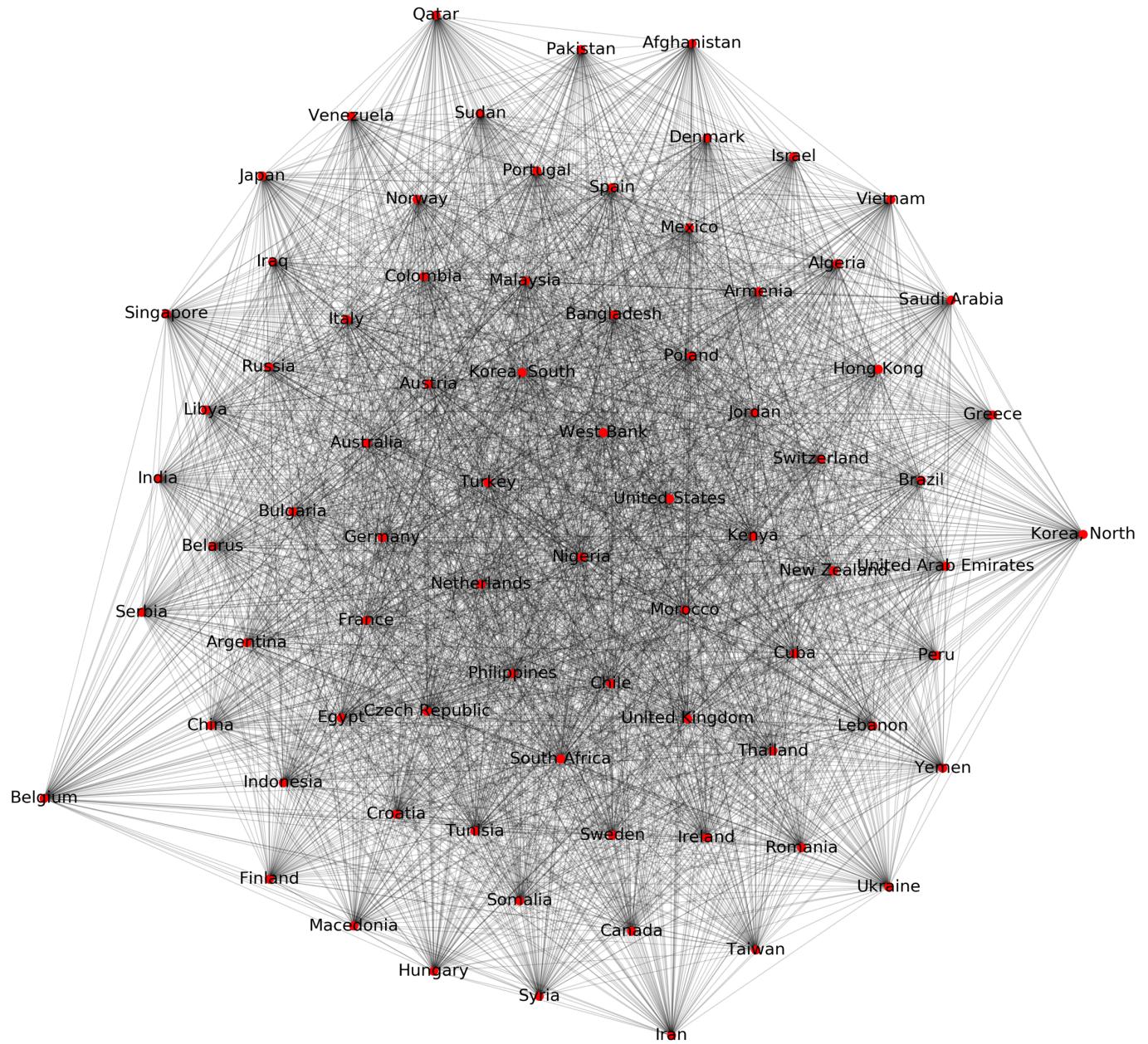
Trend analysis of
Communities

Bipartite of Countries to Observed Event Types (2015)

Bipartite of Countries to Observed Event Types (2019)



Unipartite of Countries with similar Observed Events (2019)



ASSUMPTIONS

Strength of media coverage leads to bias and limitations

Articles imply involvement of countries to themes

Each country engages to equally significant number of events

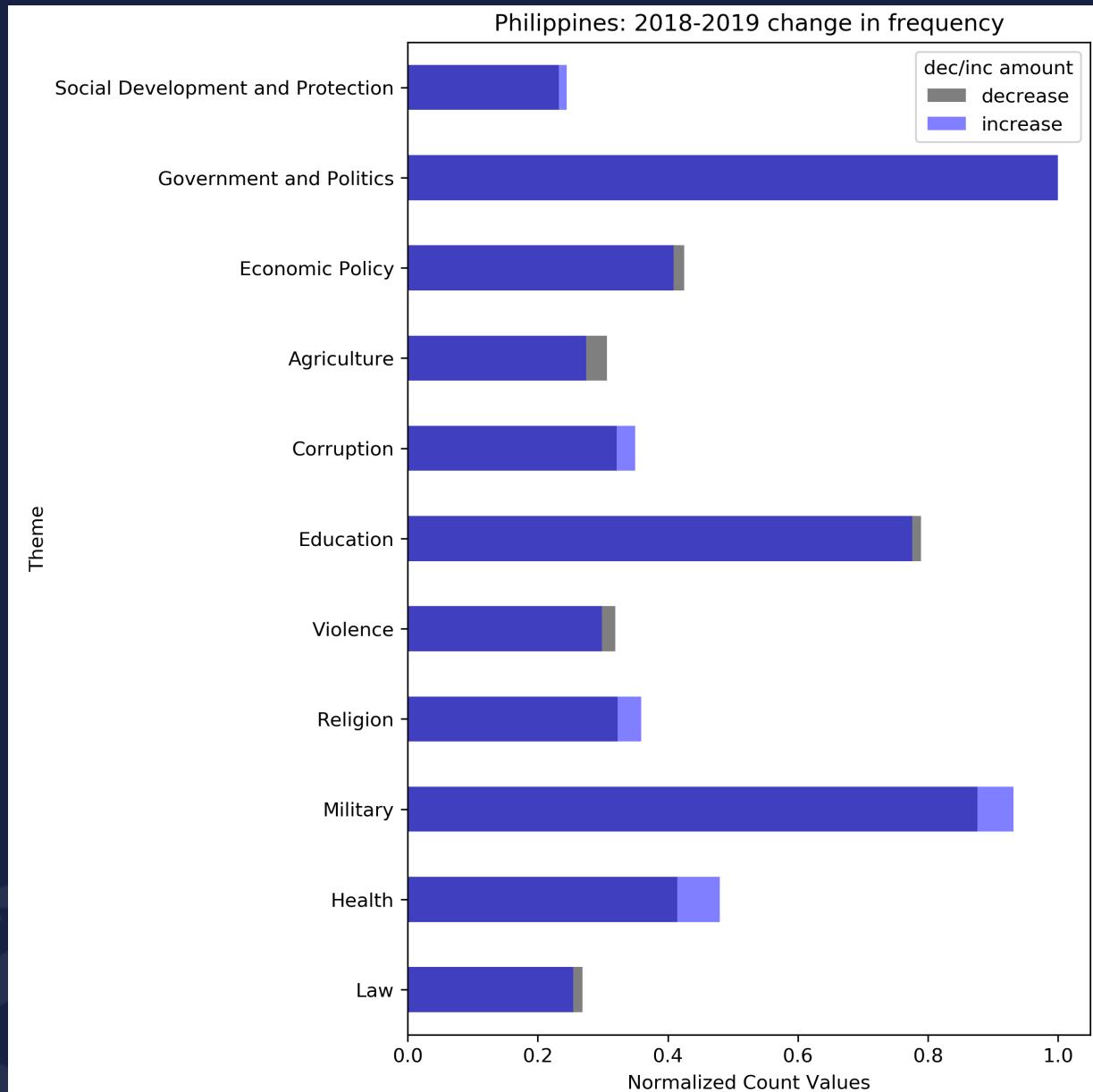
ADDITIONAL STEPS

Filter out countries with little information

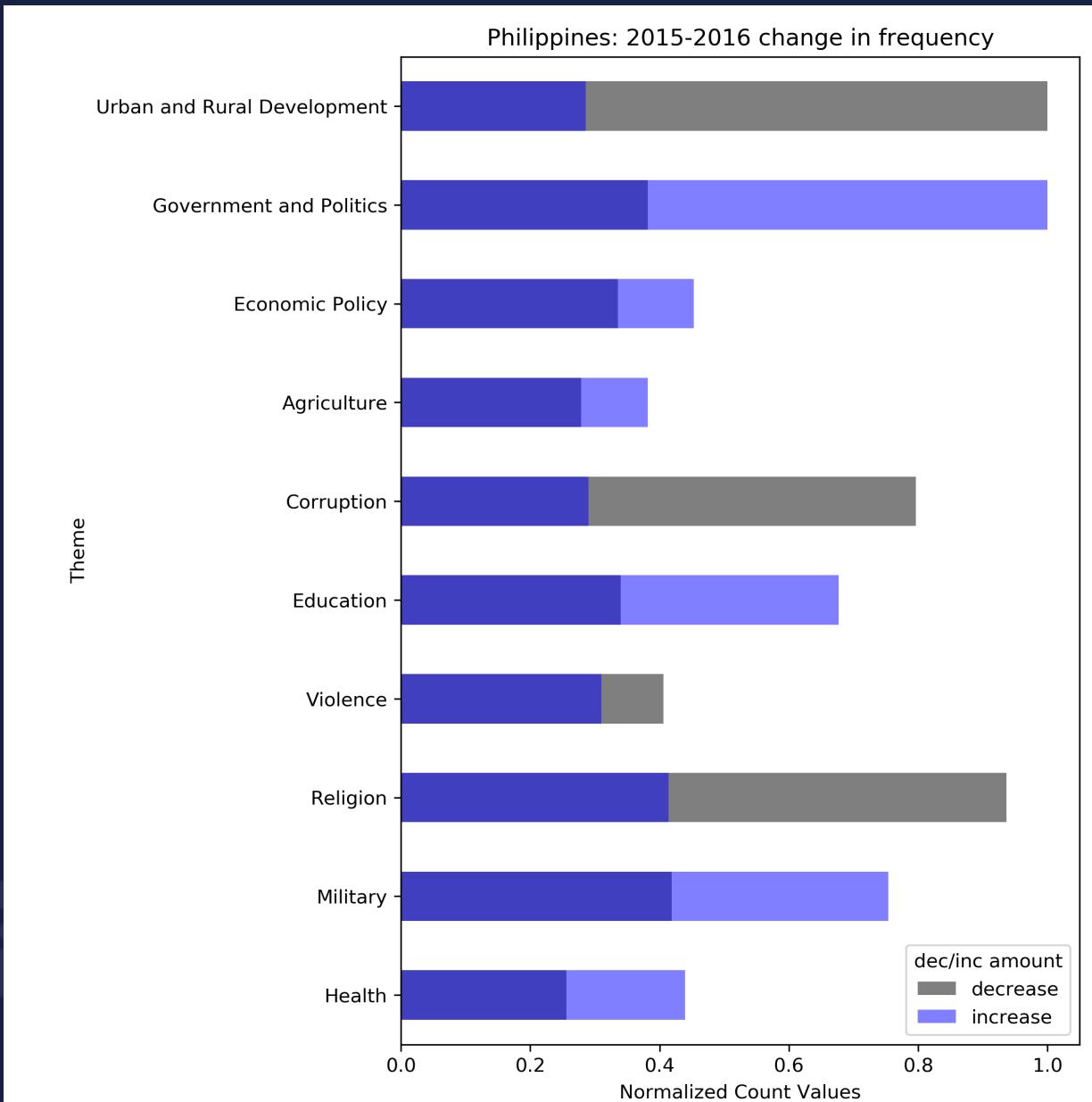
Standardize count values (MinMaxScaler)

