

# Under The Same Flag

## Exploring Protest Fragmentation With Search Query Data

Kashmiri Medhi  
UT Dallas

Anton Sobolev  
UT Dallas



PaCSS,  
Aug 2023

# Motivation

## Why do some protest campaigns succeed while others fail?

Structural factors: Economy, Technology, Regime, Demography

Strategy & Tactics: Mobilization, Violence, Collective Identity

Internal Factors: Leadership, Organization, Agenda, Unity

- Understudied
- Weak results

## Challenges [Quantification]

Access to relevant information: timely events, surveys are costly

Measurement error

Sensitive survey questions ↗ self-selection bias & “dishonesty” bias

Media bias

Misclassify multiple campaigns as a single entity

Comparability: cross-country / cross-campaign is a concern

# This Paper

- Protest Fragmentation**
  - Goals / Motivations Perspective
  - Do scholars mistakenly categorize de-facto separate campaigns as a single entity? [*"Under The Same Flag" Hypothesis*]
- Current Goal:** Method to estimate **campaign fragmentation\***
- Desired Properties**
  - Behavior-based measure:** media reports, surveys, expert opinions
  - Explicit interpretation:** Likert scale, composite measures [*Polity IV*]
  - Comparability:** cross-country / cross-campaign comparison

\* **Campaign Fragmentation** – variation in the goals / motivations of a protest campaign among protesters

# Focus

- **Sub-national differences in the demand for information related to a protest campaign**

**Assumption:** High variation ↗ High protest fragmentation

- **Correlated behaviors**

**Key idea:** Individuals who look for the same information related to a protest campaign share similar views regarding the goals of this campaign

- **Implementation**

**Input data:** Search queries (Google Trends)

**Key feature:** Ability to identify *other* search queries individuals conduct when they seek for protest-campaign information

# How scholars measure protest fragmentation

## Surveys

- General Population: Activists underrepresented  
Behavior: Reported ≠ Observed  
Self-selection | 'Dishonesty' bias  
List-experiments are hard | Costs
- Protesters: Hard timing (!) | Safety concerns  
Locations underrepresented | Costs



## 1. Identify Protesters

웃웃 웃웃웃  
웃 웃웃 웃

## 2. Identify Differences

웃웃 웃웃웃  
Corruption  
Ethnic Discrimination

# How scholars measure protest fragmentation

## □ Event Cataloging

Documents: Proclaimed goals ≠ Actual goals  
Leadership perspective  
Spontaneous collective action

News: Bias in media reporting  
'Black Box'

