

Dataset of country mentions in the UNSC debates

This dataset is based mainly on the corpus of the UNSC debates as its source, which consists of three tables:

- `meta_speeches`
- `meta_meetings`
- `raw_docs`

Our dataset comprises 14 tables:

	Table name	# columns	# rows
1	speech country mentions	3	232,729
2	speaker countries	4	72,701
3	speech country-country mentions	6	211,237
4	meeting country-country mentions	6	193,429
5	country-country mentions	10	57,977
6	yearly country mentions	12	4,130
7	events	3	182
8	speech co-mentions	9	69,316
9	speaker country co-mentions	8	61,365
10	yearly co-mentions	7	37,137
11	all-time co-mentions	6	9,901
12	yearly association rules	7	147
13	all-time association rules	6	11
14	non-permanent members	3	250

Table 1: Speech country mentions

File name: **speech_mentions**

Number of Rows: 232,729

Columns:

- **speech_id**: id of the speech (foreign key references **filename** in **meta_speeches**).
- **mentioned_country**: ISO code of a country mentioned in the speech.
- **frequency**: how many times the country was mentioned in the speech.

Table 2: Speaker countries

File name: **speaker_countries**

Number of Rows: 72,701¹

Columns:

- **speech_id**: id of the speech (foreign key references **filename** in **meta_speeches**).
- **meeting_id**: id of the meeting (foreign key references **basename** in **meta_meetings**).
- **year**: the year when the meeting took place.
- **speaker_country**: ISO code of the country of the speaker.

Table 3: Speech country country mentions

File name: **speech_country_country_mentions**

Number of Rows²: 211,237

This is a join of the tables **speech_mentions** and **speaker_countries** based on **speech_id**

Columns:

- **speech_id**: id of the speech.
- **meeting_id**: id of the meeting.
- **year**: the year when the meeting took place.
- **speaker_country**: ISO code of the country of the speaker.
- **mentioned_country**: ISO code of a country mentioned in the speech.
- **frequency**: how many times the country was mentioned in the speech.

¹ This is less than the number of speeches in the original dataset (**meta_speeches**) because some speakers are not representatives of countries, but of organizations (UN, EU, etc.,)

² This is less than the number of mentions in the country mentions table (**speech_mentions**) because some of the speeches are excluded since their speakers are not representatives of countries.

Table 4: Meeting country country mentions

File name: `meeting_country_country_mentions`

Number of Rows: 193,429

This is an aggregation of the `speech_country_country_mentions`

Columns:

- `meeting_id`: id of the meeting.
- `year`: the year when the meeting took place.
- `speaker_country`: ISO code of the country of the speaker.
- `mentioned_country`: ISO code of a country mentioned in the speech.
- `n_speeches`: in how many speeches the country was mentioned by the speaker in the meeting (mostly 1).
- `n_mentions`: how many times in total the country was mentioned by the speaker in the meeting.

Table 5: Country country mentions

File name: `country_country_mentions`

Number of Rows: 57,977

This is an aggregation of the `meeting_country_country_mentions`

Columns:

- `year`.
- `speaker_country`: ISO code of the country of the speaker.
- `mentioned_country`: ISO code of a country mentioned in the speech.
- `n_meetings`: in how many meetings the **mentioned** country was mentioned by the **speaker** country in the year.
- `n_speeches`: in how many speeches the **mentioned** country was mentioned by the **speaker** country in the year.
- `n_mentions`: how many times in total the **mentioned** country was mentioned by the **speaker** country in the year.
- `avg_mention_speech`: average number of mentions per speech.
- `avg_mention_meeting`: average number of mentions per meeting.
- `speaker_member`: one of 'P', 'R' or 'G' indicating whether the **speaker** country is a permanent member, non-permanent member, or non-member of the UNSC, respectively.
- `mentioned_member`: 'P', 'R' or 'G' indicating whether the **mentioned** country is a permanent member, non-permanent member, or non-member of the UNSC, respectively.