Remove weak connections below a threshold

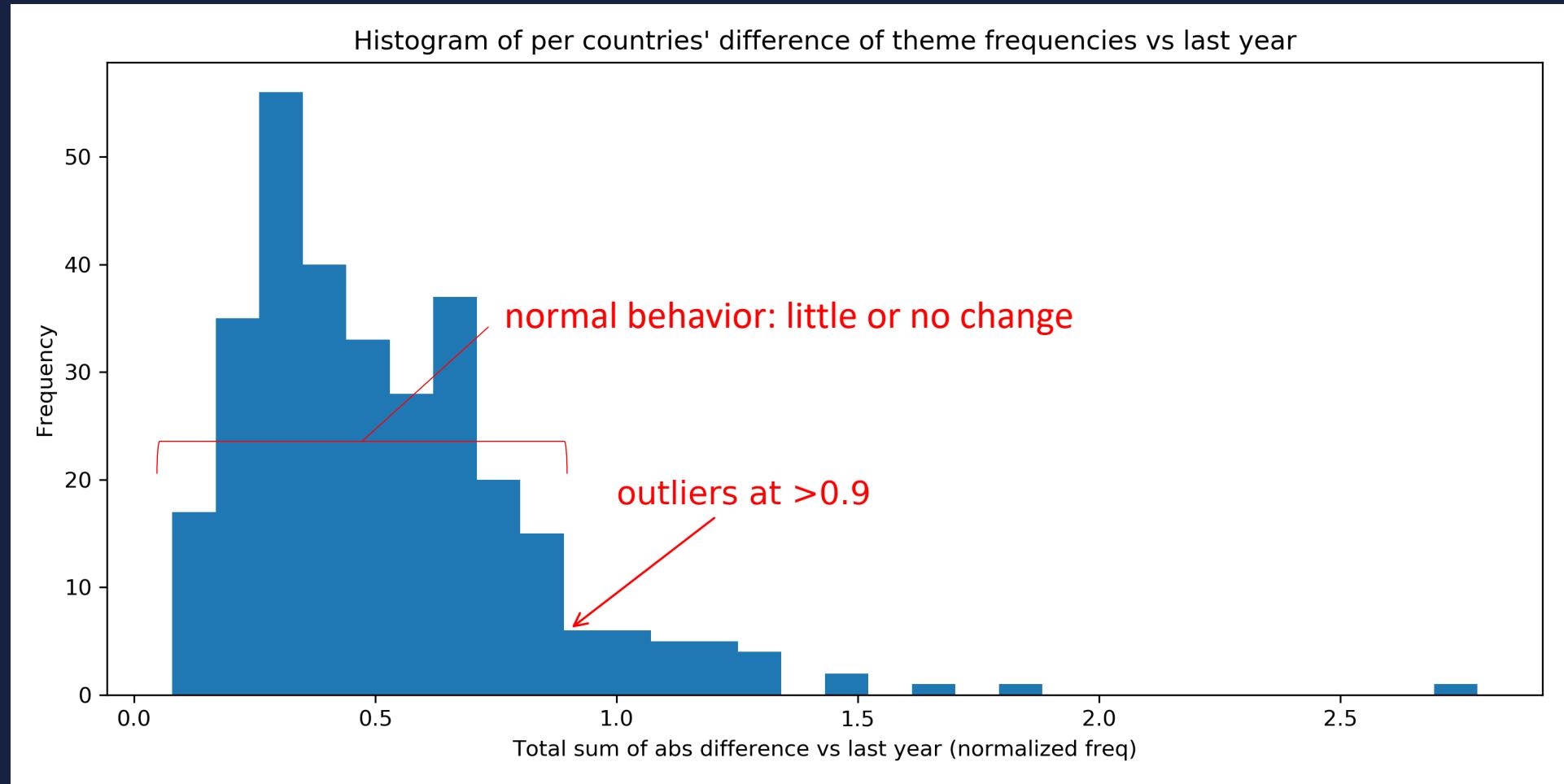
TRENDS OF ENGAGEMENTS



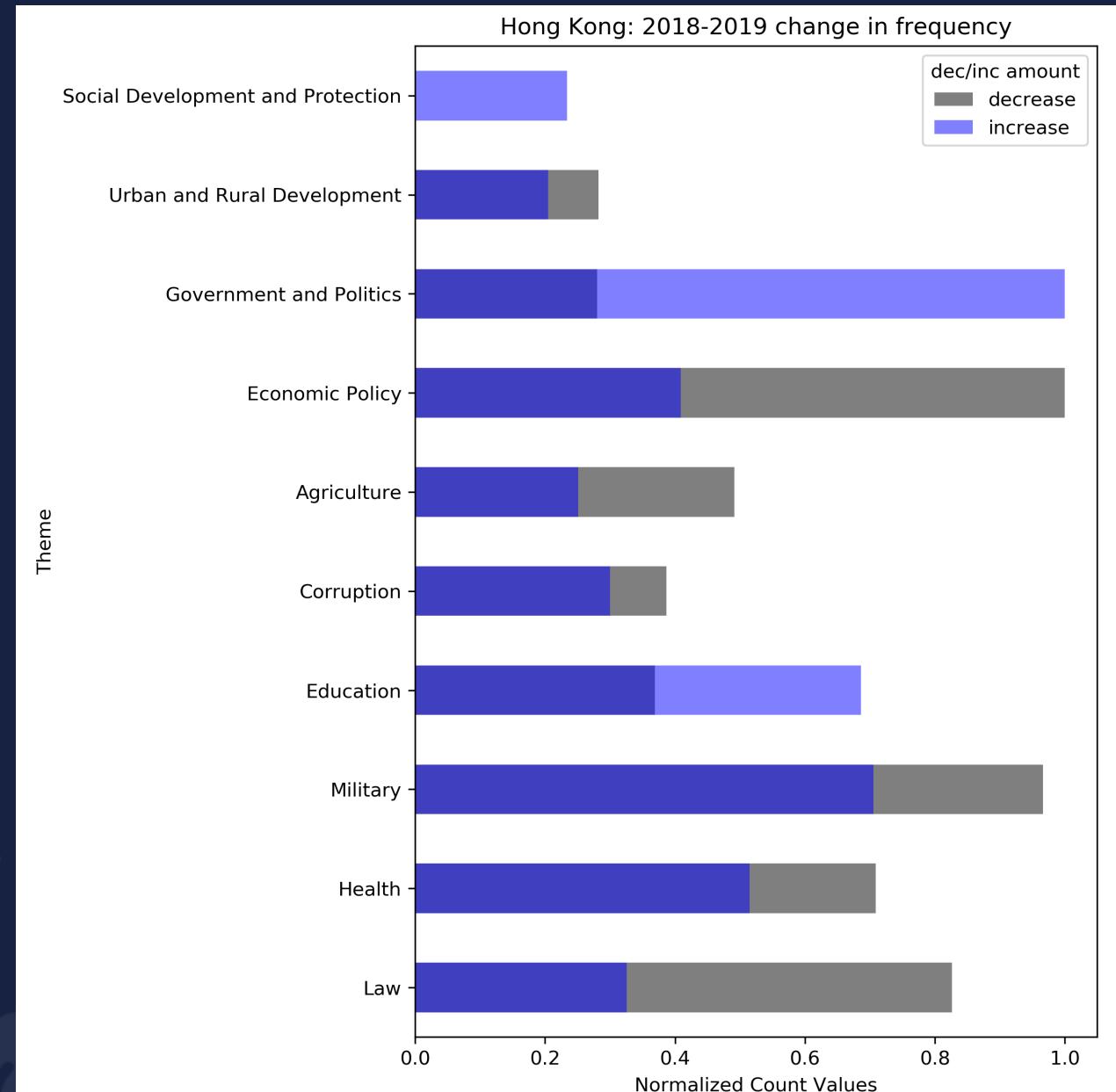
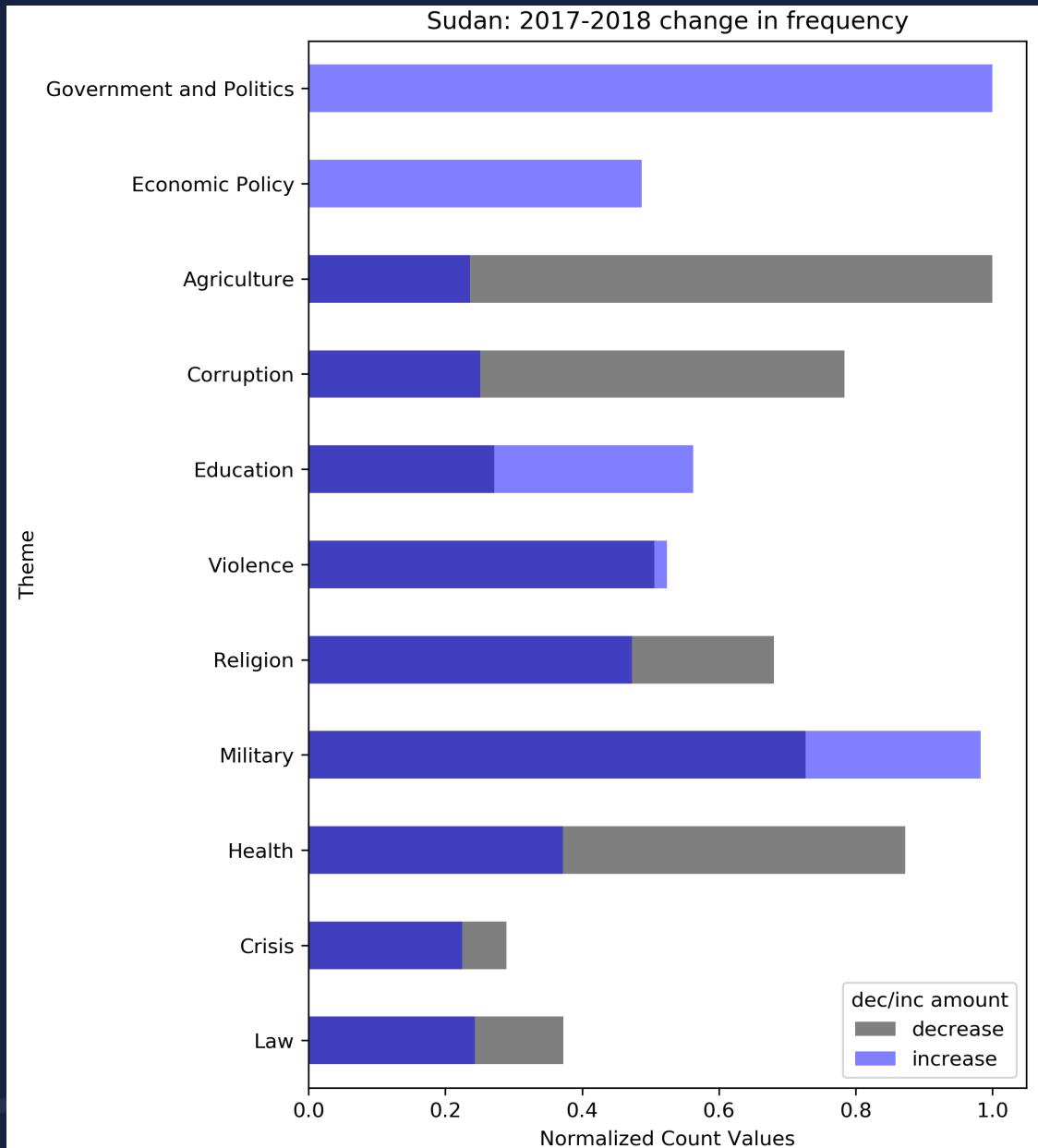
TRENDS OF ENGAGEMENTS