Multimedia: Representativeness  
Reporting Bias | Costs  
Leadership perspective

## The New York Times

So Many Faces Make Up One Crowd



A scene from the documentary filmed during protests in Ukraine. Cinema Guild



## { 1. Identify Protesters

웃웃 웃웃웃  
웃 웃웃 웃

## { 2. Identify Differences

웃웃 웃웃웃  
Corruption  
Ethnic Discrimination

# How scholars measure protest fragmentation

□ User-generated content

## 1. Identify Protesters

웃웃 웃웃웃  
웃 웃웃 웃

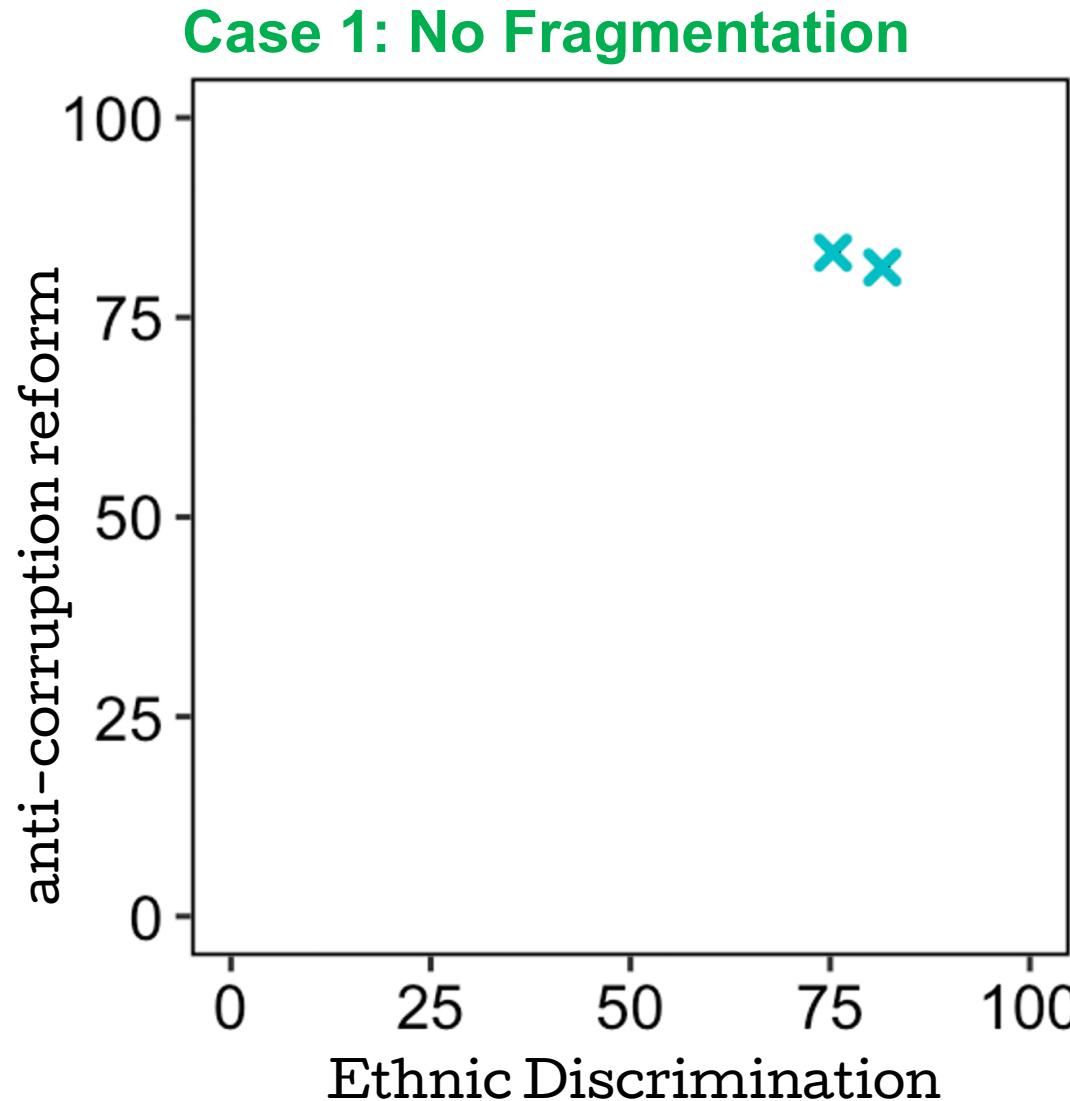
## 2. Identify Differences

웃웃 웃웃웃  
Corruption Ethnic  
Ethnic Discrimination

# Theory

- In location  $x_i$ , individuals who look for
  - └ protest campaign information *also* search:
    - └ “revolution”
    - └ “anti-corruption reform”

**Similar interest in both topics**  
→ unified protest campaign

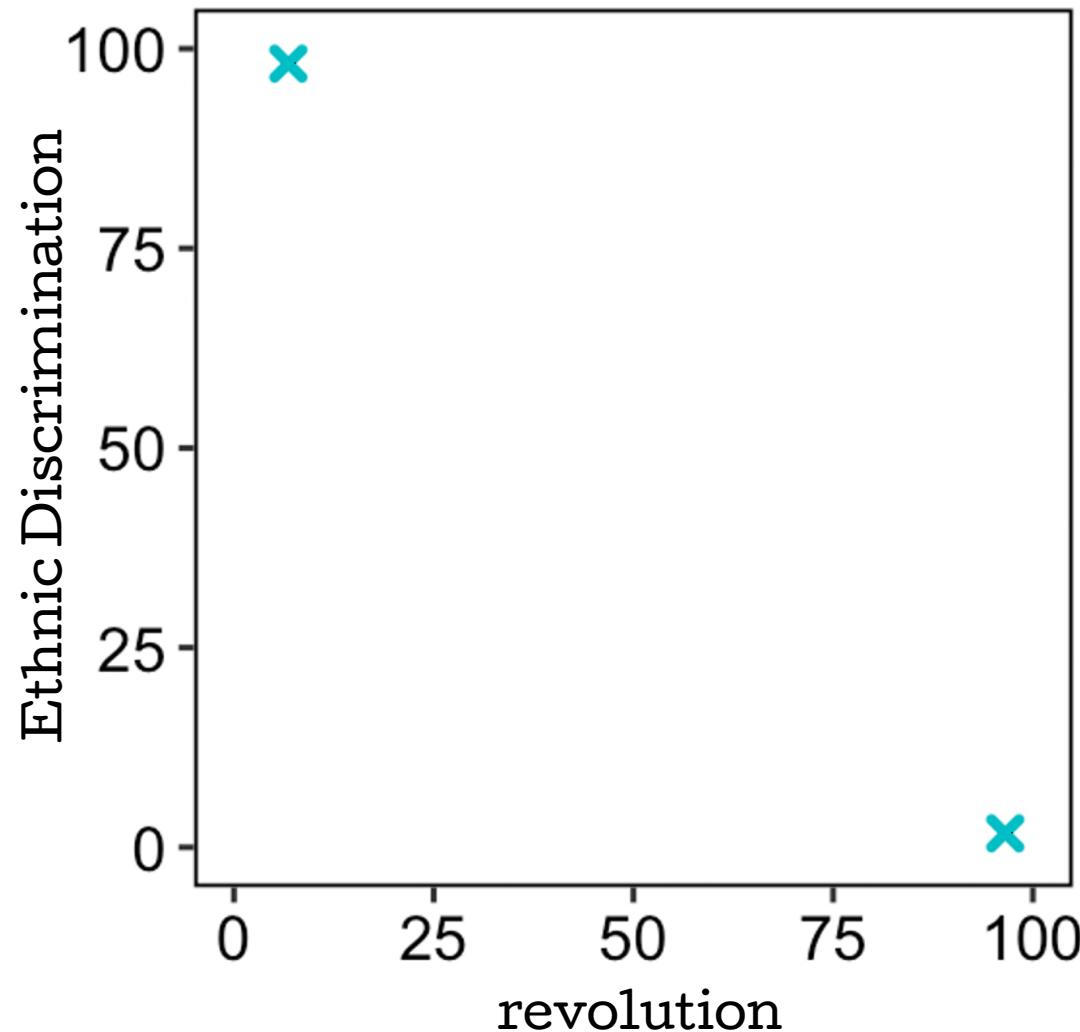


# Theory

- In location  $x_i$ , individuals who look for protest campaign information also search:
  - └ “revolution”
  - └ “anti-corruption reform”

