Contentions and Alliances in International Climate Negotiations*

Federica Genovese
University of Essex

Richard McAlexander
Columbia University

Johannes Urpelainen†
Johns Hopkins SAIS

February 1, 2019

Abstract

The United Nations Framework Convention on Climate Change (UNFCCC) plays a crucial role in global climate politics despite its complexity and slow pace of deliberation. In defiance of the great importance of this international organization, there is little systematic understanding of the arguments countries raise at these global meetings. Consequently, there is little cross-temporal understanding of how consistent the dominant topics are, who 'owns' these topics, and if cleavages in the UNFCCC statements reflect contentions and coalitions specific to this organization or embedded in broad international politics. Using a natural language text-analysis approach, this paper explores the evolution and association of countries' plenary speeches at the UNFCCC from 2010 until today. We seek to make two contributions. First, by presenting a granular description of the national statements at recent climate negotiations, we provide a comprehensive picture of the main divisions characterizing these climate meetings in the post-Kyoto protocol era. Second, by focusing our attention on the language associations between developed and developing countries, we explore to what extent this division manifests itself and if other sources of international collaboration can predict similarity across their speeches. Our data indicates where countries stand in these verbal conflicts, and suggests a strong time stickiness of countries' positions. Additionally, we find that speech associations are in part determined by the UNFCCC Annex division but also in part by other factors of international affiliation and geopolitical concerns. The new empirical evidence has implications for the climate negotiation 'gridlock' and the scholarly debate on meaningful climate policy cooperation.

Keywords: climate change politics; United Nations; industrialized countries; developing countries; text analysis.

*We thank Raghav Anand for research assistance.

†Corresponding author. Address: Rome Building, 4th Floor. 1619 Massachusetts Avenue, NW. Washington, DC 20036, USA. Tel: +1-202-779-1352. Email: JohannesU@jhu.edu.

1 Introduction

Observers of international climate politics largely agree that effective solutions to global climate change require some sort of international climate regime (Keohane and Victor, 2016; Underdal, 2017). To date, the international climate regime has mostly been defined by the multilateral negotiations at the United Nations Framework Convention on Climate Change (UNFCCC). While many believe in the faults and failures of this multilateral body of international climate policy, the UNFCCC is still the main climate forum for the international community, even after the Paris accord that established bottom-up voluntary pledges. Therefore, studying country positions at the UNFCCC can offer valuable insights into global climate politics. A systematic review of countries' official claims at the UNFCCC can help scholars understand the drivers of ambition in mitigation (Rietig, 2014) and the role of equity in global burden sharing (Audet, 2013). It can also help explore the importance different governments attach to North-South politics, and – as we investigate in this paper – whether issues across developed and developing countries point to old or new climate 'alliances'.

To assess political associations of governments at UNFCCC, we propose an empirical evaluation of positions at the international climate negotiations using state-of-the-art quantitative text analysis tools. Two observations motivate our investigation. First, a recent wave of UNFCCC studies has focused on the positions of negotiation groups and ideal 'clubs' rather than nation states. Efforts to empirically map the national positions across all UNFCCC members are still scarce, yet nation-states are crucially important for global climate policy, let alone effective climate leadership (Stern, 2015). Second, recent measurements of national positions at the UNFCCC do not fully address associations and similarities across national positions. In other words, we still know relatively little about what nation states talk about when meeting at the global climate talks, how their speeches overlap or diverge, and how that feeds into our understanding of contentions and coalitions in global climate politics.

This knowledge gap is not only important to scholarly work, but also for policy. Without a granular understanding of the national positions at the UNFCCC and their interconnectedness over time and across countries, the international community lacks priors for cross-national policy

connections and potential issues of cooperation or disagreement. Along these lines, the literature has frequently and energetically stressed the important institutional divide between developed (Annex 1) and developing (Non-Annex 1) countries at these negotiations (Kasa, Gullberg, and Heggelund, 2008; Castro, Hörnlein, and Michaelowa, 2014). But without high-quality indicators of positions and their similarity, researchers have little material to validate to what extent the Annex division matters today, under which cases it matters less, and what other international mechanisms may as well capture variation across positions.

We propose new indicators of UNFCCC national positions and a systematic analysis of their associations. We present a text-as-data analysis of all the high-level segment statements from seven most recent UNFCCC negotiations following the Copenhagen talks (2010-2016). Using state-of-the-art methods for frequentist text analysis (Grimmer and Stewart, 2013), we show the general positioning of countries' statements. We then apply an unsupervised machine learning algorithm that generates 'similarity scores' across countries (Le and Mikolov, 2014). The new similarity scores map the networks of UNFCCC speeches, which we investigate with descriptive and econometric analyses. We specifically use the similarity scores to evaluate if and to what extent country positions are mainly clustered by divisions such as the Annex 1/Non-Annex 1 groups or are also captured by other (more or less natural) sources of international cooperation and partnership.

Our new data yield a number of important findings. To begin with, we find that the wording and topics of the countries' statements are stable over time. The 'temporal stickiness' of the national positions is in part a function of the relatively short time-span under analysis, but it also indicates a lack of drastic changes to the agenda issues at the table. It is also particularly relevant if put in contrast with studies that depict observable changes in countries' recent negotiations positions (Andonova and Alexieva, 2012; Hochstetler and Viola, 2012).

Second, we find confirmatory evidence that Annex 1 and Non-Annex 1 statements are systematically different, in line with what other studies on international climate negotiations have found (Castro, Hörnlein, and Michaelowa, 2014). However, the data also points to important nuances. We find that the statements of Annex 1 countries tend to be more similar to each other compared to the Non-Annex 1 – a result that challenges the claim that developing countries have consistently

been more unified than developed ones (Kasa, Gullberg, and Heggelund, 2008). Furthermore, we find some important cases of countries that speak more similarly to peers from the opposite group. Our data allows us to qualitatively describe this similarity across the UNFCCC countries, but we also run regression models to corroborate if associations in speeches are fundamentally determined by the institutional divide between annexes, or if it is also a by-product of other sources of international alignment. Our results confirm the importance of the Annex 1/Non-Annex 1 division, but also point to other forms of international affinity with equal effect size. Most noticeably, we find that the similarity of UNFCCC statements is mostly related to geographic proximity and similar preferences on foreign policy measured by United Nations Assembly General votes.

In light of the empirical findings, the study has several implications for climate policy. While the data indicates that institutional divisions embedded in the UNFCCC Annex are sticky and potentially damaging to a constructive debate of climate policy issues, the distributional conflict also has other subtle forms within the developed versus developing countries' groups. The results imply that, because of the affinity and homogeneity of their positions, Annex 1 countries may be the credible leaders of the post-Copenhagen climate regime structure, as some experts have indicated (Victor, 2011; Hovi et al., 2017). At the same time, countries with mixed interests and interlinked positions exist and are indeed identifiable. The interest of these 'mixed' countries should presumably become a focus of cooperation if they can bridge the abrasive Annex 1/Non-Annex 1 divide. That said, it is up to debate if the 'mixed' countries' interest are truly linked to climate change issues or if they are only signals more broadly related to international affairs, as our findings partly suggest.

2 Distributional Conflicts and Position Affinity at International Climate Negotiations

To set the stage for empirical analysis, we first provide a brief overview of international climate negotiations, summarize the literature on country positions in those negotiations, and specify our hypotheses.

2.1 State Interactions at International Climate Negotiations

Observers of international climate negotiations have long discussed the pitfalls of global climate conventions, pointing to the complications international meetings face when dealing with such a multilayered collective action problem (Keohane and Victor, 2011; Falkner, 2016). Skeptical remarks about the UNFCCC process came already in the early 2000s, before the Kyoto Protocol entered into force. At that point, it had become clear that, due to the conflictual nature of global warming and the pervasive costs unevenly distributed across the UN member states, the Kyoto Protocol negotiations had not yielded effective carbon targets. Commentators pointed to lack of punishment and low costs of compliance as problems that needed to be rectified. To complicate the process from the start, amendments to the Kyoto Protocol would have to pass consensus-based voting, given the unanimity rule of the Convention (Barrett and Stavins, 2003; Victor, 2011).