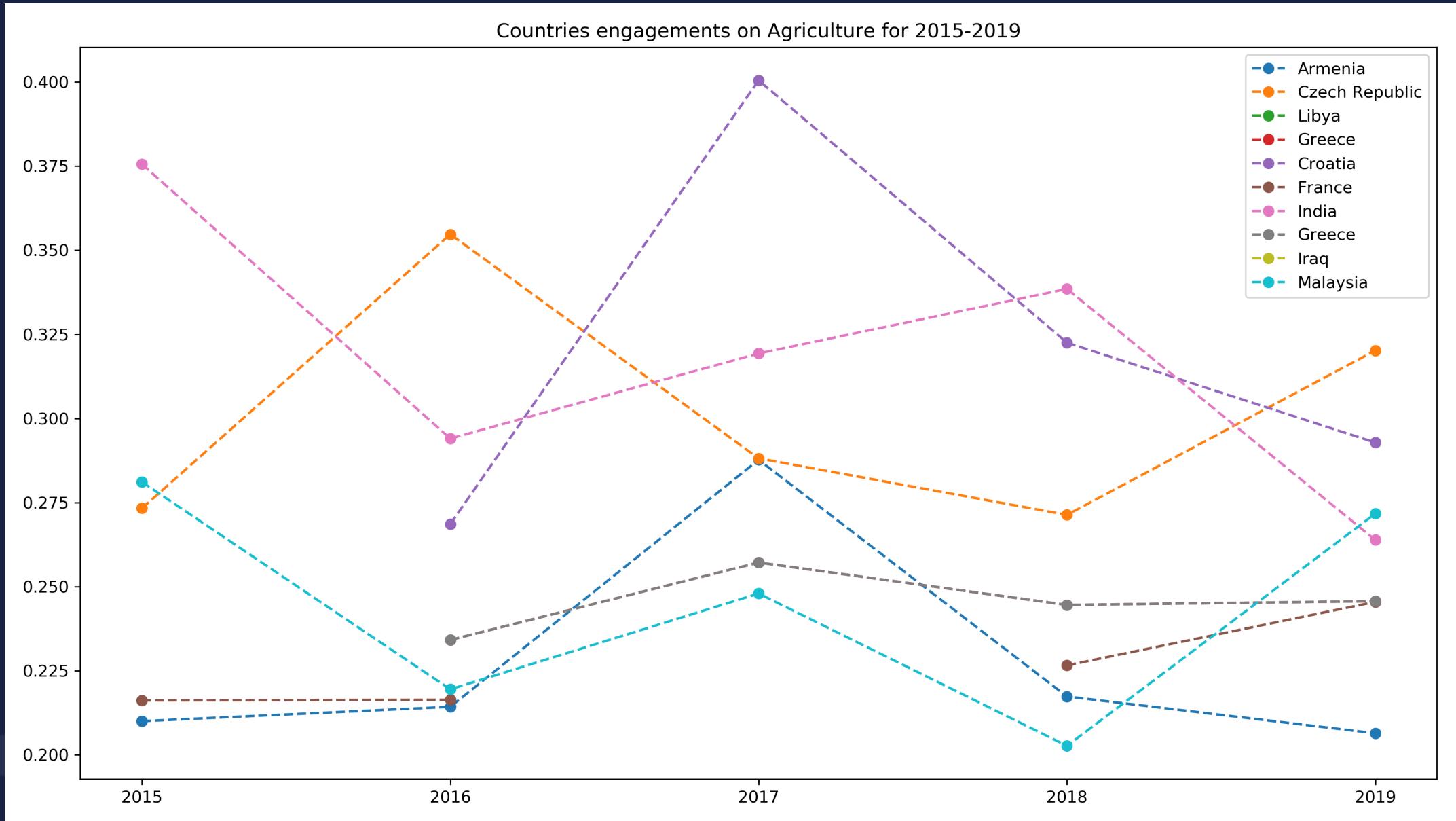
TRENDS OF ENGAGEMENTS



TRENDS WITH SIGNIFICANT CHANGE



TRENDS OF ENGAGEMENTS

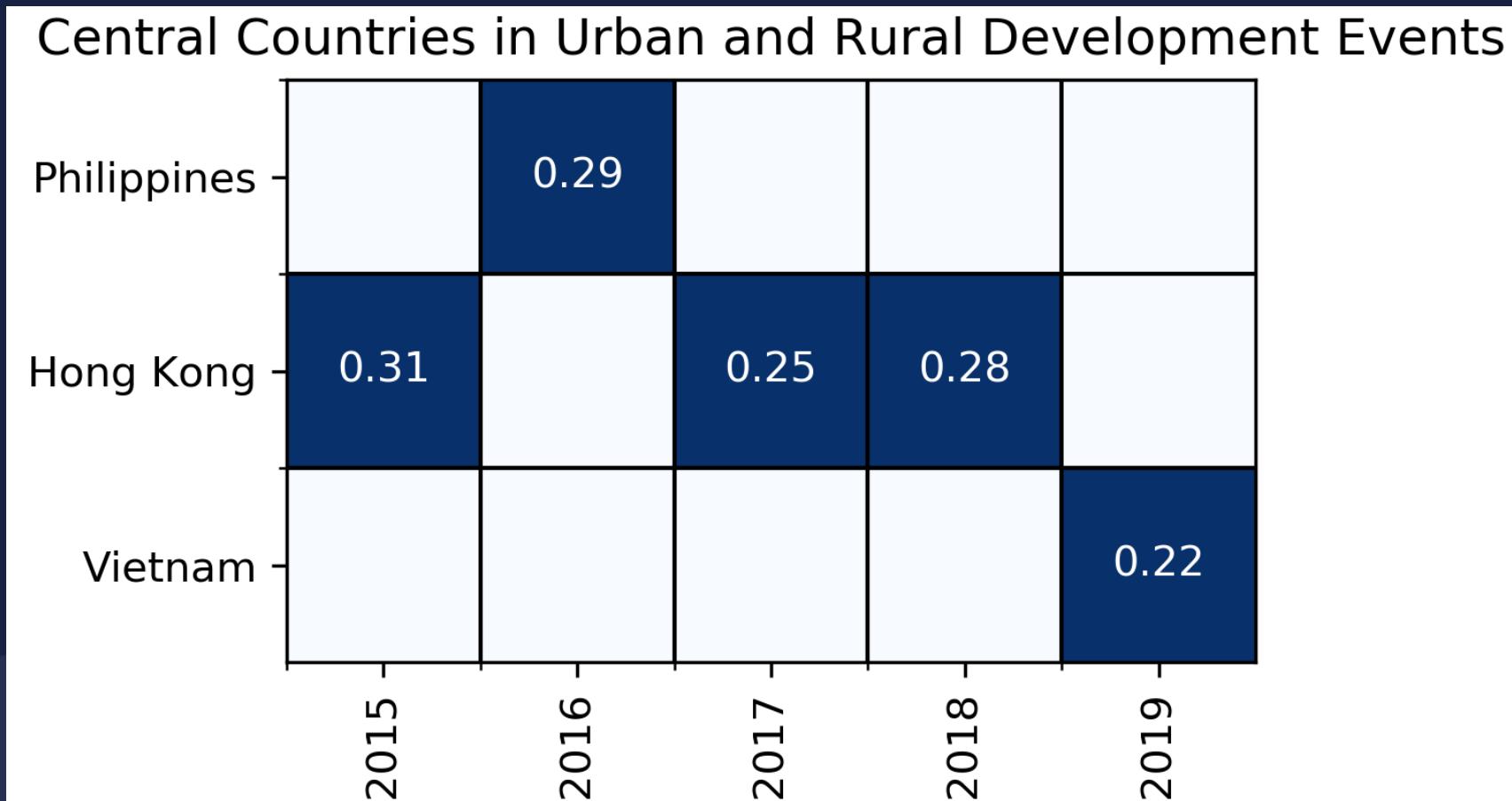


INFLUENCE: NETWORK CENTRALITY

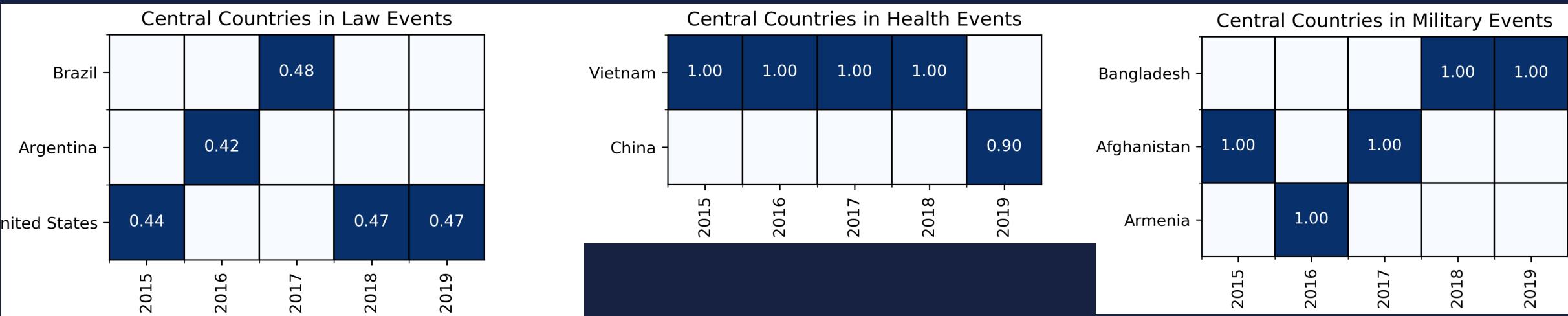
Network Centrality: Country with the highest prioritization to a theme of events based on the standardized count values.

INFLUENCE: NETWORK CENTRALITY

Network Centrality: Country with the highest prioritization to a theme of events based on the standardized count values.