**Dissimilar interest in topics**  
⇒ **High protest campaign fragmentation**

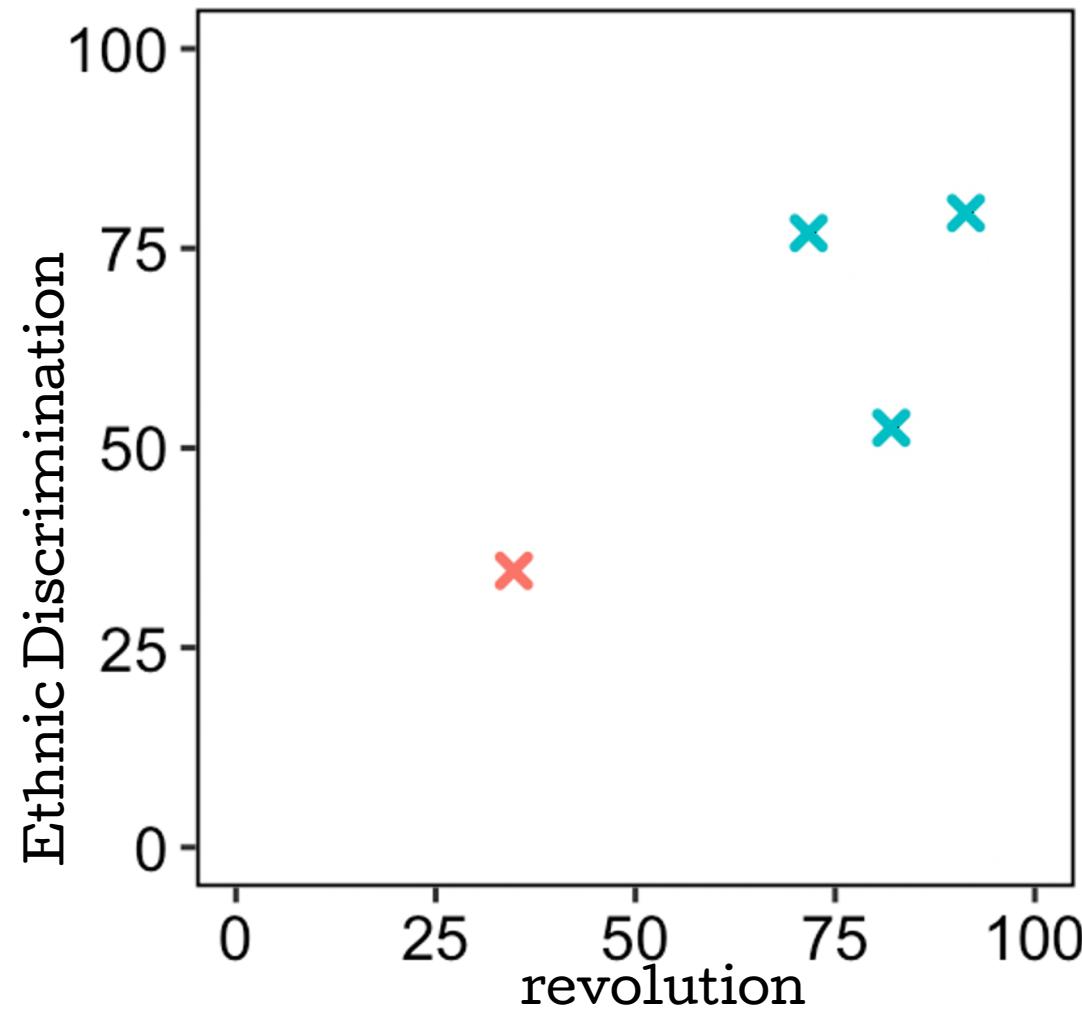
**Case 2: High Fragmentation**



# Theory

- In location  $x_i$ , individuals who look for protest campaign information *also* search:
  - “revolution”
  - “anti-corruption reform”