Against this background, early reviewers claimed that the inefficiency of the UNFCCC negotiations was embedded in the institutional set-up of the convention (Gupta, 2014). In 1995, at the very first Conference of the Parties to the UNFCCC, the so-called Berlin Mandate specified that only industrialized countries would commit to emission reduction targets in what was to be the 1997 Kyoto Protocol. A critical mass of researchers today agree that this division between Annex 1 countries (i.e. the industrialized countries listed in the Annex 1 of the UNFCCC) and the Non-Annex 1 group (i.e. the non-industrialized countries excluded from legally binding greenhouse targets) turned out to be critical for future negotiations, as the developing country bloc insisted on retaining this distinction and opposed any commitments.

This analysis of the faults of the international climate regime has continued on par with the negotiations throughout the years. Public commentators have discussed the pervasive consequences of the Annex 1/Non-Annex 1 divide in the late 2000s (Ott, Sterk, and Wataname, 2008). This discussion heated up at the outset of the 2009 Copenhagen conference (COP15). That year, the international community was expected to produce a successor agreement to the then-outdated Kyoto Protocol. One major expectation was the discussion of new, measured mitigation efforts by the Non-Annex 1 countries. Ultimately, none of the expectations materialized, mainly because the Copenhagen negotiations continued focusing on legal global agreements that were too rigid for

many countries to accept. Along these lines, some scholars claim the Copenhagen talks were made difficult by the corrosive nature of the distributional conflict between developed and developing countries embedded in the UNFCCC (Keohane and Victor, 2016).

Against this background, the output of the 2015 Conference of the Parties (COP21) led some to hope that global climate change governance had now turned away from the complex international legal regime focused mainly on state commitments (Bernstein and Hoffmann, 2018). The 2015 Paris Agreement paved the ground to voluntary pledges to be reviewed internationally, with the aim to increase a virtuous mechanism of global climate accountability and silently remove the Annex 1/Non-Annex 1 dichotomy (Falkner, 2016). But despite the belief that the 2015 agreement may neutralize the complex interactions in the international negotiation process, the accord is still deeply embedded in expectations of 'naming and shaming' that can only exist within the framework of monitoring and transparency of the UNFCCC. In other words, the institutions set in Paris heavily rely on the sustenance of the UN climate negotiations, which provide the main forum for reviewing the national pledges and committing to future ones. Thus, it is up to debate whether the Annex 1/Non-Annex 1 division matters for positions in the post-Copenhagen era. Investigators of international climate policy still need a careful understanding of the conflicts featuring these negotiations and the dimensions these take.

2.2 Country Positions and Affinity

Precisely because of the persistent need to understand the conflictual history of global climate talks, much of the literature on the international climate negotiations has focused on the Annex 1/Non-Annex 1 disputes underlying the UNFCCC debates. Earlier reviews of the international climate conventions point to the importance of contrasting coalitions between developed and developing countries (Roberts and Parks, 2007). For example, Kasa, Gullberg, and Heggelund (2008) have argued that this contrast has been productive for the cohesiveness of non-Annex countries. Other researchers have adapted this view to the interpretation of post-Copenhagen international climate politics, hence focusing on North-South clashes and these coalitions' most salient positions. On that end, several empirical studies have focused on the distinct preferences of the Annex 1 and Non-Annex 1 countries looking at these groups as the central 'blocks' of analysis. Castro, Hörnlein,

and Michaelowa (2014), for example, analyze a dataset of self-coded government statements during the negotiations between 2007 and 2009, and find that belonging to the Annex 1 makes member countries less likely to associate themselves to statements made by Non-Annex 1 countries. In a similar vein, Tobin et al. (2018) make inference on Paris pledge ambition based on manually selected clusters of member states that generally conform with Annex divisions.

We acknowledge the importance of understanding conflict of UNFCCC positions through the Annex 1 and Non-Annex 1 lenses, and we expect this legacy to matter in the years after the Paris Agreement. However, we also anticipate the distributive conflict at the UNFCCC to go beyond this dichotomy and originate from other areas of international politics. We believe expecting a complex interconnectedness of countries' positions at the UNFCCC is warranted for two reasons. First, there is limited evidence that Annex 1/Non-Annex 1 groups are homogenous (Betzold, Castro, and Weiler, 2012) nor that the understanding of their issues is unconflicting and unambiguous (Parker et al., 2012). For example, the findings in Tobin et al. (2018) suggest that Non-Annex 1 countries do not socialize much and have relatively low internal similarity. This has implications for our text-based empirical analysis. Our goal is for our nation-level measurements of UNFCCC positions to give fine-grained evidence of which countries are more or less 'naturally' clustered together. At the same time, we seek to identify which countries are more ambivalent or stretched across topics that 'belong' to different groups (Dubash, 2013; Blaxekjaer and Nielsen, 2015).

Second, political and economic factors outside of the Annex 1/Non-Annex 1 dichotomy make it compelling to look at diverse 'non-climate' sources of international socialization at the UNFCCC (Hovi et al., 2017). Overall, scholars are still undecided on whether conflicts at the climate negotiations extend to other broader concerns in international politics. Some argue that the UNFCCC faces similar fundamental political issues to, for example, World Trade Organization negotiations (Depledge, 2005) and United Nations General Assembly (Baturo, Dasandi, and Mikhaylov, 2017). Others have instead indicated the special institutional characteristics of the climate regime (Colgan, Keohane, and Van de Graaf, 2012). To date, it is unclear how global climate change debates are extendible to conflict in other international issue areas. Against this background, we claim that the literature can benefit from well-specified indicators of positions at the nation-state level that

can be analyzed in network fashions. This is especially compelling if the goal is to map the affinity of positions over time and across countries according to natural linguistic connections, as we do in this paper.

In sum, we claim that, while the UNFCCC Annexes may go a long way to explain similarities and differences across positions over UNFCCC countries, there may be other dimensions of similarities and differences worth exploring from a country-level perspective. We argue that all these dimensions may be measured and validated through national speeches.

2.3 Empirical Expectations

Building on our assessment of the literature above, our premise is that socialization within Annex 1/Non-Annex 1 groups is relevant and should continue to shape country positions on global climate cooperation to this date. However, and differently from the scholars who believe the Annex 1/Non-Annex 1 division to be powerful 'over and above country characteristics' (Castro, Hörnlein, and Michaelowa, 2014: p.114), we believe variation across countries' positions at the UNFCCC is more spread and subtle than this dichotomy can possibly capture. We hold that the deeply held contentions at the UNFCCC are in part institutional and in part due to other sources of international disputes only indirectly related to Annex 1/Non-Annex 1 divisions.

This argument leads us to three empirical expectations. First, because of the deep multidimensional arguments that we expect to capture from UNFCCC speeches, we predict to see a significant amount of temporal consistency – what we call 'time stickiness' – in the UNFCCC texts. We contend that Annex divisions may accurately explain some of this stickiness. At the same time, we expect Annex membership alone to predict only a portion of country speeches and their similarity.

Expectation 1: Country positions are 'sticky' over time, but Annex divisions only explain a fraction of similarity patterns over time.

Second, we anticipate that, while Annex 1 and Non-Annex 1 member countries will produce speeches that are more similar to their 'peers', the two groups may not be equally similar. As previous case studies have indicated (Betzold, Castro, and Weiler, 2012; Roger and Belliethathan, 2016; Watts and Depledge, 2018), Annex 1 countries may be overall more compact due to a clear

set of shared issues. Vice versa, Non-Annex 1 countries may be more heterogeneous and fragmented overall. We expect to observe these patterns in the full sample of countries speaking at the UNFCCC.

Expectation 2: Country positions are starkly divided between Annex 1 and Non-Annex 1 groups, but Annex 1 countries are more unified in their positions than Non-Annex 1 countries.

Finally, we expect that Annex 1/Non-Annex 1 divisions will only go some way to explain the dense map of associations across speeches at the UNFCCC, and that some countries in each of these groups actually have closer positions to members of the other group. So, thirdly, we conjecture that some of the similarity across national positions may be systematically related to other international connections and constructions that are only indirectly feeding in the UNFCCC agenda. We concentrate here on fundamental drivers of dyadic relations frequently acknowledged in international relations research. Empirically, we will focus on shared geography (Weidmann, Kuse, and Gleditsch, 2010), common historical challenges e.g. climate risks (Hochstetler and Viola, 2012; Vihma, 2011), and other sources of international affinity such as similar income levels and foreign policy votes (Bailey, Strezhnev, and Voeten, 2017; Baturo, Dasandi, and Mikhaylov, 2017).