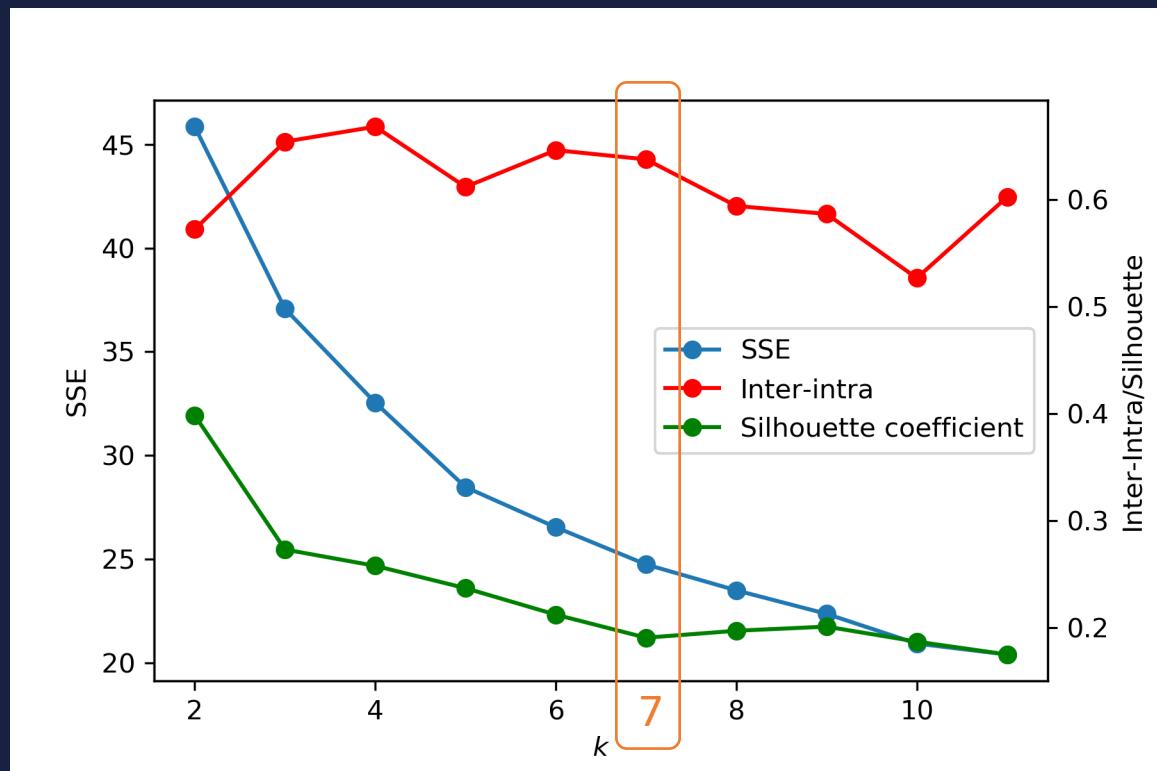


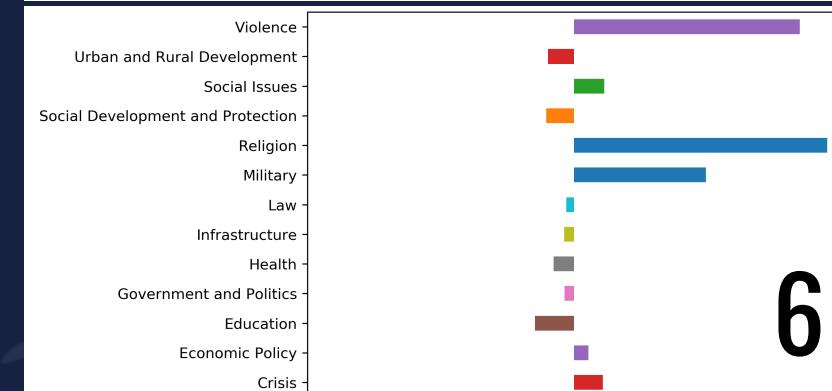
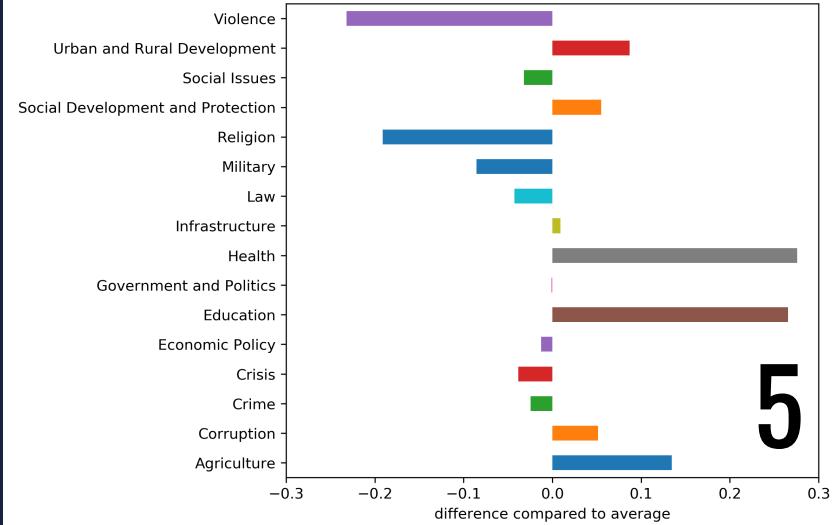
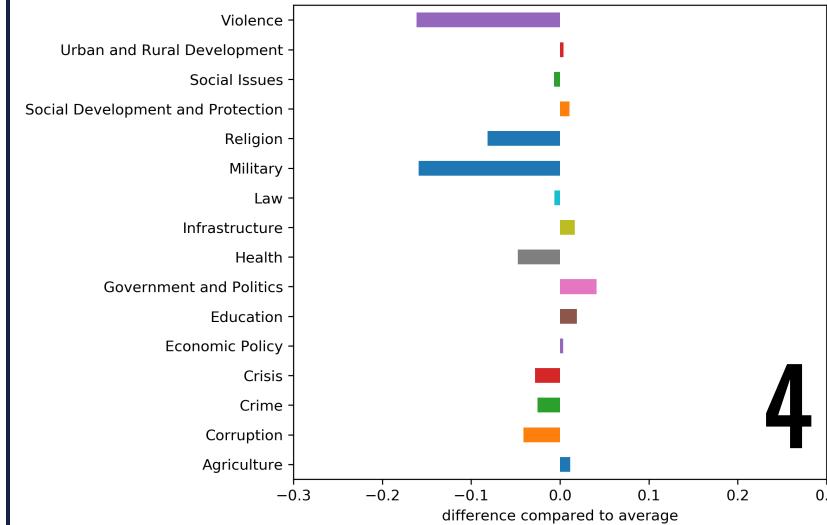
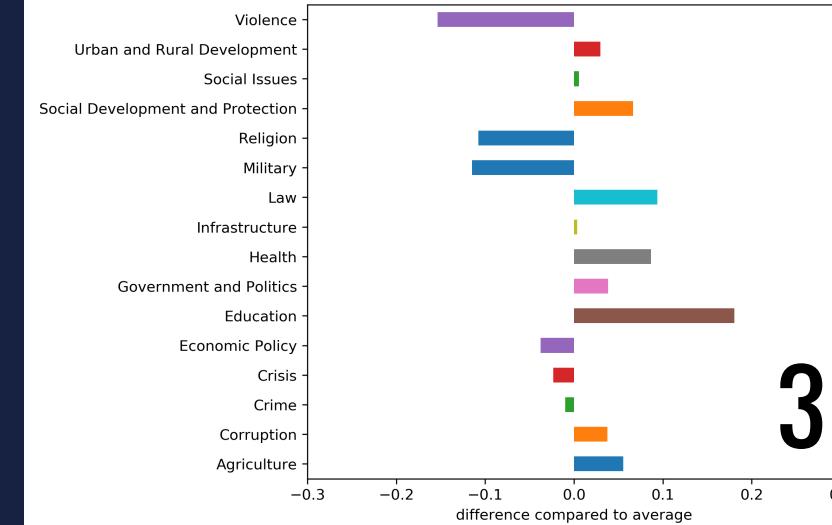
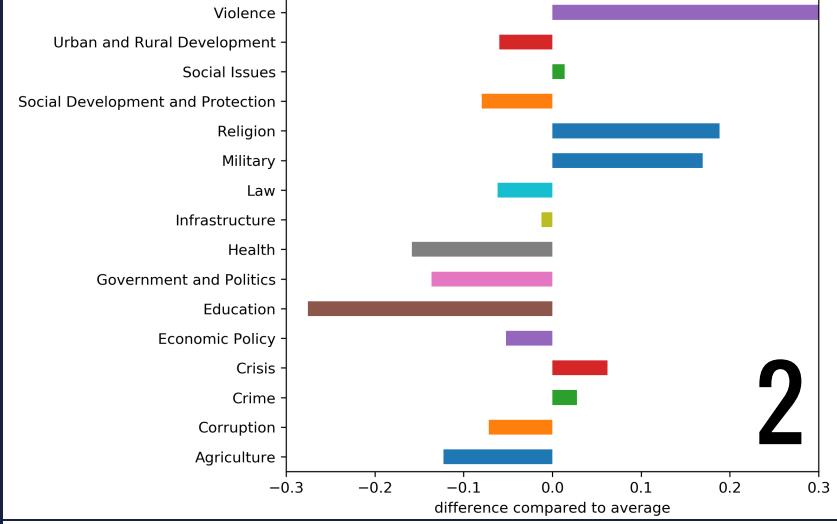
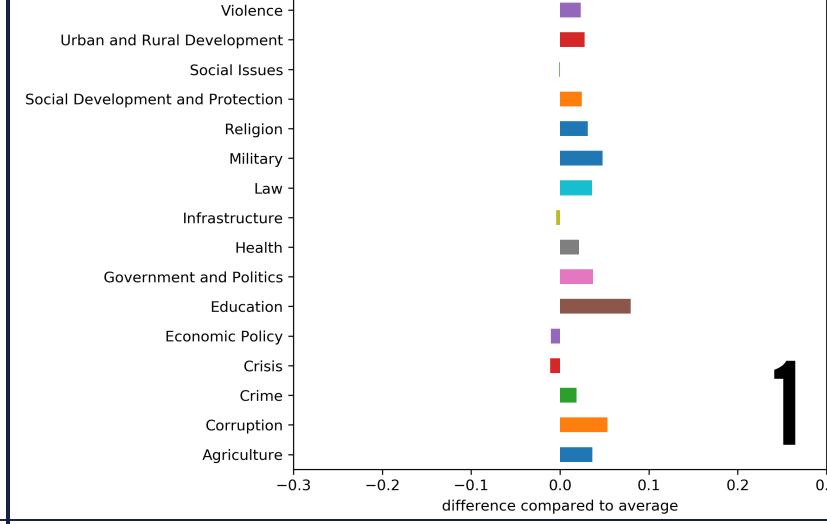
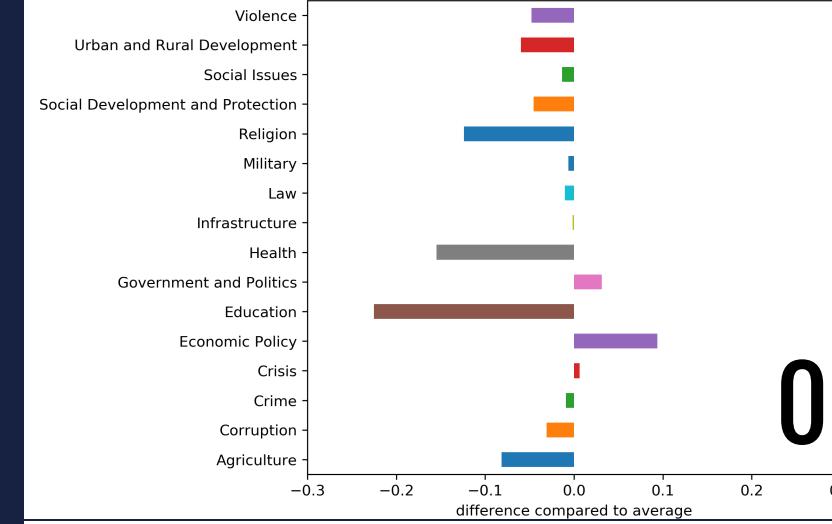
INFLUENCE: NETWORK CENTRALITY