Table 6: Yearly country mentions

File name: **yearly_country_mentions**

Number of Rows: 4,130

This is an aggregation of the country_country_mentions (excluding self-mentions)

Columns:

- **year**.
- **mentioned_country**: ISO code of mentioned country.
- **n_meetings**: in how many distinct meetings the country was mentioned in the year.
- **n_speeches**: in how many speeches the mentioned country was mentioned in the year.
- **n_mentions**: how many times the mentioned country was mentioned in the year.
- **n_speaker_countries**: how many times distinct countries mentioned the country.
- **avg_speech_country**: average number of speeches per country
- **avg_speech_meeting**: average number of speeches per meeting
- **avg_mention_country**: average number of mentions per country
- **avg_mention_meeting**: average number of mentions per meeting
- **avg_mention_speech**: average number of mentions per speech
- **mentioned_member**: 'P', 'R' or 'G' indicating whether the mentioned country is a permanent member, non-permanent member, or non-member of the UNSC, respectively.

Table 7: Events

File name: **events**

Number of Rows: 182

Columns:

- **mentioned_country**: ISO code of mentioned country.
- **year**.
- **peak_n_mentions**: how many times the mentioned country was mentioned in the year.

Table 8: Speech co-mentions

File name: **speech_co-mentions**

Number of Rows: 69,316

Columns:

- **speech_id**: id of the speech.
- **meeting_id**: id of the meeting.
- **year**: the year when the meeting took place.
- **speaker_country**: ISO code of the country of the speaker.
- **country1**: ISO code of one mentioned country
- **country2**: ISO code of another mentioned country
- **freq1**: number of mentions of country1
- **freq2**: number of mentions of country2
- **score**: harmonic mean of freq1 and freq2

Table 9: Speaker-country co-mentions

File name: **speaker_country_co-mentions**

Number of Rows: 61,365

This is an aggregation of the speech_co-mentions table (grouped by: year, speaker country, and the two mentioned countries)

Columns:

- **year**: the year when the meeting took place.
- **speaker_country**: ISO code of the country of the speaker.
- **country1**: ISO code of one mentioned country
- **country2**: ISO code of another mentioned country
- **n_speeches**: number of speeches
- **freq1**: number of mentions of country1 (sum over speeches)
- **freq2**: number of mentions of country2 (sum over speeches)
- **score**: harmonic mean of freq1 and freq2 (sum over speeches)

Table 10: Yearly co-mentions

File name: **yearly_co-mentions**

Number of Rows: 37,137

This is an aggregation of the speech_co-mentions table (grouped by: year, and the two mentioned countries)

Columns:

- **year**: the year when the meeting took place.
- **country1**: ISO code of one mentioned country
- **country2**: ISO code of another mentioned country
- **n_speeches**: number of speeches
- **freq1**: number of mentions of country1 (sum over speeches)
- **freq2**: number of mentions of country2 (sum over speeches)
- **score**: harmonic mean of freq1 and freq2 (sum over speeches)

Table 11: Alltime co-mentions

File name: **alltime_co-mentions**

Number of Rows: 9,901

This is an aggregation of the speech_co-mentions table (grouped by the two mentioned countries)

Columns:

- **country1**: ISO code of one mentioned country
- **country2**: ISO code of another mentioned country
- **n_speeches**: number of speeches
- **freq1**: number of mentions of country1 (sum over speeches)
- **freq2**: number of mentions of country2 (sum over speeches)
- **score**: harmonic mean of freq1 and freq2 (sum over speeches)

Table 12: Association rules, Yearly

File name: **association_rules_yearly**

Number of Rows: 147

Columns:

- **year**.
- **LHS**: Left-hand side (antecedent) of the rule (a set of countries)
- **RHS**: Right-hand side (consequent) of the rule (a set of countries)
- **rel_support**: Relative support of the association rule.
- **support**: Support of the association rule.
- **confidence**: Confidence of the association rule
- **lift**: Lift of the association rule.

Table 13: Association rules, Alltime

File name: **association_rules_alltime**

Number of Rows: 11

Columns:

- **LHS**: Left-hand side (antecedent) of the rule (a set of countries)
- **RHS**: Right-hand side (consequent) of the rule (a set of countries)
- **rel_support**: Relative support.
- **support**: in how many distinct meetings the country was mentioned in the year.
- **confidence**: in how many speeches the mentioned country was mentioned in the year.
- **lift**: how many times the mentioned country was mentioned in the year.

Table 14: non-permanent members

File name: **non-permanent-members**

Number of Rows: 250

Columns:

- **year**
- **country**: name of the country (a non-permanent member of the UNSC in that year).
- **code**: ISO code of the country.