9th International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Measures of Importance: a Network of UNGA Speech Mentions

Keywords: Natural Language Processing, Network Analysis, UN Speeches

Extended Abstract

United Nations General Assmebly (UNGA) speeches are an under-explored source of information on government policy preferences over time. Unlike with UNGA votes, each member state is able to address the issues that it deems important, irrespective of an exogenously set or pre-determined agenda. In this paper, I aim to measure the salience that States and other actors have in the international system over time by identifying the number of times they are mentioned by UN member states in any given year between 1970 and 2020, as well as the tone in which they were mentioned.

To this effect, I rely on a combination of natural language processing (NLP) and social network analysis. This results in a ranking of the most mentioned countries in a given year and a list of edges which enumerates the number of times a country i mentioned a country j in a year t. Furthermore, I classify each of these mentions into three categories: neutral, positive, and negative, using Naive Bayes algorithm in order to exclude non-meaningful country mentions. This relational approach can capture a higher degree of granularity on State preferences and stances than comparable existing data sets as mentions data shows higher levels of variation over time (Baturo et al., 2017a) than alliances or diplomatic relations (Duque, 2018).

Though there is a growing literature using NLP and SNA in IR, there is to date little overlap between the two.[1] This stems in part from a lack of new data sets on theoretically relevant cross-border relations. This paper seeks to fill this gap by offering a new data set and outlining reproducible steps on how to combine these two approaches to study a broader range of relevant patterns in the international system.

The data used in this paper comes from the United Nations General Debate Corpus, a compilation of UNGA speeches by world leaders at the annual general debate sessions made available in machine-readable format by Baturo et al., 2017b. It includes all UNGA speeches between 1970 and 2020 consisting of 8,481 separate documents; each containing one speech per-country per-year. The documents are the official English-language translations of the original speeches.

The mentions data from this corpus is extracted by detecting the number of times a relevant actor j was mentioned by a State i in a year t.[2] The data is presented in the form of an *edgelist* of mentions and a *nodelist* of relevant actors. The combined result is a temporal, weighted and directed network of mentions where nodes are countries and other relevant actors in international relations and edges are mentions. Table 1 shows an random sample of the UNGA mentions edgelist.

Subsequently, I classify these mentions into three categories (neutral, positive, and negative) using a combination of substantive knowledge and natural language processing techniques.

[1]For a notable exception see: Peacock et al. (2019). Similarly, Sazak, 2020 uses social network analysis to explore the mechanism of brokerage in the Turco-German alliance during World War I. This approach combines the use of small-n quantitative dyadic data and qualitative archival data (text).

[2]The entire workflow is developed using the R statistical programming language.

Specifically, I draw on the expertise of graduate students and faculty of IR to develop a training data set that captures the nuances of country mentions in international discourse. Next, I train a Naive Bayes algorithm to classify all mentions of countries in the corpus into one of the three categories. The algorithm uses a bag-of-words representation of the text data, with each word or phrase in the text represented as a binary feature. The algorithm then calculates the probability of each mention belonging to one of the three categories based on the frequency of each feature in the training data set.

To date, most measures of country importance in the international system have relied on metrics of their attributes. When relational aspects have been measured (Duque, 2018), these have utilized dyadic data that does not fully capture the dynamic nature of international relations. Figure 1 displays the weighted, directed network of UNGA speech mentions for the year 2000. Unlike with alliance or diplomatic representation dyadic data, mentions are more dynamic. The most mentioned countries change from year to year (Tuvalu has not yet reached the top ten position again since 2000) as the most salient issues on the international agenda shift.

Despite numerous debates on the promises and perils of network analysis for structural approaches in international relations, most empirical studies have relied on a limited pool of available data sets. Studies carried out during the first wave of network analysis in IR (1960s and 70s) examined cross-country ties on trade, organization and alliance membership, and diplomatic exchanges. Though these data sets have since improved in both size and quality, few new options have been added to the menu. This paper address this gap by combining SNA and NLP on a theoretically relevant *corpus* of documents.

Bibliography

Baturo, A., Dasandi, N., & Mikhaylov, S. J. (2017a). Understanding state preferences with text as data: Introducing the un general debate corpus. *Research & Politics*, *4*(2), 1–9. https://doi.org/10.1177/2053168017712821

Baturo, A., Dasandi, N., & Mikhaylov, S. J. (2017b). Understanding state preferences with text as data: Introducing the UN general Debate corpus. *Research and Politics*, *4*(2). https://doi.org/10.1177/2053168017712821

Duque, M. G. (2018). Recognizing international status: A relational approach. *International Studies Quarterly*, *3*(62), 577–592.