Expectation 3: Country positions are explained by the Annex 1 and Non-Annex 1 division as much as they are explained by other external sources of international affinity.

3 Research Design

To test the validity of our propositions, we engage with an identification of UNFCCC national positions and their ties with text-as-data methods. Applying text-as-data practices to the measurement of national positions at the UNFCCC is increasingly popular (Weisser, 2014). One of major advantages of these approaches is the automatization of the mapping of large numbers of texts. Moreover, these applications allow for an efficient, unbiased and comprehensive overview of representative documents. In the past, several such frequentist approaches have been used to provide indicators of relative positions of countries involved at the UNFCCC climate negotiations (Genovese, 2014; Bagozzi, 2015). However, little systematic text-as-data work has been proposed to

explore socialization and similarities across countries' positions.[1] Furthermore, despite the interest in understanding post-Copenhagen international climate politics, virtually no study has yet used this type of methods to document cross-national positions after the 2009 negotiations. We seek to fill this lacuna and propose a new empirical survey of national positions and their connections using post-Copenhagen UNFCCC texts.

3.1 Data on UNFCCC National Statements

We collected the official high-level segment speeches at the seven annual UNFCCC conferences between 2010 (the earliest conference for which all countries' national statements are reported in readable format on the UNFCCC website) and 2016. We converted all the statements into English-language machine-readable texts, handling them as a corpus following common text-as-data practices (see Appendix for more details). The result is 959 statements for 169 countries. Noticeably, this sample is representative of the whole community of UNFCCC members.

The texts are suitable for different types of text-as-data learning approaches. A common way to handle this data is via extraction and classification strategies. These involve measuring and scaling known words from a vocabulary – the so-called bag-of-word approach. An example of a bag-of-word estimation processes is the Wordfish algorithm (Slapin and Proksch, 2008). This estimator represents each text as a vector of word counts, and then estimates document and word parameters assuming that word frequencies are generated by a Poisson process.[2] Like other approaches, this algorithm compares texts efficiently and is widely used in empirical studies of political positions. However, Wordfish works under certain conditions – including a sufficient (but unknown) number of documents and unique words. It also has difficulties estimating its parameters in sparse, high-dimensional spaces.

While we engage with the family of bag-of-word approaches to build the baseline identification of the data and describe the basic word patterns, we concentrate our investigation on a second type of approach, called word embedding (Mikolov et al., 2013). In contrast to the bag-of-words

[1] Various empirical research relies on texts to code similar positions at the UNFCCC, such as Castro, Hörnlein, and Michaelowa (2014); Tobin et al. (2018). However, these deductively hand code positions from other texts.

[2] Another simple type of bag-as-word approach is the Naive Bayes algorithm. For a past application of this type of method to UNFCCC documents (specifically, the National Communications), see Genovese (2014).

approach, word embeddings encode semantic relationships among words. Put more simply, this algorithm learns the meaning of each word by their context.[3] Because our central motivation is to learn about relative positioning across UNFCCC members, we concentrate on the differences in these embeddings across countries to explore semantic distances in national statements.

A preliminary description of the collected statements provides some initial but relevant observations. On average, the national statements at the High-Level Segment of the UNFCCC are short (844 average words in length) but vary substantively (360 standard deviation). While well beyond the majority of the UNFCCC member states present at least one statement in the course of the seven meeting identified in our study, not all propose a statement every year. In fact, the countries with more frequent speeches (i.e. with more than 5 national statements out of 7 conferences) are 73, of which 39 are developed countries (Annex 1). The texts vary substantively across countries; however, they do not seem to vary much across time. Figure 1 shows the key terms of central issues of redistribution and support to developing countries affected by climate change. If, like some commentators (Falkner, 2016) have indicated, the problems of 'insurance', 'fairness' and 'solidarity' have had increasing salience at the negotiations, these key words should be increasingly more referred to during the national statements at the annual negotiations. However, we find that these terms are relatively steady across time. They are also much less preponderant that more generic concepts such as 'ambition'/'ambitiousness'. We further assess the generalizability of this pattern later in the paper.

[Figure 1 about here.]

3.2 Measuring Similarity of UNFCCC Statements

The central measurement proposed in this study is a score of similarity across all the countries presenting national statements at the High-Level Segment of the UNFCCC negotiations. To generate this score, we treat each country's statement for the entire time period as a single document, because - as we indicated earlier and further validate below - the texts vary little across time. We

[3]For example, a properly trained set of word vectors can produce a representation of words where the distance between 'man' and 'king' is the same as the distance between 'woman' and 'queen'. For a more detailed description of the steps undertaken by the word embedding algorithm, see the Appendix.

only include countries that participate at five or more meetings to limit the estimates to countries that are regularly involved in the negotiations. Our word embedding algorithm operates on each country's collapsed texts, which are represented numerically in one long vector (Le and Mikolov, 2014).[4] This algorithm finds the corresponding vector for a document by maximizing the likelihood of the predicted words in the text. Similarity scores are calculated by measuring the difference between each pair of country-vectors[5]

Figure 2 presents the standardized distribution of these scores computed for each pair of countries. The scores are distributed normally. Overall there are several pairs of countries' cumulative statements that have very little in common with each other. The pattern at the top of Figure 2 warrants against cross-national plagiarism. Put differently, there are several substantive differences across the documents – including some very different texts captured by the tails at the left-end of the distribution. At the same time, the plot also indicates some substantive connections across the texts. More specifically, the histogram at the bottom of Figure 2 indicates that there is a more noticeable similarity among the texts of Annex 1 countries (red bars) than Non-Annex 1 countries (green bars), as shown by the separate histograms. But why do countries' dyads vary with respect to similarity of UNFCCC statements? And which countries tend to be more similar or more different than the rest of their group? We investigate these questions with the dyadic regression analysis described below.

[Figure 2 about here.]

3.3 Covariates of Similarity

We analyze the country-pair similarity scores in a regression framework, in order to explore more precisely what predicts the similarities between countries' speeches at the UNFCCC. Based on the exploration of the raw documents above, Annex 1 and Non-Annex 1 countries seem to be strictly divided, and so should be their construction of UNFCCC statements. Because of the institutional status of Annex 1 states at the UNFCCC and the magnitude of Non-Annex 1 states, it is also

[4] We use the `doc2vec` algorithm to generate these vectors.

[5] We use the cosine similarity to calculate the similarity between different documents. See Appendix for more technical notes.

plausible that there is less heterogeneity among speeches made at the UNFCCC by Annex 1 states than Non-Annex 1 states. To formally explore this claim, we include three binary indicators for whether both states in the dyad are Annex 1 states, whether both are Non-Annex 1 states, and whether both states are in the same Annex.

In addition, we are interested in exploring to what extent our similarity scores are linked to other (more or less 'climate-related') explanations of positions in international politics. In order to explore these propositions, we introduce in the regression equation whether climate-related vulnerability matters for different types of states, assuming that similar vulnerable countries may speak in similar ways. To measure vulnerability, we use the Climate Risk Index (Kreft et al., 2013), which is a time-varying measure of each country's vulnerability in terms of deaths and income losses to weather-related loss events (e.g. storms, floods, and heat waves).[6] Additionally, we include four measures of sources of international political preferences that we expect to be relevant in international climate politics: the similarity of votes in the essential UN decision-making chamber, the General Assembly; geographic distance among the focal countries; shared security treaties; as well as whether countries have similar levels of income. To measure the similarity of votes in the UN assembly, we use the absolute value difference in the ideal points estimated from voting patterns at the United Nations General Assembly from Bailey, Strezhnev, and Voeten (2017). To capture geographic distance, we include the log of distance between capital cities (Weidmann, Kuse, and Gleditsch, 2010). To measure security alliances (including defense commitments and military treaties), we use the Correlates of War formal alliance dataset (Gibler and Sarkees, 2004). This variable takes on a 1 if the two countries had any formal treaty during the entire time period. We finally use the difference in GDP per capita as measured by the World Bank Indicators.

Note that we collapse the time dimension and take the average of all non-missing values to produce a cross-section where the unit of analysis is the dyad. Both the similarity measures and the difference scores are standardized to aid interpretation. For our correlational analysis, we use OLS to estimate normal dyadic linear models.

[6] To convert this to dyadic data, we calculate the absolute value of the difference between each country's CRI score.