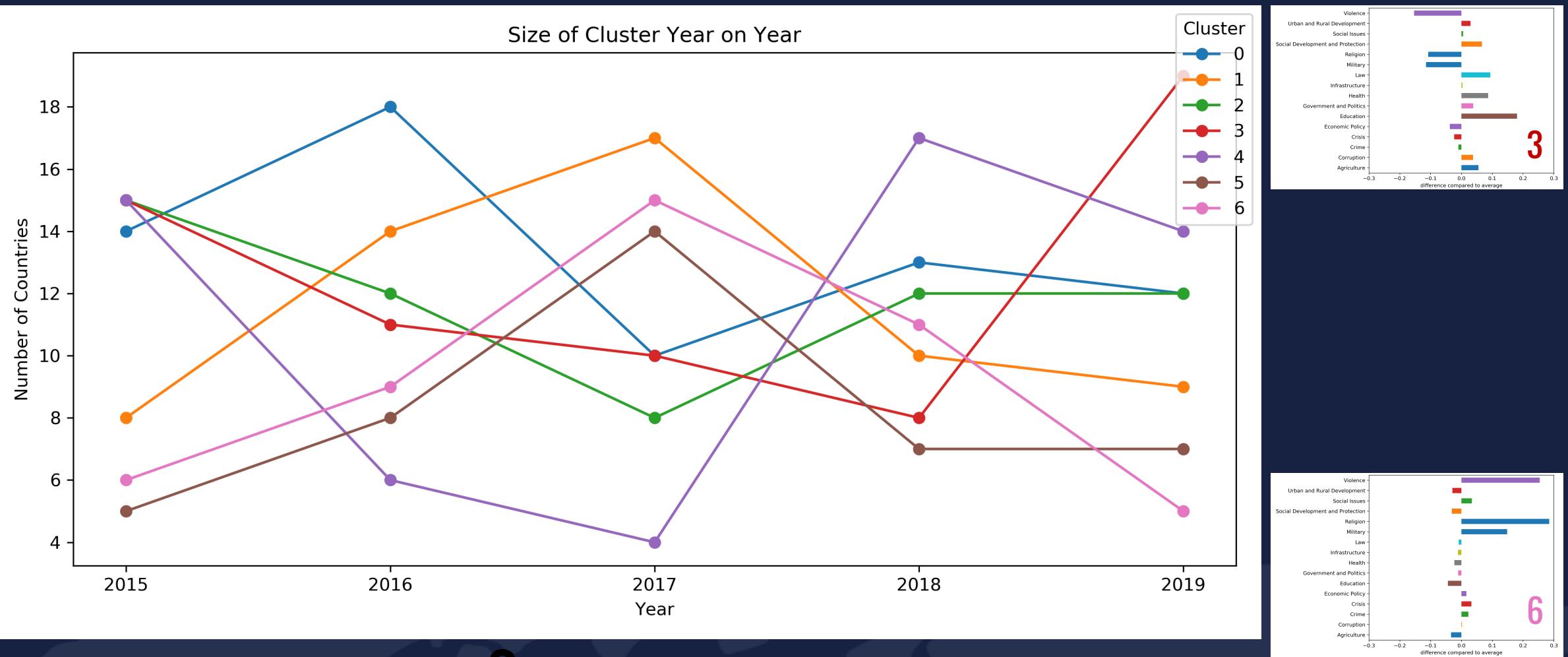
COMMUNITIES AND TREND ANALYSIS

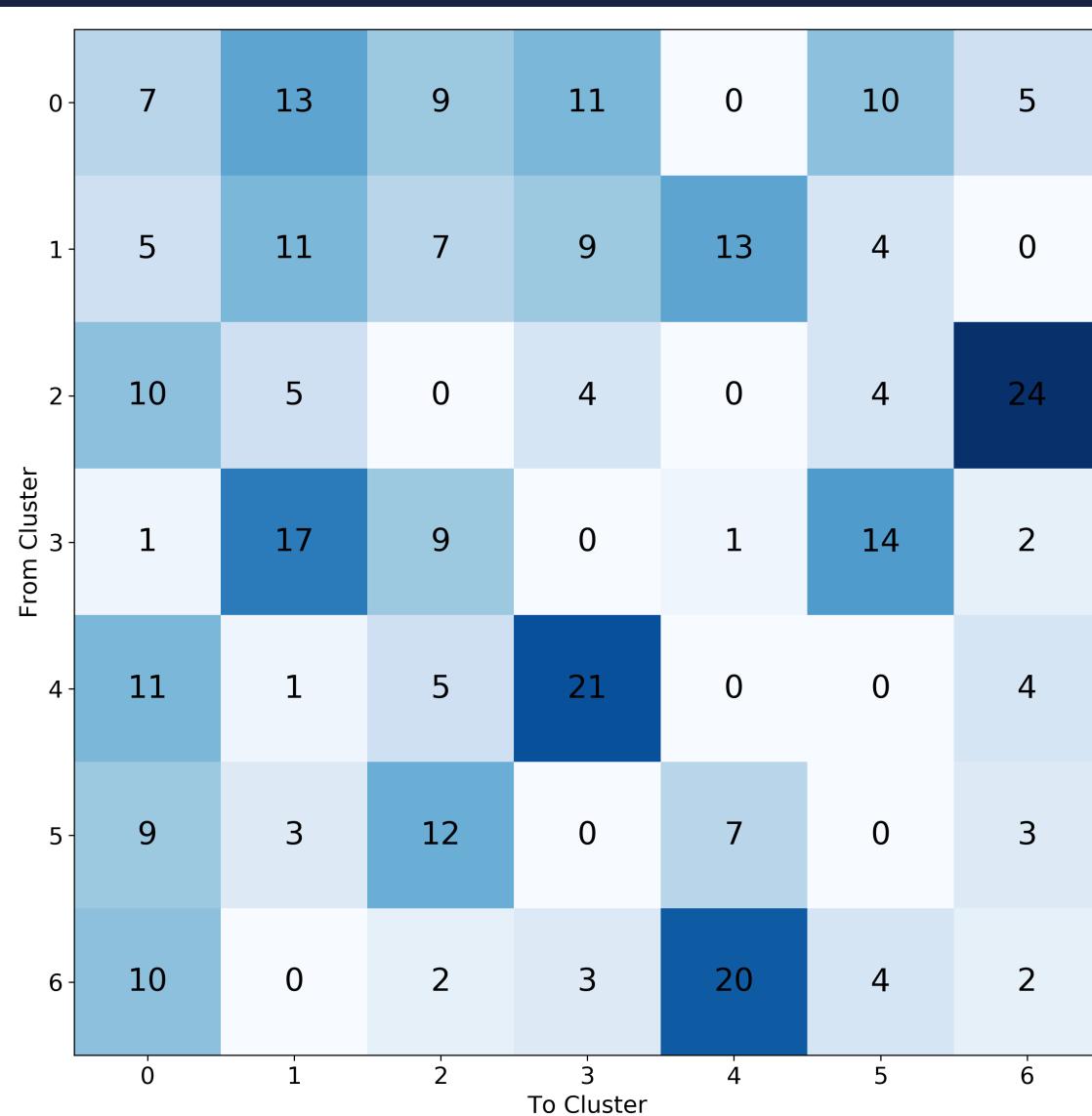
Community: A cluster of countries with similar behaviors
KMeans Clustering using standardized counts as features

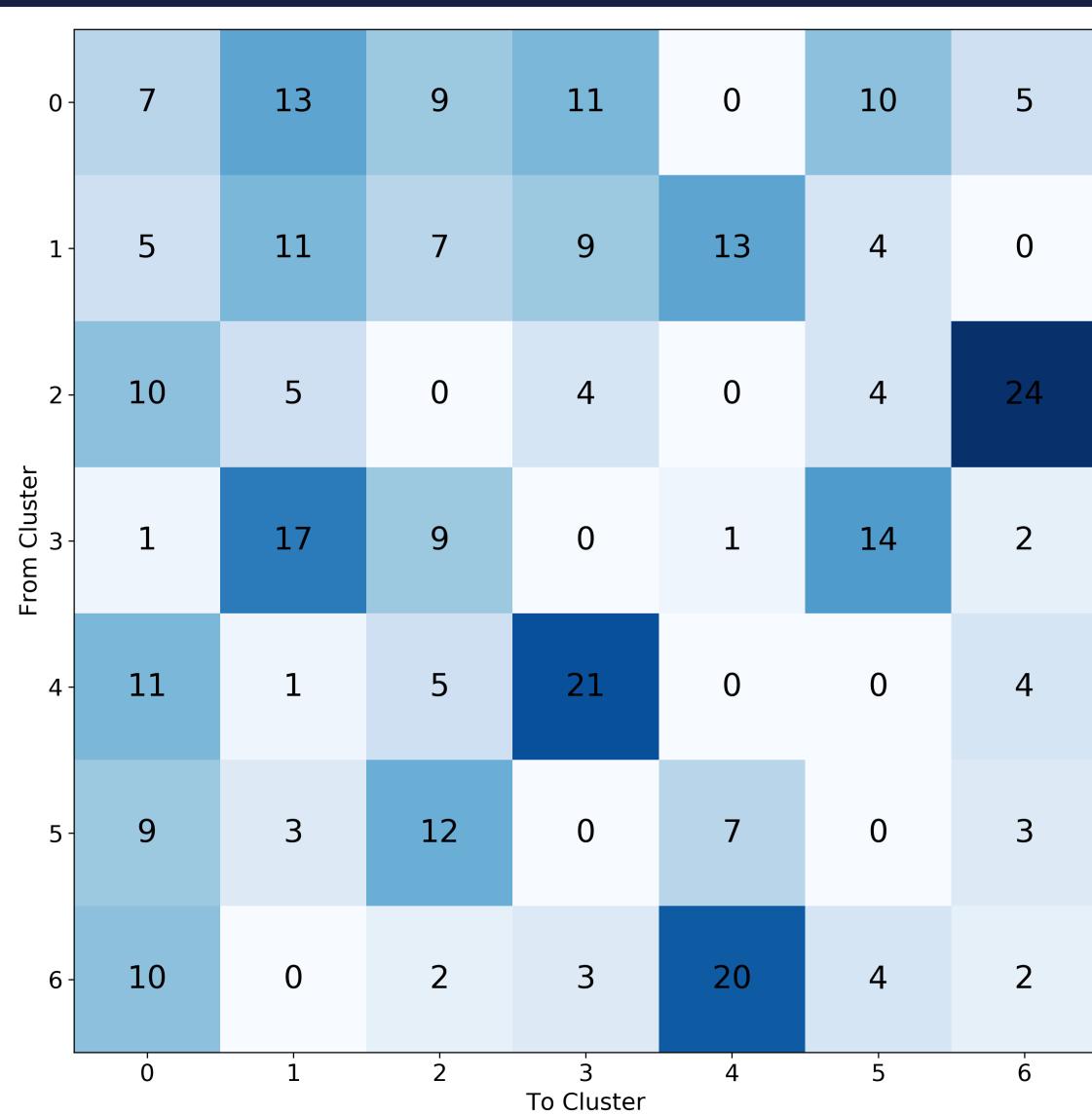




COMMUNITIES AND TREND ANALYSIS









FURTHER STUDY

Domain Expert

Prediction Modeling

Sentiment Analysis

Refined Identification of Interaction

Other ways to analyze community structures

THANK YOU

MAPPING GLOBALIZATION THROUGH EVENT THEMES

Tan, Jefferson; Nebres, Elisa Jr.; Legara, Erika; Monterola, Christopher
Asian Institute of Management, Philippines

Presented on CCS 2019, NTU, Singapore; September 30, 2019

EXTRACT RECORDS USING GOOGLE BIG QUERY

Row	GKGRECORDID	DATE	SourceCommonName	Counts	Themes	Locations
1	20170303050000-2210	20170303050000	kagstr.com	KIDNAP#2#EMERGENCY TEXAS DIVISION#2#Texas, United States#US#USTX#31.106#-97.6475#TX;	TAX_FNCACT;TAX_FNCACT_CHILD;KIDNAP_CRISISEX_CRISISEXREC_SECURITY_SERVICES;TAX_FNCACT_POLICE_CRISISEX_C07_SAFETY_TAX_DISEASE_TAX_DISEASE_ABDOMEN_TAX_ETHNICITY_TAX_ETHNICITY_BLACK_WB_1921_PRIVATE_SECTOR_DEVELOPMENT_WB_405_BUSINESS_CLIMATE_WB_2531_INSPECTIONS_LICENSING_AND_PERMITS_WB_2530_BUSINESS_ENVIRONMENT_LEGISLATION_EPU_POLICY_EPU_POLICY_LAW_WB_696_PUBLIC_SECTOR_MANAGEMENT_WB_840 JUSTICE_WB_1920_FINANCIAL_SECTOR_DEVELOPMENT_WB_328_FINANCIAL_INTEGRITY_WB_1014 CRIMINAL JUSTICE_WB_2082 LAW_ENFORCEMENT_TAX_FNCACT_OFFICIALS_CRISISEX_T01_CAUTION_ADVICE_WB_678_DIGITAL_GOVERNMENT_WB_694_BROADCAST_AND_MEDIA_WB_133_INFORMATION_AND_COMMUNICATION_TECHNOLOGIES;	3#Anderson County, Texas, United States#US#USTX#31.8168#-95.6669#1383786;2#Arkansas, United States#US#USAR#34.9513#-92.3809#AR;1#Israel#IS#IS#31.5#34.75#IS;2#Texas, United States#US#USTX#31.106#-97.6475#TX;