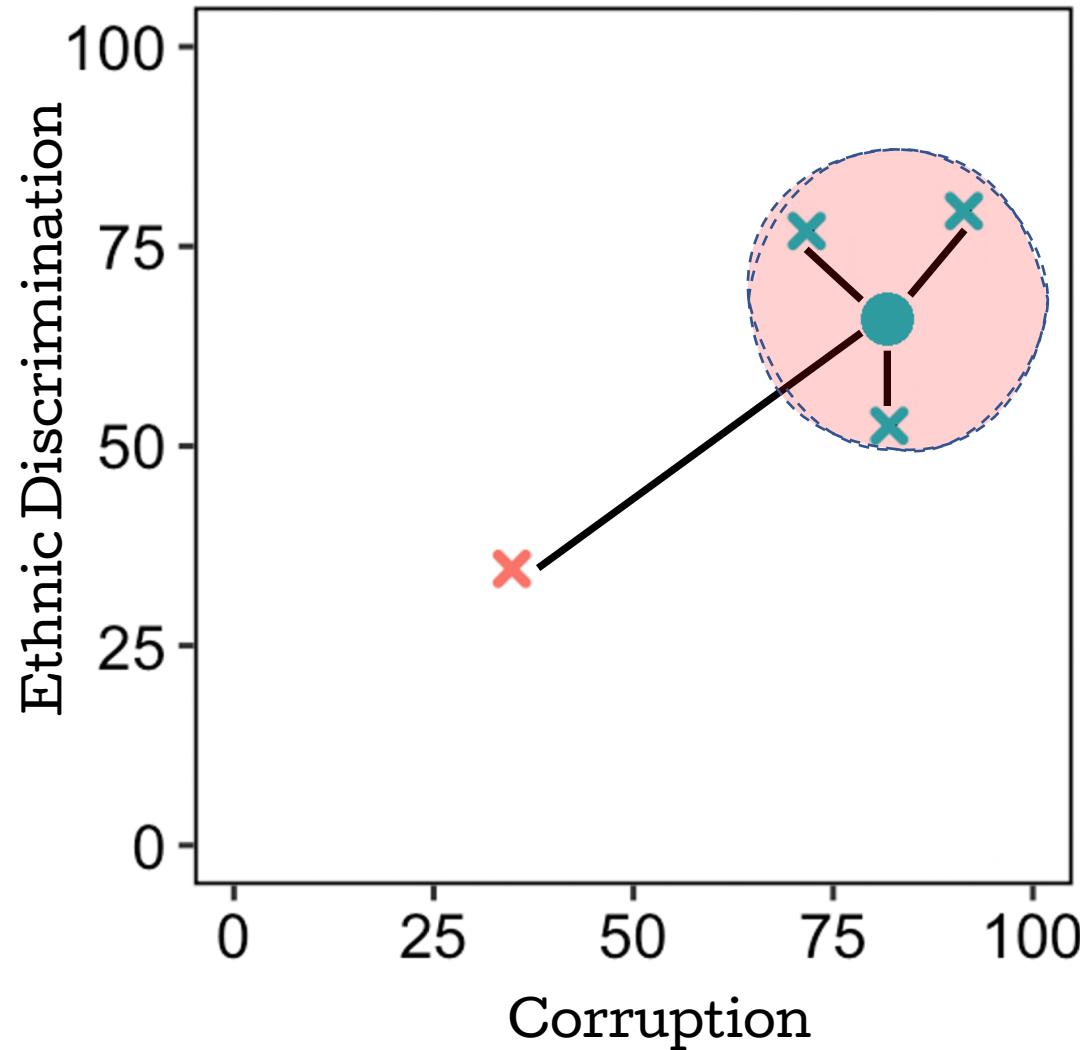
Case 3: Multiple Locations



# Theory

- In location  $x_i$ , individuals who look for protest campaign information *also* search:
  - “revolution”
  - “anti-corruption reform”
  
- Proposed Approach**
  - Identify the largest cluster [robust to outliers]
  - Calculate cluster’s centroid [*n*-dimensional space]
  - Fragmentation Score:** average distance to the centroid [Manhattan distance]

Case 3: Multiple Locations



# Search Queries



Initial Q: How do people search for protest-related information?

- Identify: Search queries by a keyword  
Similar queries

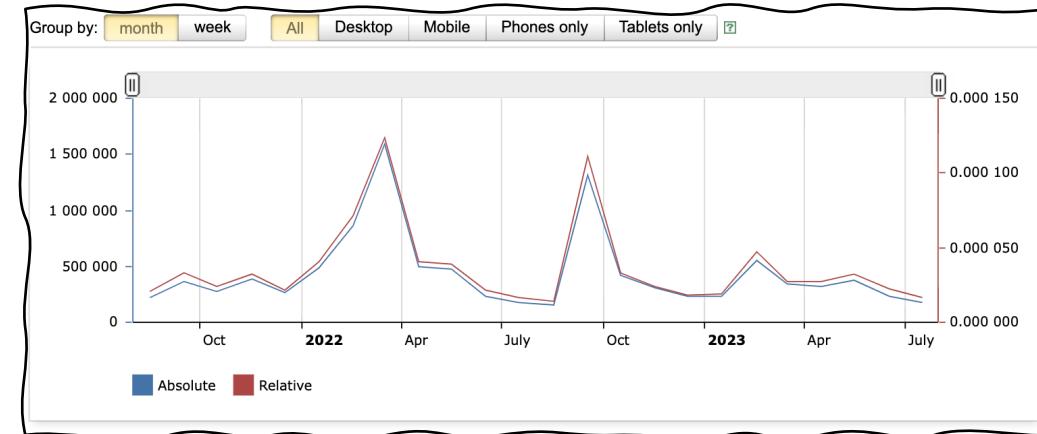
- Classification method: Clicking same websites

The screenshot shows the Yandex Keyword statistics interface. The search term entered is 'МИТИНГ'. The results are grouped by 'Statistics by keyword' and 'Displays per month'. The 'Displays per month' section lists the following data:

| Keyword          | Displays per month |
|------------------|--------------------|
| МИТИНГ           | 123,966            |
| митинг 2023      | 8,672              |
| МИТИНГ +в москве | 8,212              |
| митинг против    | 7,296              |
| митинги августа  | 6,935              |
| митинг сегодня   | 6,758              |

Below this, there is a section titled 'Requests, similar to «МИТИНГ»' which lists other search terms and their display counts.

Raw volume of search queries containing keyword



| Period                  | Absolute  | Relative          |
|-------------------------|-----------|-------------------|
| 01.08.2021 - 31.08.2021 | 216 811   | 0.000 020 736 036 |
| 01.09.2021 - 30.09.2021 | 363 898   | 0.000 033 018 338 |
| 01.10.2021 - 31.10.2021 | 277 206   | 0.000 023 673 519 |
| 01.11.2021 - 30.11.2021 | 391 215   | 0.000 032 698 807 |
| 01.12.2021 - 31.12.2021 | 265 974   | 0.000 021 635 545 |
| 01.01.2022 - 31.01.2022 | 491 261   | 0.000 040 198 758 |
| 01.02.2022 - 28.02.2022 | 867 256   | 0.000 071 654 695 |
| 01.03.2022 - 31.03.2022 | 1 587 907 | 0.000 123 160 982 |
| 01.04.2022 - 30.04.2022 | 494 952   | 0.000 040 661 098 |
| 01.05.2022 - 31.05.2022 | 471 760   | 0.000 039 129 490 |
| 01.06.2022 - 30.06.2022 | 237 453   | 0.000 021 874 484 |
| 01.07.2022 - 31.07.2022 | 172 360   | 0.000 016 349 320 |