4 Results

We present our findings following the order of our hypotheses. We first present our results on the 'time stickiness' of the UNFCCC speeches, and then move to discuss the sources of interconnectedness and similarity of these texts.

4.1 Stickiness of Statements over Time

Evaluating how country positions vary over time is important because the extent to which they vary would indicate how volatile issues and positions are (Jasny et al., 2018; Kammerer and Namhata, 2018). Additionally, time consistency of country positions may or may not be a function of country's converge to their groupings (e.g. the Annex they belong to). To explore these questions, we fist assess whether the UNFCCC country speeches are at all sticky across time.

To systematically estimate the time consistency of the UNFCCC speeches, we calculated the predictiveness of the statements across the seven years. Using the original document vector model, we trained the model on texts from all but one of the meetings. Then, we use the model to predict the country that the text belongs to. The misclassification error for each meeting is remarkably stable: if we remove an entire meeting from the training set we can predict with about 0.6 accuracy which country is speaking in the held-out meeting (the misclassification error ranges from 39% to 47%).[7] This is evidence that 'time stickiness' is significant and substantial and does not pertain only the key words highlighted earlier.

Additionally, we evaluate the extent to which time consistency is dependent on countries' group affiliation and - specifically - their division between Annex 1 and Non-Annex 1 countries. In order to do this, we contrast two forecasting models based on the similarity scores of the country dyads described above. One model holds out a meeting and employs the similarity scores for each dyad to predict the scores for a held-out meeting. The alternative model only employs Annex 1/Non-Annex 1 membership as the main predictor of the similarity scores.

We constructed a list of all dyads from countries that appear in five or more of the meetings. We then created similarity scores from document vectors using the speeches from each meeting.

[7] This is actually quite good given that we have almost 200 countries in our dataset. Previous studies with much simpler data structures and more observations (e.g., Peterson and Spirling, 2018) ranged at about a 0.8 accuracy.

This created a similarity scores for each dyad-meeting pair. For each meeting, we took the average of all the other similarity scores for the dyad-meeting observations.[8] For each dyad-meeting, we then created a variable that takes the average of the remaining dyad-meeting similarity scores and use that as a predictor.[9] We used that average as the predictor in a classification model – namely, a random forest model – where the outcome takes a value of 1 if each similarity score is above the mean of all the similarity scores for each focal meeting, and 0 otherwise. Following this binary classification, we can then plot a receiver operating characteristic (ROC) curve for each random forest model (i.e. for each meeting).[10] In order to infer the extent to which time dependencies is caused by Annex 1/Non-Annex 1 socialization, we also calculated a 'naive' model where only the binary shared Annex variables are used.

The ROC curves for these models are presented in Figure 3. The plots show the ROC curves for each meeting, where each ROC curve is plotted in a different color to show the similarity/differences across meetings. The predictors are binary variables that indicate if the dyad is in the same annex and the average of the remaining meeting similarity scores. The plot on the left shows the ROC curves for each meeting using the held-out meeting?s similarity scores. The plot on the right shows the ROC curves for the naive model.

The results highlight two important points. One is that the ROC curves for each meeting using the held-out similarity scores are predictors are extremely similar. The other is that the held-out similarity scores greatly increase the predictive accuracy of the models: while the naive model on the right predicts better than an uninformed 'coin-flipping' forecast, it is hardly as accurate as the time dependency model on the left. Altogether, our findings are that the speeches are stable over time, and that this stability is only partially related to Annex 1/Non-Annex 1 country divisions.

[Figure 3 about here.]

[8] For example, for the US-UK dyad, we have the following similarity scores: US-UK-Doha, US-UK-Durban, US-UK-Paris, US-UK-Cancun, US-UK-Marrakech and US-UK-Lima.

[9] So, for US-UK-Doha, we create an average of US-UK-Durban, US-UK-Paris, US-UK-Cancun, US-UK-Marrakech and US-UK-Lima

[10] A ROC curve illustrates the diagnostic ability of a binary classifier system by plotting the true positive rate against the false positive rate at various threshold settings. It is a widely used tool in forecasting and classification machine learning.

4.2 Similarity Patterns in Country Positions

Despite the limited movement of statements across time, there are substantive cross-national differences of wordings across the statements. Here we evaluate if, as a significant part of the literature points out, these distinctions are mainly embedded in the Annex 1/Non-Annex 1 division. For descriptive purposes, we first resort to a simple estimation of document-parameters using the Wordfish algorithm. Wordfish captures the position of documents in a unidimensional space after having specified the direction for the positions. We identify an Annex 1 country to be to the right of a Non-Annex 1 country, but the results are robust to different document anchoring.

Figure 4 shows the document-level estimates for the statements collapsed at the country level. As delineated in the literature, we find a significant divide in the distribution of the country estimates: most of the Annex 1 countries stand on the right-side of the plots, while the majority of the Non-Annex 1 countries is on the left-side. This evidence supports the argument of the institutional conflict implanted in the UNFCCC. It is also important for the purpose of identifying networks and overlapping issues across countries. The country estimates close to the zero line constitute cases of statements that cross the Annex 1/Non-Annex 1 divide, thus potentially 'bridging' countries at the negotiations. These include emerging economies such as India and Indonesia – i.e. countries with increasingly similar economic structures as developed countries.

[Figure 4 about here.]

We next analyze our outcome variable in the dyadic matrix fashion, to explore the determinants of interdependence of the UNFCCC countries' texts collapsed across time. Figure 5 illustrates our similarity scores in a heatmap format, along with the country name and lines indicating the division between Annex 1 and Non-Annex 1 countries. The individual values correspond to the similarity between two countries and are represented as changing colors.

[Figure 5 about here.]

The heatmap has a few important implications for the understanding of similar interests and different positions at the UNFCCC. First, the data indicates that Annex 1 and Non-Annex 1

countries are divided into two clusters. This is evinced by the red dots grouping on the upper right triangle (where scores for dyads where both countries are in the Annex 1 are presented) and the lower left triangle (with scores for dyads where both countries are not in the Annex 1). At the same time, the plot shows that the speeches of the Annex 1 countries are overall more similar than the Non-Annex 1 countries, suggesting a more cohesive group of similarly minded countries. Furthermore, some 'mixed pairs' (mix of Annex 1/Non-Annex 1 countries) also seem to converge on similar speeches, as shown by some red spots in the otherwise whiter square in the graph. On the Annex 1 side, these relatively similar mixed pairs involve smaller European countries such as Czech Republic, Latvia and Cyprus. On the Non-Annex 1 side, the countries whose statements 'cross border' are richer ones such as Israel and Saudi Arabia but also emerging economies identified in the Wordfish estimations, such India and the Phillippines.[11] Following our initial argument, these countries may be better connected to others outside of their group by means of similar structural or geographic characteristics.

To further understand the connections unveiled by the similarity scores, we move to the regression analysis of these measures on the previously discussed covariates of interdependence. The results of our linear estimations are presented in Table 1. We present three models. One model follows the baseline theory for discord and cleavages at the UNFCCC and that only includes dummies for whether the two countries of each dyad are both Annex 1 or Non-Annex 1 countries. One model ignores the institutional Annex 1/Non-Annex 1 distinction and explores the explanatory power of classical determinants of international politics, specifically the difference in vulnerability between the paired countries, the UN ideal point variation, the geographic distance, the difference in GDP per capita and whether both countries are signatories to the same international security treaties. The last model is the full specification that includes all the mentioned covariates.

Our first model evinces the importance of the connections embedded in the Annex 1/Non-Annex 1 institutional groups. We find that two countries belonging to the same group are significantly more likely to have more similar speeches - and thus, presumably, to cover more similar topics. Our second model however shows that the other covariates matter just as well. As the geographic

[11] For example, according to our estimates India's statements are as similar to statements by Singapore and Namibia as to the speeches by Denmark and Switzerland.

distance or UN ideal point difference increases, UNFCCC speeches become less similar.[12] Regarding our last model, we find that all such results hold but that coefficient of the Non-Annex 1 countries loses precision. In other words, Annex 1 states seem to be on average more similar to each other that Non-Annex 1 states.

[Table 1 about here.]