2	20170303050000-929	20170303050000	theadocate.com.au	null	TAX_FNCACT;TAX_FNCACT_ADMINISTRATOR,ELECTION_APPOINTMENT_WB_2670_JOBS_WB_1467 EDUCATION_FOR_ALL_WB_470 EDUCATION_WB_2131 EMPLOYABILITY_SKILLS_AND_JOBS_WB_1484 EDUCATION_SKILLS_DEVELOPMENT_AND LABOR_MARKET_TAX_ETHNICITY_TAX_ETHNICITY AUSTRALIANS;	1#Australia#AS#AS#-27#133#AS
3	20170303050000-626	20170303050000	inverelltimes.com.au	null	MEDIA_MSM_TAX_ECON_PRICE_CRISISEX_C07_SAFETY_UNGP_FORESTS_RIVERS_OCEANS_TAX_FNCACT;TAX_FNCACT_WOMAN_WB_678_DIGITAL_GOVERNMENT_WB_694_BROADCAST_AND MEDIA_WB_133_INFORMATION_AND_COMMUNICATION_TECHNOLOGIES;	1#South Africa#SF#SF#-29#24#SF
4	20170303050000-1860	20170303050000	wollondillyadvertiser.com.au	null	TAX_FNCACT;TAX_FNCACT_POLICE;TAX_FNCACT_CRIMINAL;	4#Illawarra, Victoria, Australia#AS#AS07#-37.0667#142.7#-1579342.4#Sydney, New South Wales, Australia#AS#AS02#-33.8833#151.217#-1603135;4#Woonona, New South Wales, Australia#AS#AS02#-34.3333#150.917#-1610869
5	20170303050000-1930	20170303050000	theadocate.com.au	null	TAX_FNCACT;TAX_FNCACT_POLICE;TAX_FNCACT_CRIMINAL;	4#Illawarra, Victoria, Australia#AS#AS07#-37.0667#142.7#-1579342.4#Sydney, New South Wales, Australia#AS#AS02#-33.8833#151.217#-1603135;4#Woonona, New South Wales, Australia#AS#AS02#-34.3333#150.917#-1610869
6	20170303050000-1816	20170303050000	engadget.com	null	KILL_CRISISEX_CRISISEXREC_CRISISEX_T02_INJURED_CRISISEX_T03_DEAD;	1#United Kingdom#UK#UK#54#-2#UK
7	20170303050000-1401	20170303050000	sheknows.com	KILL#2014#rat_x#4#Milan, Lombardia, Italy#IT#IT09#45.4667#9.2#-121726#NATURAL_Disaster_DROWNED#2014#rat_x#4#Milan, Lombardia,	TAX_RELIGION;TAX_RELIGION_CHRISTIAN_TAX_ETHNICITY_TAX_ETHNICITY_CHRISTIAN_TAX_WORLDLANGUAGES;TAX_WORLDLANGUAGES_SIRIANO_KILL_NATURAL_Disaster_NATURAL_Disaster_DROWNED;TAX_WORLDMAMMALS;TAX_WORLDMAMMALS_RAT;	2#New York, United States#US#USNY#42.1497#-74.9384#NY;4#Milan, Lombardia, Italy#IT#IT09#45.4667#9.2#-121726;4#London, London, City Of, United Kingdom#UK#UKH9#51.5#-0.116667#-2601889