Raw data over time

# Search Queries



## G-Trends: **Queries aggregated by topics**

- Identify:
  - “Co-related” queries | topics
  - `Protest topic` ~ `Related topics`

**Classification method:** Wikipedia  
[Reverse Engineering]



Anti-CAA Protests (India, 2019)

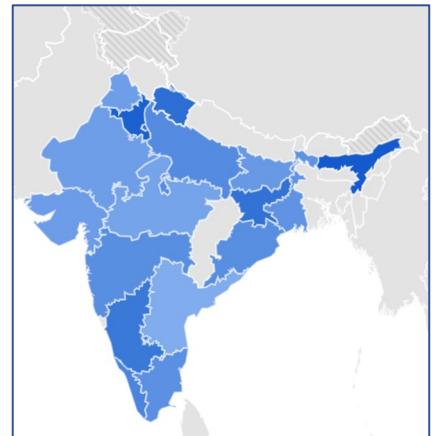
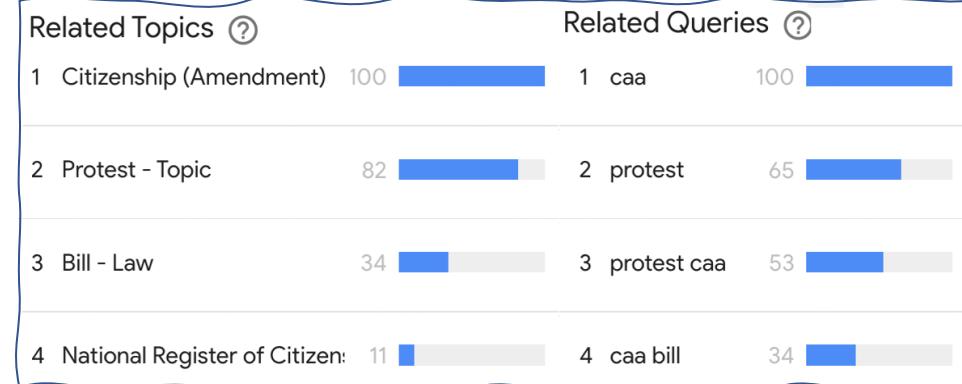
## Citizenship (Amendment) Bill protests

Citizenship (Amendment) Bill protests

Search term

Citizenship (Amendment) Bill protests

Topic



# Implementation with G-Trends

Example: Citizenship Amendment Act protests (India, 2019)

## Google trends

- Score [0-100] based on the volume of search queries
- Provides data for separate queries and queries aggregated into topics
- Identifies queries / topics correlated with the initial query / topic
- Subnational level data

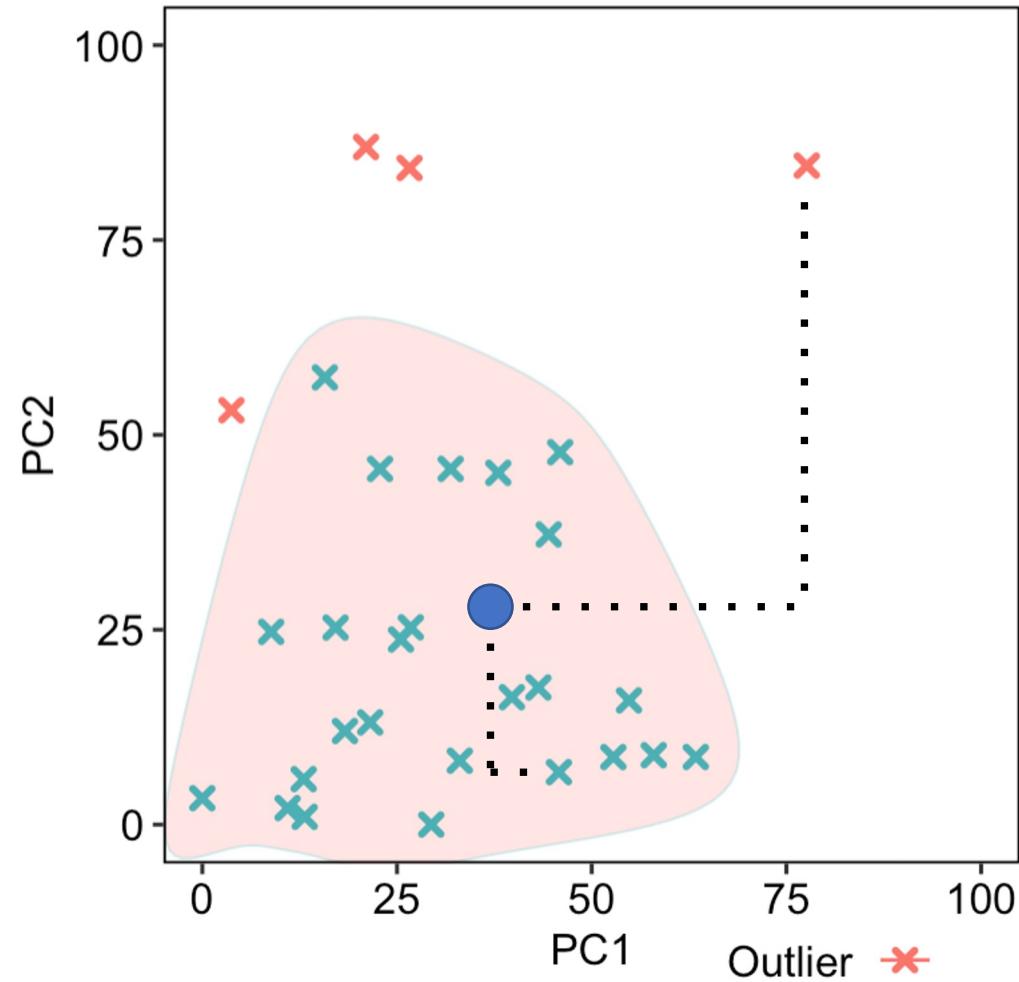
## Fragmentation due geographic variation