Our main findings are robust to a number of tests and model specifications reported in the Appendix. In other graphical analyses, we show that the two groups of countries share similar words, although Annex 1 countries share more frequently their more similar words (e.g. 'cooperation' and 'respect'). We also run a basic Latent Dirichlet Allocation algorithm to calculate the topics the emerge more prominently from the Annex 1 and Non-Annex 1 texts. The results indicate that Non-Annex 1 countries talk more frequently about disasters (as evinced by Topic#3 in the Appendix, where words such as 'typhoon', 'philippines', 'adaptation' and 'islands' are prominent). Vice versa, Annex 1 countries are more focused on 'energy' and 'carbon'. At the same time, there are lots of overlapping topics across Annex 1 and Non-Annex 1 countries that seem centered on international law and institutional governance (captured by 'president', 'agreement', and 'action' across different topics). Again, this analysis suggests more cross-cutting language issues than the Annex 1/Non-Annex 1 division may possibly explain.

Finally, and in light of the topic model analysis, we explore alternative heterogenous effects of Annex 1 and Non-Annex 1 divisions. In the Appendix, we implement interaction models of the Annex 1/Non-Annex 1 dyads and the CRI difference variable. The results indicate that the interaction is insignificant for the Annex 1 dyads. However, it is significant for Non-Annex 1 countries. This suggests that, while Annex 1 countries may speak more similarly to each other on average, Non-Annex 1 countries with high ecological vulnerability to climate change speak more similarly. Presumably, this is the case because they similarly tend to speak about resilience and disasters (as the topic models seems to suggest).

Altogether, we think this evidence points to, on the one hand, the fundamental role of Annex

[12] Countries with any treaty are more likely to have similar speeches, and states that are similarly vulnerable to climate change also have more similar speeches. However, these results are small and not statistically significant.

1/Non-Annex 1 divides for UNFCCC debates, and on the other hand, the influence of other sources of interdependence beyond the climate per se. We find support for the perniciousness of the North-South politics at the UNFCCC. At the same time, we identify countries and topics that may be pivotal to bridge or at least mediate some of these divisions.

5 Conclusion

Understanding conflict and cooperation at international climate conferences is of fundamental importance for the effective tackling of global climate change. While the stop-and-go patterns of the multilateral negotiations at the UNFCCC have triggered much criticism, the international talks continue to matter for the future of climate policy even under the Paris pledge-based accord. Consequently, the international community needs careful, up-to-date measures of national positions at these negotiations. Understanding how these positions are related and to what extent they overlap or diverge is also critical. While the major institutional division of Annex 1/Non-Annex 1 countries is virtually common wisdom, a validation of such cleavage or, alternatively, of other possible associations across the UNFCCC members requires sustained fine-grained analyses.

Tackling this research agenda, in this paper we presented new indicators of national positions at the UNFCCC that allow us to investigate these UNFCCC patterns in years following the 2009 Copenhagen meeting. The data confirms the strong power of the Annex 1/Non-Annex 1 categories to predict differences in statements. At the same time, our empirics also presents subtle features of national positions that are somewhat mixed to this institutional divisions. We also show the varying degree of associations across national positions, and how similarity of positions is explained by institutional UNFCCC categories as well as other international political determinants

What does our research offer to the future study of international climate policy? Our results clearly suggest that past cross-national conflicts will continue mattering. Against that light, the distinction of Annex 1 and Non-Annex 1 could in principle create incentives for both groups of parties for free-riding. However, we also observe variation within and across these two groups that could lead the future of international climate cooperation in different directions. Future creative forms of the international climate regime will have to confront - and possibly exploit - this variation for the sake of reaching effective policies for decarbonization (Eckersley, 2012). For example, it may

be worth focusing on specific countries whose positions are mixed and share similarity across the spectrum of countries. But to leverage the mixed positions of these countries for the common good, it will be critical to look into their domestic preferences, their internal economic structures, and assess how these could be catalyzed for global climate policy diffusion (Kammerer and Namhata, 2018).

References

Andonova, Liliana B, and Assia Alexieva. 2012. "Continuity and change in Russia's climate negotiations position and strategy." *Climate Policy* 12 (5): 614–629.

Aronow, Peter M, Cyrus Samii, and Valentina A Assenova. 2015. "Cluster - robust variance estimation for dyadic data." *Political Analysis* 23 (4): 564–577.

Audet, R. 2013. "Climate justice and bargaining coalitions: a discourse analysis." *International Environmental Agreements: Politics, Law and Economics* 13 (3): 369–386.

Bagozzi, Benjamin E. 2015. "The Multifaceted Nature of Global Climate Change Negotiations." *The Review of International Organizations* 10 (4): 439–464.

Bailey, Michael A, Anton Strezhnev, and Erik Voeten. 2017. "Estimating dynamic state preferences from United Nations voting data." *Journal of Conflict Resolution* 61 (2): 430–456.

Barrett, Scott, and Robert Stavins. 2003. "Increasing Participation and Compliance in International Climate Change Agreements." *International Environmental Agreements: Politics, Law and Economics* 3: 349–376.

Baturo, Alexander, Niheer Dasandi, and Slava J. Mikhaylov. 2017. "Understanding state preferences with text as data: Introducing the UN General Debate corpus." *Research and Politics* 2 (1-9).

Bernstein, Steven, and Matthew Hoffmann. 2018. "The politics of decarbonization and the catalytic impact of subnational climate experiments." *Policy Sciences* 51 (2): 189–211.

Betzold, Carola, Paula Castro, and Florian Weiler. 2012. "AOSIS in the UNFCCC negotiations: from unity to fragmentation?" *Climate Policy* 12 (5): 591–613.

Blaxekjaer, L./O., and T.D. Nielsen. 2015. "Mapping the narrative positions of new political groups under the UNFCCC." *Climate Policy* 15 (6): 751–766.

Castro, Paula, Lena Hörnlein, and Katharina Michaelowa. 2014. "Constructed peer groups and path dependence in international organizations: The case of the international climate change negotiations." *Global Environmental Change* 25: 109–120.

Colgan, J., R. Keohane, and T. Van de Graaf. 2012. "Punctuated equilibrium in the energy regime complex." *The Review of International Organizations* 7 (2): 117–143.

Depledge, Joanna. 2005. *The organization of global negotiations: Constructing the climate change regime.* Earthscan.

Dubash, Navroz K. 2013. "The politics of climate change in India: Narratives of equity and cobenefits." *Wiley Interdisciplinary Reviews: Climate Change* 4: 191–201.

Eckersley, Robyn. 2012. "Moving Forward in the Climate Negotiations: Multilateralism or Minilateralism?" *Global Environmental Politics* 12 (2): 24–42.

Falkner, Robert. 2016. "The Paris Agreement and the New Logic of International Climate Politics." *International Affairs* 92 (5): 1107–1125.

Genovese, Federica. 2014. "States' interests at international climate negotiations: New measures of bargaining positions." *Environmental Politics* 23 (4): 610–631.

Gibler, Douglas M, and Meredith Reid Sarkees. 2004. "Measuring alliances: The correlates of war formal interstate alliance dataset, 1816–2000." *Journal of Peace Research* 41 (2): 211–222.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297.

Gupta, Joyeeta. 2014. *The History of Global Climate Governance.* Cambridge University Press (CUP).

Hochstetler, Kathryn, and Eduardo Viola. 2012. "Brazil and the politics of climate change: Beyond the global commons." *Environmental Politics* 21 (5): 753–771.

Hovi, Jon, Detlef F. Sprinz, Hakon Saelen, and Arild Underdal. 2017. "The Club Approach: A Gateway to Effective Climate Co-operation?" *British Journal of Political Science* .

Jasny, L, Dewey AM, Robertson AG, Yagatich W, Dubin AH, and Waggle JM. 2018. "Shifting echo chambers in US climate policy networks." *PLoS ONE* 13 (9).

Kammerer, Marlene, and Chandreyee Namhata. 2018. "What drives the adoption of climate change mitigation policy? A dynamic network approach to policy diffusion." *Policy Sciences* Forthcoming.

Kasa, Sjur, Anne T Gullberg, and Gorild Heggelund. 2008. "The Group of 77 in the international climate negotiations: Recent developments and future directions." *International Environmental Agreements: Politics, Law and Economics* 8: 113–127.

Keohane, Robert O., and David G. Victor. 2011. "The Regime Complex for Climate Change." *Perspectives on Politics* 9 (1): 7–23.

Keohane, Robert O., and David G. Victor. 2016. "Cooperation and discord in global climate policy." *Nature Climate Change* 2016 (6): 570–575.