AGGREGATE THE RECORDS

Group by YEAR, COUNTRY, and THEME

Location 1
Location 2
Location 3



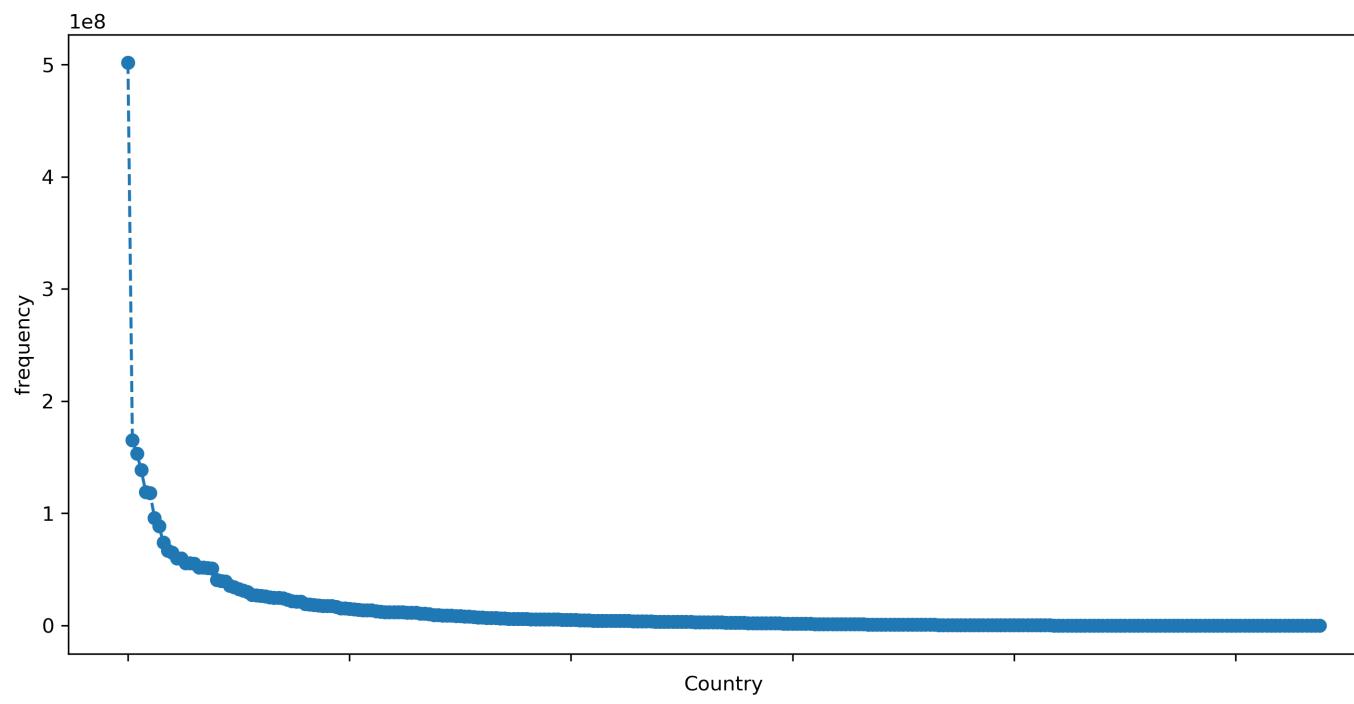
News Article/Record 1



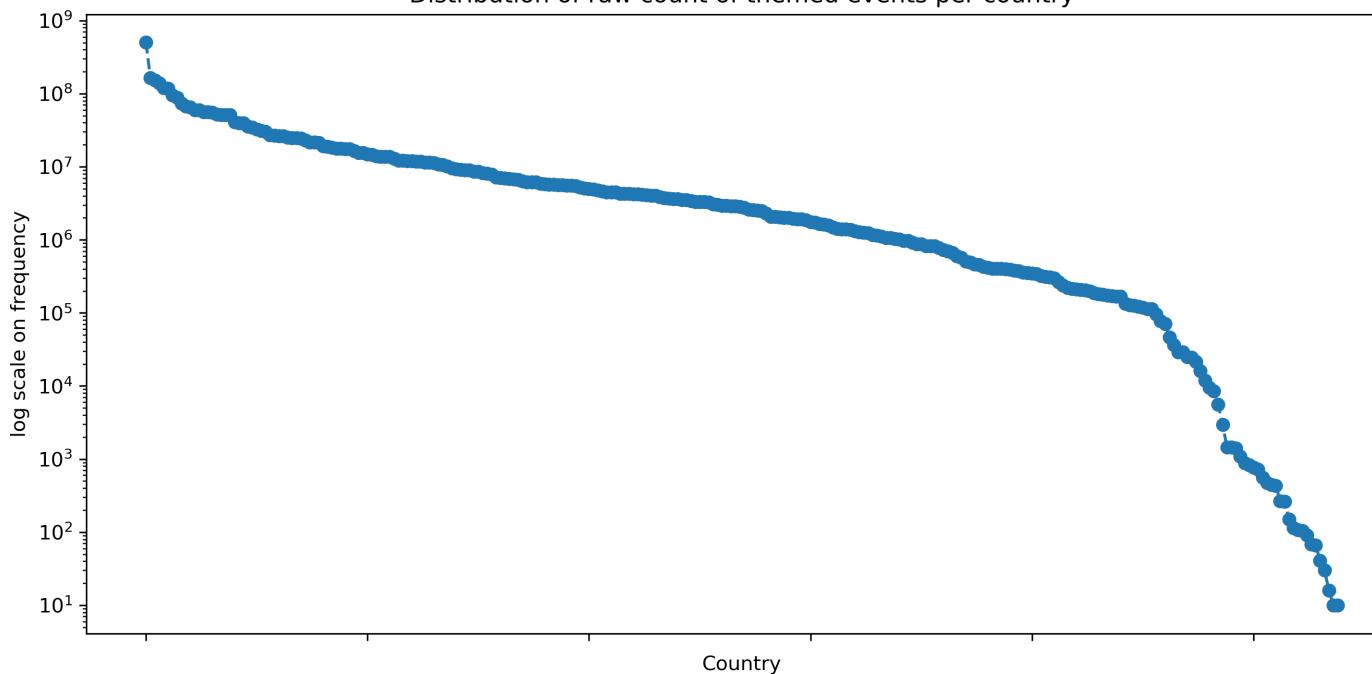
Theme 1
Theme 2
Theme 3

ID	Themes	Locations
Record 1	Theme 1	Location 1
Record 1	Theme 1	Location 2
Record 1	Theme 1	Location 3
Record 1	Theme 2	Location 1
Record 1	Theme 2	Location 2
Record 1	Theme 2	Location 3

Flattened Locations	Flattened Themes	Count
Location 1	Theme1	1
Location 1	Theme 2	1
Location 1	Theme 3	1
Location 2	Theme1	1
Location 2	Theme 2	1
Location 2	Theme 3	1