# Implementation with G-Trends

Example: Citizenship Amendment Act protests (India, 2019)

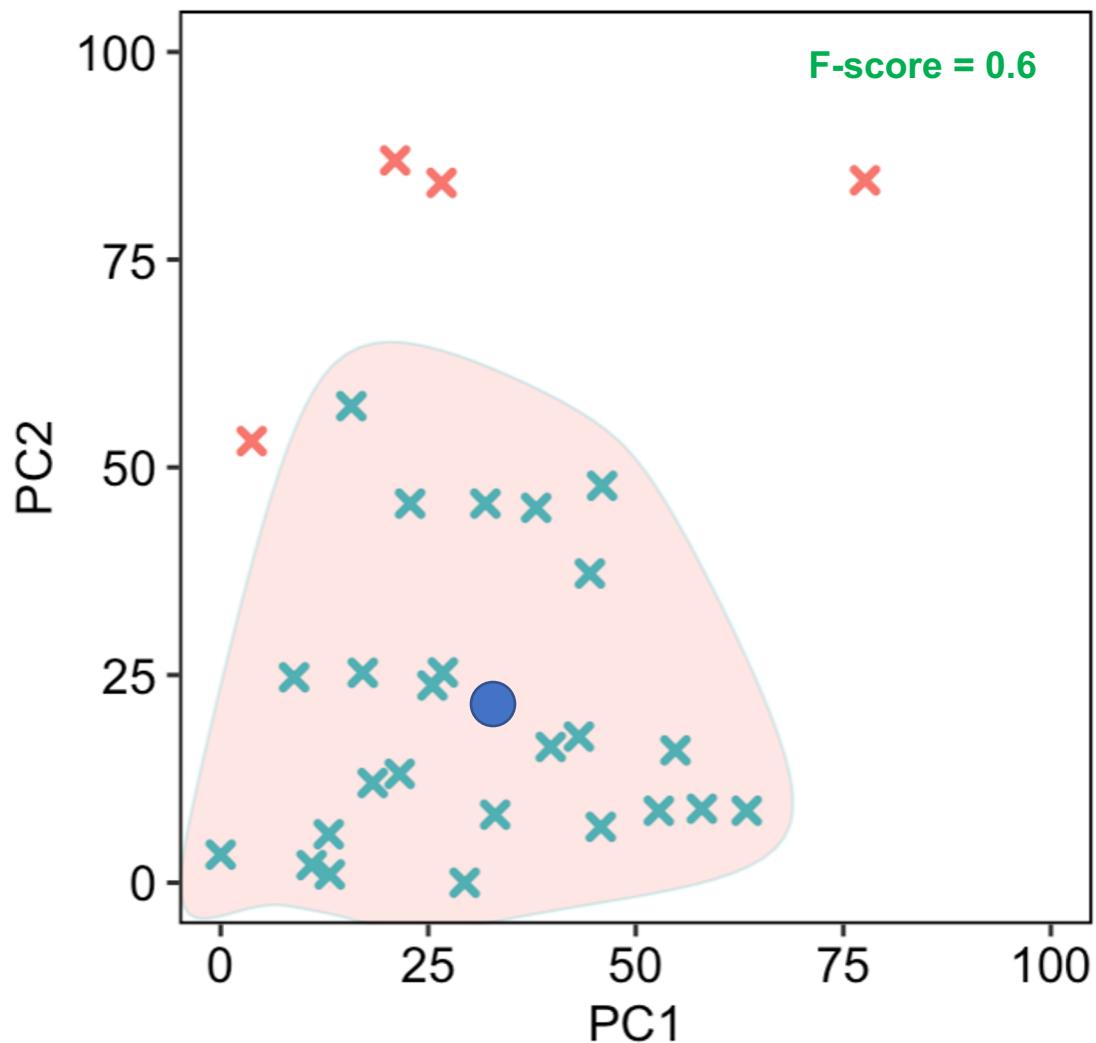
- Citizenship Amendment Act protests (India, 2019)**
  - Identify the protest movement topic
  - Identify first 10 correlated topics
  - Identify largest cluster [via DBSCAN]
  - Calculate centroid
  - Calculate mean distance  $D$  [Manhattan]
  - Adjust  $(100 - \frac{D}{2})/100$