Kreft, Sönke, David Eckstein, Lisa Junghans, Candice Kerestan, and Ursula Hagen. 2013. "Global climate risk index 2014." *Who suffers most from extreme weather events* pp. 1–31.

Le, Quoc, and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML'14 JMLR.org pp. II–1188–II–1196.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality." In *Advances in Neural Information Processing Systems 26*, ed. C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. Curran Associates, Inc. pp. 3111–3119.

Ott, Hermann E., Wolfgang Sterk, and Rie Wataname. 2008. "The Bali roadmap: new horizons for global climate policy." *Climate Policy* 8: 91–95.

Parker, C. F., C. Karlsson, M. Hjerpe, and B.-O. Linnér. 2012. "Fragmented climate change leadership: Making sense of the ambiguous outcome of COP-15." *Environmental Politics,* 21 (2): 268–286.

Peterson, Andrew, and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26: 120–128.

Rietig, Katharina. 2014. "Reinforcement of multilevel governance dynamics: Creating momentum for increasing ambitions in international climate negotiations." *International Environmental Agreements: Politics, Law and Economics* 14 (4): 371–389.

Roberts, J. Timmons, and Bradley C. Parks. 2007. *A climate of injustice: global inequality, North–South politics, and climate policy.* Cambridge, MA: MIT Press.

Roger, Charles, and Satishkumar Belliethathan. 2016. "Africa in the global climate change negotiations." *International Environmental Agreements: Politics, Law and Economics* 16 (1): 91–108.

Slapin, Jonathan B, and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–722.

Stern, Nicholas. 2015. *Why Are We Waiting? The Logic, Urgency, and Promise of Tackling Climate Change.* MIT Press.

Tobin, P., N. M. Schmidt, J. Tosun, and C. Burns. 2018. "Mapping states' Paris climate pledges: Analysing targets and groups at COP21." *Global Environmental Change* 48 (11-21).

Underdal, A. 2017. "Climate Change and International Relations (After Kyoto)." *Annual Review of Political Science* 20: 169–188.

Victor, David G. 2011. *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet.* Cambridge University Press (CUP).

Vihma, Antto. 2011. "India and the Global Climate Governance: Between Principles and Pragmatism." *The Journal of Environment and Development* 20 (1): 69–94.

Watts, Joshua, and Joanna Depledge. 2018. "Latin America in the climate change negotiations: Exploring the AILAC and ALBA coalitions." *Wiley Interdisciplinary Reviews: Climate Change* (13): 533.

Weidmann, Nils B, Doreen Kuse, and Kristian Skrede Gleditsch. 2010. "The geography of the international system: The CShapes dataset." *International Interactions* 36 (1): 86–106.

Weisser, Florian. 2014. "Practices, politics, performativities: Documents in the international negotiations on climate change." *Political Geography* 40: 46–55.

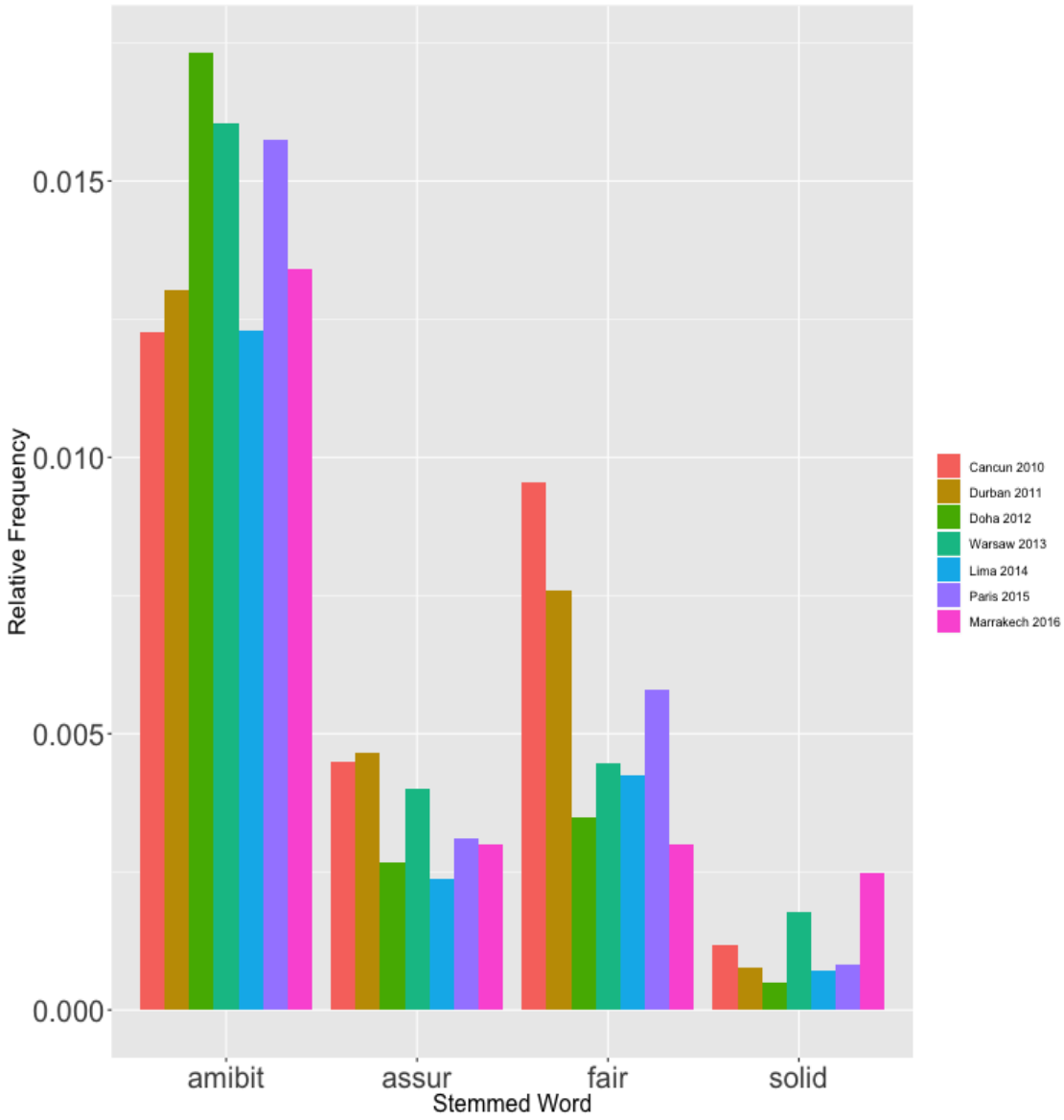

Figure 1: This heatmap shows the standardized frequency of selected word frequencies.