Lin Pedersen, T. (2018). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. https://cran.r-project.org/package=ggraph

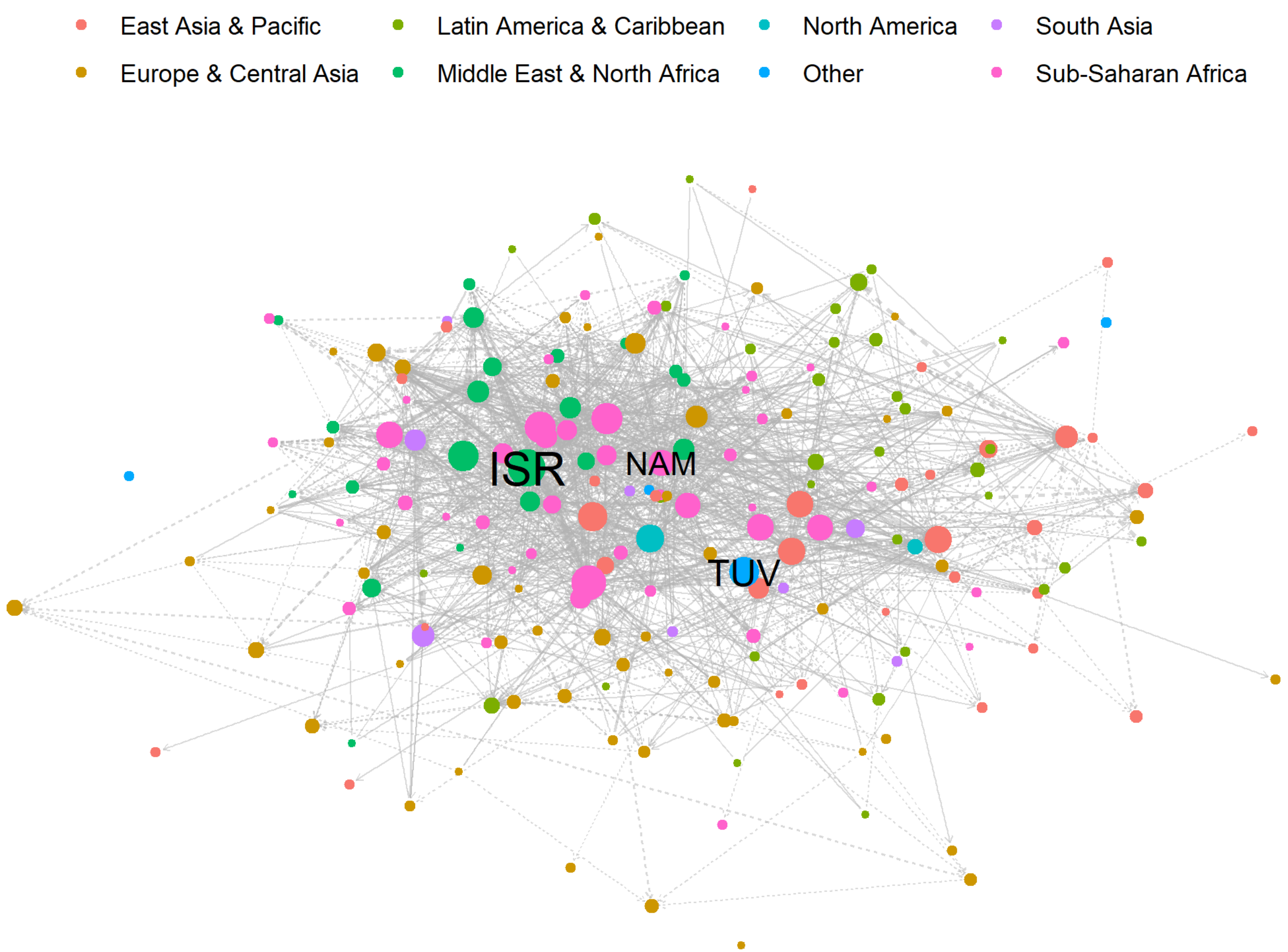
Peacock, C., Milewicz, K., & Snidal, D. (2019). Boilerplate in International Trade Agreements. *International Studies Quarterly*, *63*(4), 923–937. https://doi.org/10.1093/isq/sqz069

Sazak, S. C. (2020). Bad influence: Social networks, elite brokerage, and the construction of alliances. *European Journal of International Relations*, *26*(1_suppl), 64–90. https://doi.org/10.1177/1354066120938839

Table 1: Edgelist Random Sample

year	source	target	weight
1995	TTO	KHM	1
1974	GIN	PRK	4
1994	TUR	ZAF	4
1978	GTM	COG	1
2017	COG	BDI	1
1987	KHM	PAK	2
2005	NGA	MLI	3
1982	GHA	MOZ	1
2007	BOL	USA	4
1990	JOR	IRQ	3
1975	CHN	CHN	8
1988	GNB	ESH	1
1974	GAB	ZAF	1
2008	ECU	NIC	1
2015	AUS	CAF	1

Figure 1: UNGA Mentions Network in 2000



Notes: Node and edge size are determined by weighted in-degree. Dashed edges represent *inter*-regional while solid edges represent *inter*-regional mentions. The network is weighted and directed.

Sources: Baturo et al. (2017a) and author's calculations using *tidygraph* (Lin Pedersen, 2018)