**Fragmentation Score = 0.6**

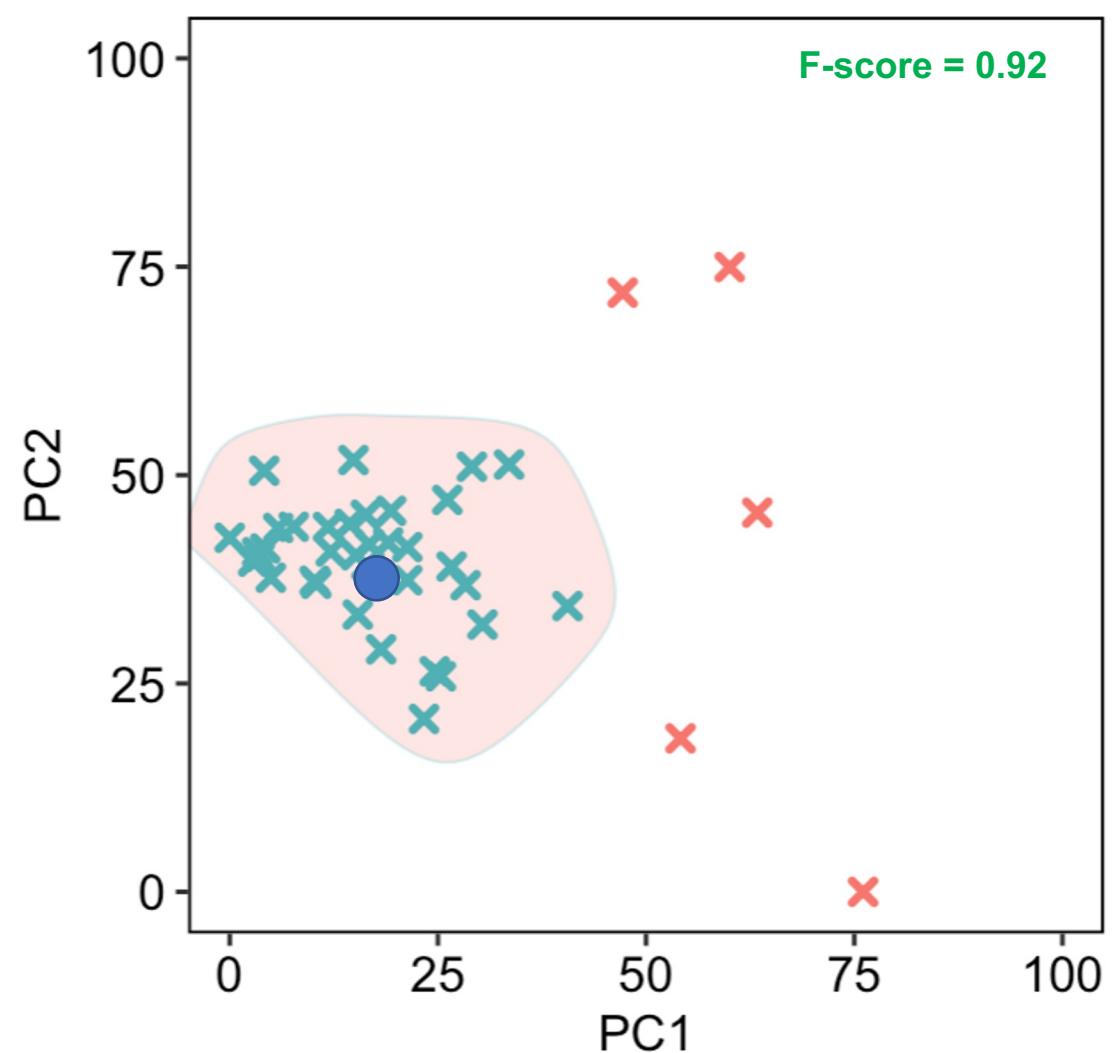


\* Principal Components are used for illustration purposes

# Comparison India-2019 VS US-2017

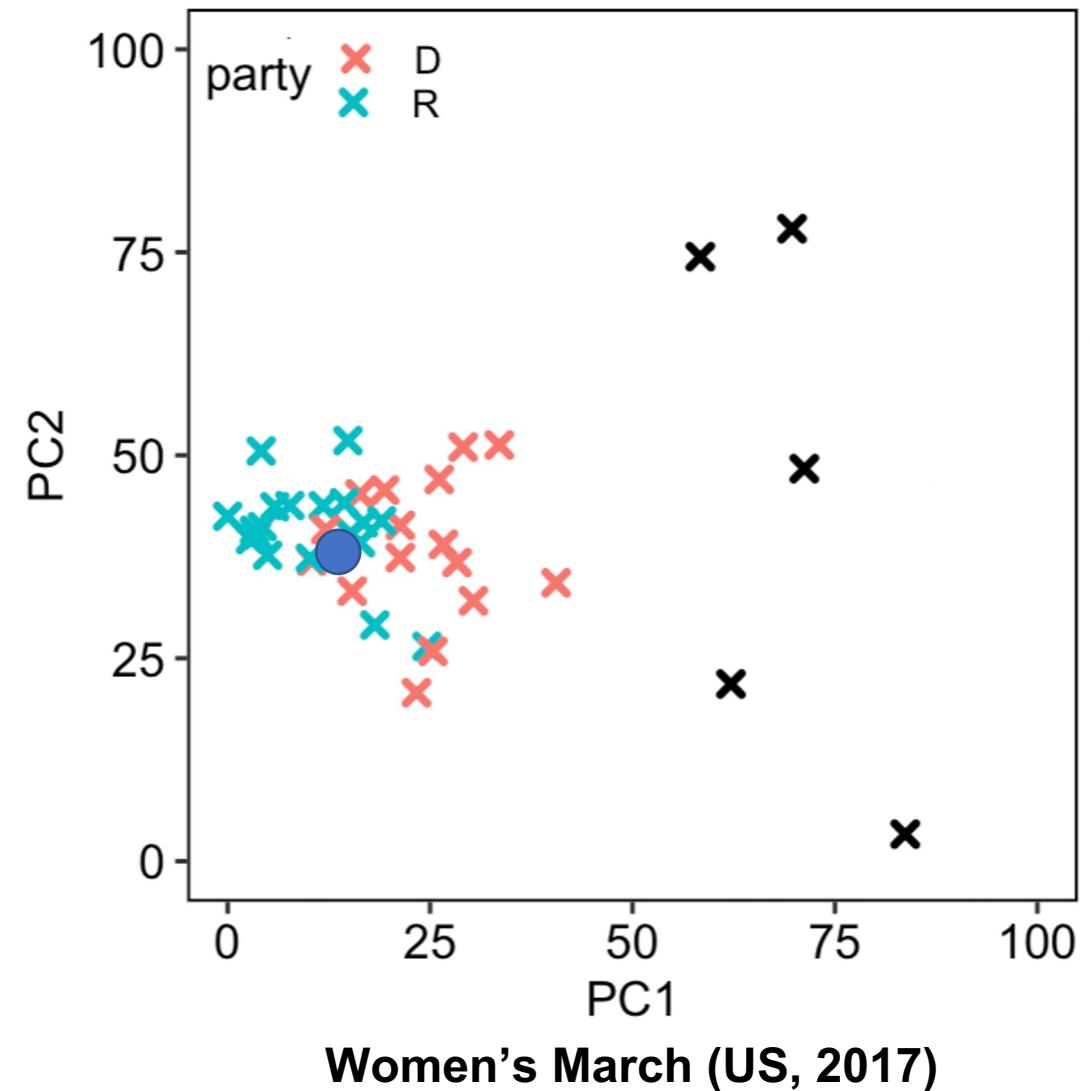
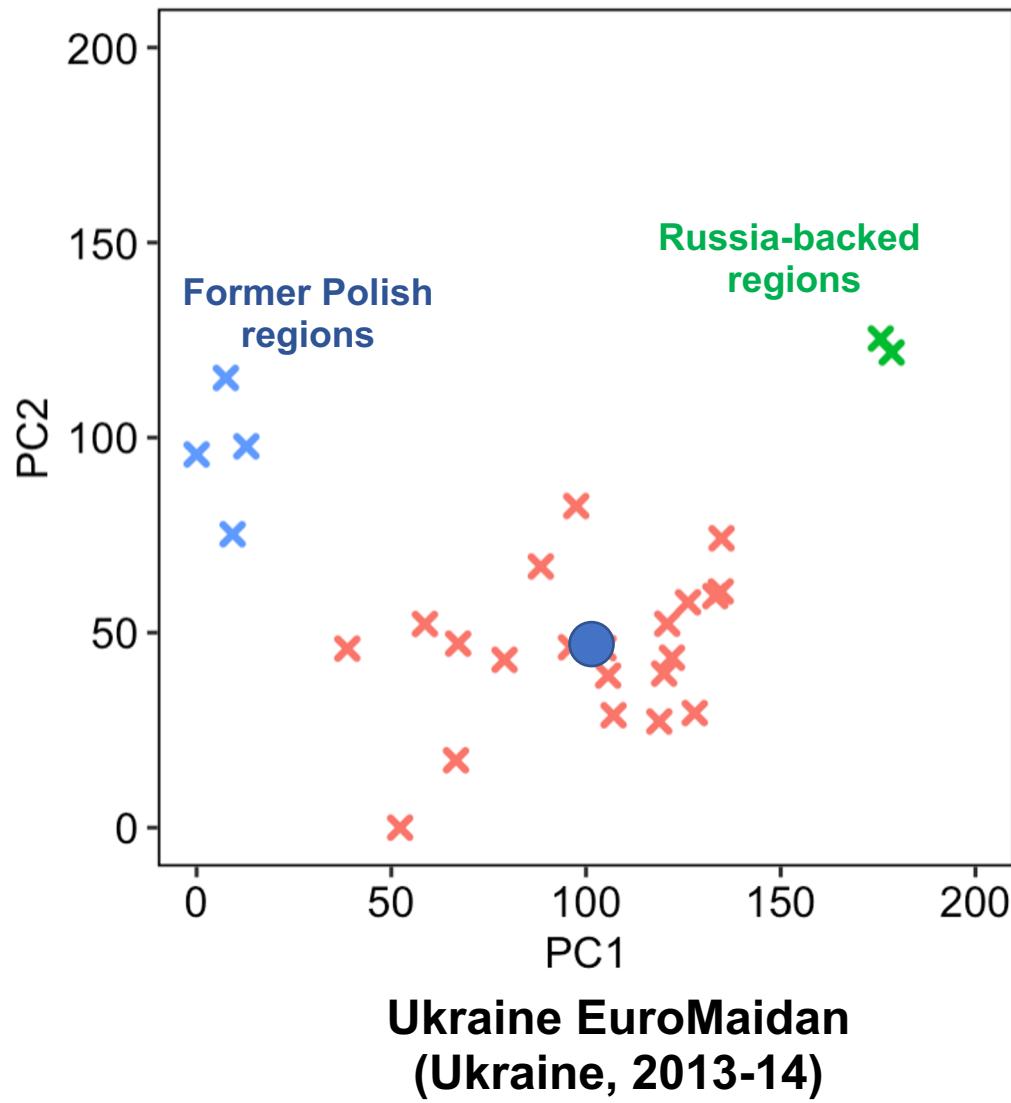


Citizenship Amendment Act  
protests (India, 2019)



Women's March (US, 2017)

# Validity Check Potentially Important Variables



# Concluding thoughts

Thank  
you!

## Not an ideal measure

- └ Benefits:
  - observed behavior
  - comparability
  - fine-grained data

## Applications

- └ Protest Fragmentation snapshot
- └ Solving misclassification issue (single VS multiple campaigns)
- └ Track temporal dynamics

## Issues

- └ Internet coverage: Adjust for the region population / internet users?
- └ Specific Censorship Regimes:

# Feedback, please



Thank  
you!

## Next steps

└ Robustness:  Correlations with other existing measures

Sensitivity to: Alternative clustering / list of topics