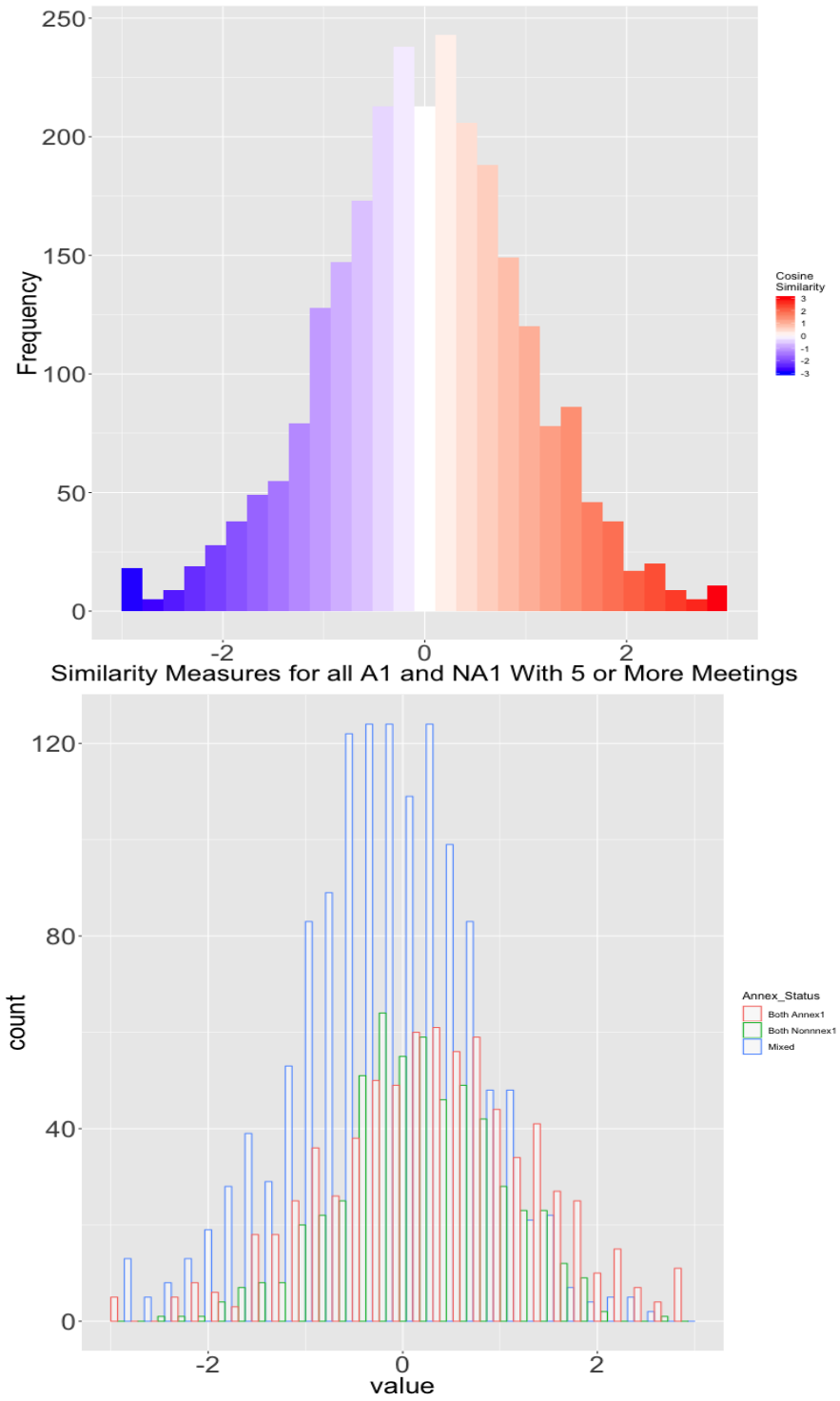

Figure 2: The histograms show the distribution of the similarity scores computed from document vectors for each country's statement. The histogram at the top shows the similarity scores for all the countries that presented at least 5 of the 7 UNFCCC meetings. The histogram at the bottom shows the distribution of the similarity scores divided by Annex.

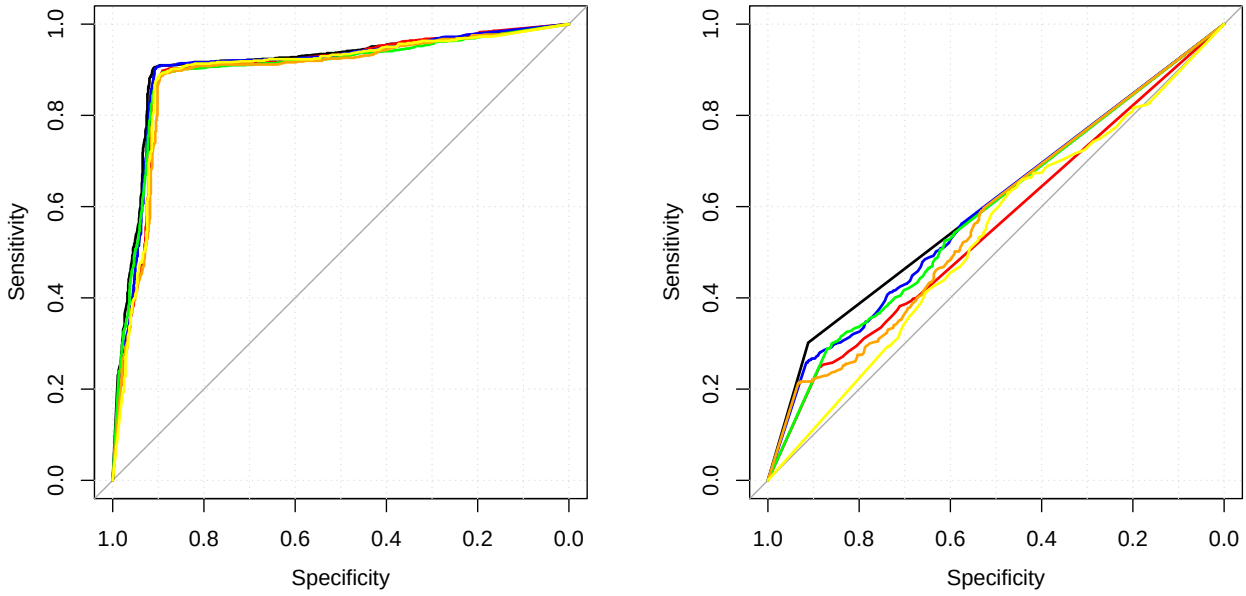

Figure 3: The plot on the left shows the ROC curves for each meeting using the held-out meeting's similarity scores. The plot on the right shows the ROC curves for the naive model, where the only predictors are variables indicating if both countries in the dyad are in the same annex. Since the individual ROC curves on the left are similarly good at predicting each held-out meeting, this shows the stickiness of the speeches and the similarity scores over time. Also, the ROC on the left plot cover more area-under-the curve than the ROC on the right plot, and this indicates that the model on the left is more accurate at predicting out-of-sample statements. Note: sensitivity (also called the true positive rate) and specificity (also called the true negative rate) are statistical measures of the performance of a binary classification test.

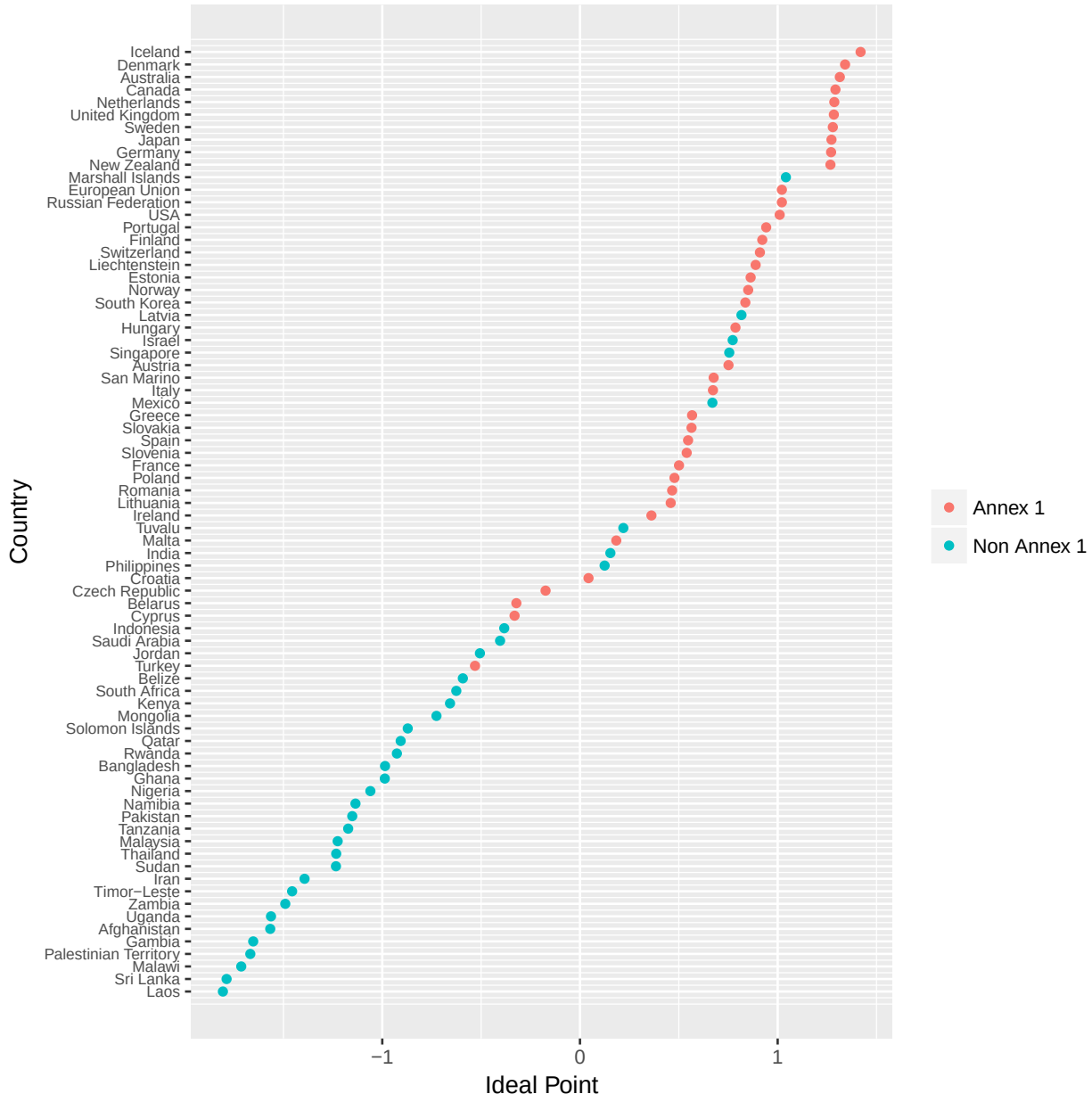

Figure 4: This dot plot shows the distribution of the document-level Wordfish estimates for the sample that contains all the countries that produced speeches at at least 5 of the seven UNFCCC meetings.