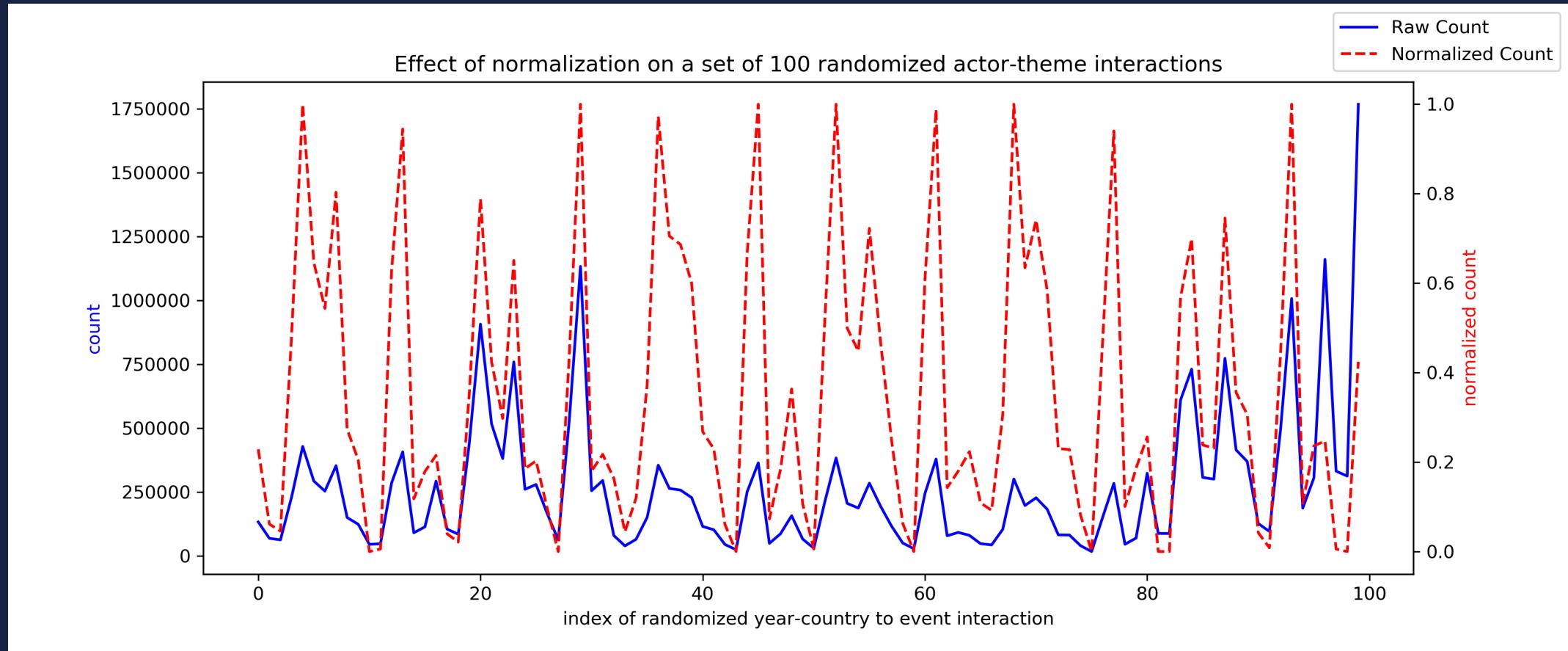
Distribution of raw count of themed events per country



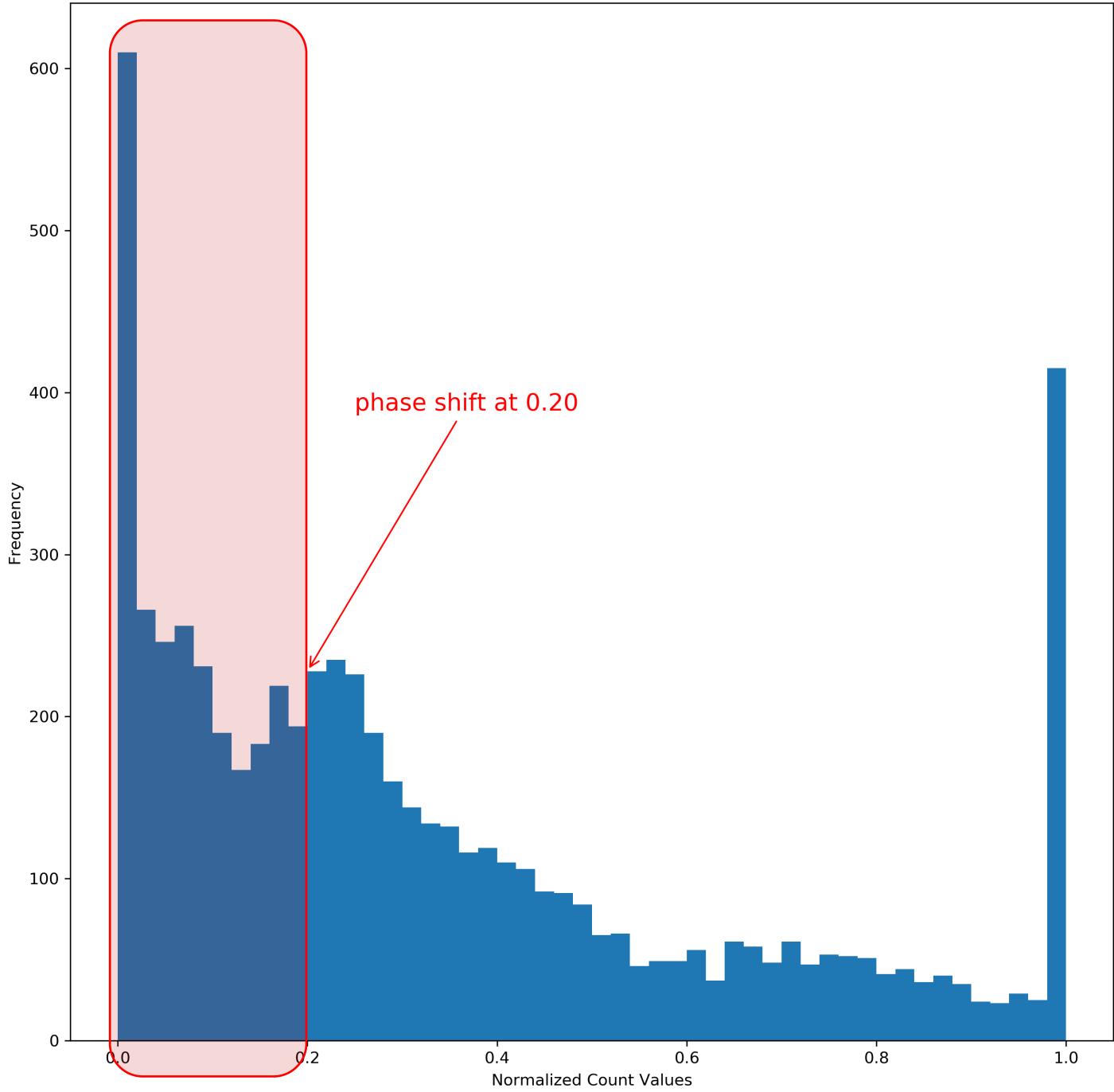
Filter countries based on the distribution of raw mentions of themes.

1. Countries with little media coverage
2. Countries that may not really be countries (such as ocean territories, etc.)

EFFECT OF NORMALIZATION

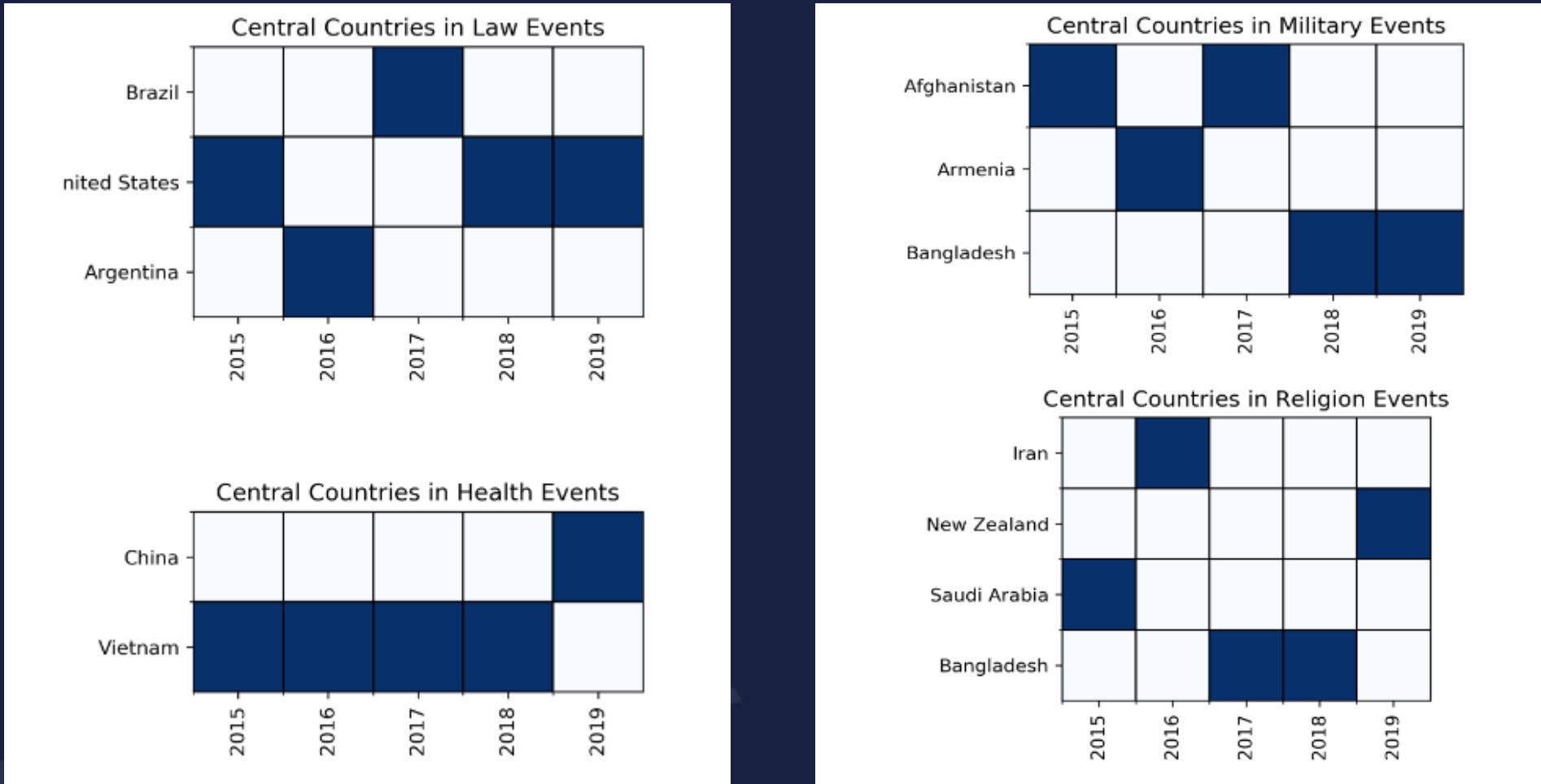


Distribution of norm counts



MOVING OF WEAK CONNECTORS

Results



Results