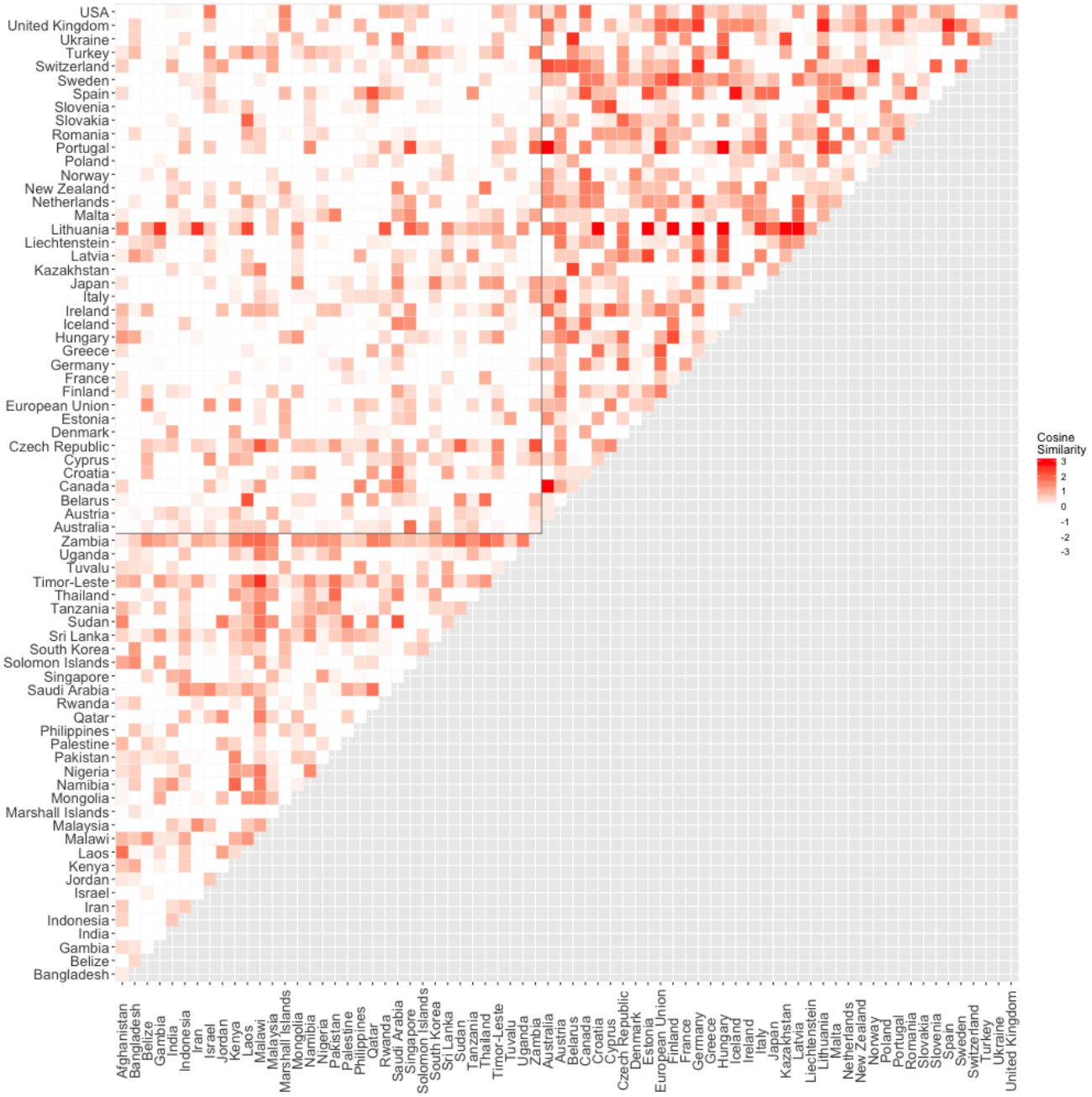

Figure 5: This heatmap shows the similarity scores. Pairs of more similar statements are indicated by red, while less similar statements are indicated by white. The L-shaped lines demarcates the border between Annex 1 and Non-Annex 1 dyads. Dyads between Annex 1 and Non-Annex 1 countries are in the upper left square, and show lower similarity scores. The upper right triangle shows scores for dyads where both countries are in the Annex 1, and the lower left triangle shows scores for dyads where both countries are not in the Annex 1.

	Dependent variable:		
	Similarity Scores		
	(1)	(2)	(3)
Both non-Annex 1	0.130^{**}		0.073
	(0.051)		(0.056)
Both Annex 1	0.623^{***}		0.325^{***}
	(0.090)		(0.108)
CRI Difference		−0.021	−0.021
		(0.020)	(0.019)
UN Ideal Point Difference		-0.080^{***}	-0.053^{**}
		(0.020)	(0.023)
Geographic Distance		-0.263^{***}	-0.240^{***}
		(0.033)	(0.036)
GDP Per Capita Difference		0.001	0.006
		(0.018)	(0.018)
Treaty Indicator		0.121	0.104
		(0.121)	(0.116)
Observations	9,180	9,180	9,180
R^2	0.024	0.058	0.063
Adjusted R^2	0.024	0.058	0.063
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 1: This table shows the result of OLS models where the dependent variable is the similarity score for each dyad. All continuous variables, including the dependent variable, are standardized. Standard errors are clustered at the dyad using the method of (Aronow, Samii, and Assenova, 2015).

Contentions and Alliances at International Climate Negotiations
Supporting Information

Federica Genovese
University of Essex

Richard McAlexander
Columbia University

Johannes Urpelainen*
Johns Hopkins SAIS

February 1, 2019

*Corresponding author. Address: Rome Building, 4th Floor. 1619 Massachusetts Avenue, NW. Washington, DC 20036, USA. Tel: +1-202-779-1352. Email: JohannesU@jhu.edu.

Data Management and Calculations

All of the statements are in the form of a PDF. To convert them into machine-readable text, we OCRed each file. Many countries supplied the English translations of their statement alongside the statement in their native language. In those cases, we only extracted the supplied English translation. If no English translation was supplied, we used the Google Translate API to automatically translate the text into English.[1] Many of the texts from the earlier conferences were of poor quality. To ensure that the output from the OCR program matched the PDF file, we manually examined each text and corrected any discrepancies. This resulted in a set of 959 machine-readable English-language statements. To avoid errors induced by overprocessing the data, we only remove punctutation and convert all characters to lower-case (Denny and Spirling, 2018).

We chose to use word embeddings to examine these texts (Mikolov et al., 2013). Word embeddings are a powerful tool that is widely used in the machine learning community but has remained underutilized in applied political science research. Word embeddings learn the meaning of each word by their context. The meaning of each word is represented geometrically by a N-dimensional vector estimated from the training of a neural net. In this neural net, the predicted value is either the missing word within a specific range of words, or the context words for each word. The result is a n-dimensional vector associated with each word. Since similar words appear in similar contexts, word embeddings can learn the meanings of words. Moreover, since the vector has a geographical interpretation, word embeddings can engage in analytical reasoning. For example, a properly trained set of word vectors can produce a representation of words where the distance between 'man' and 'king' is the same as the distance between 'woman' and 'queen'.

This approach has a number of important differences with topic models, which are currently the most widely used tools used to analyze large amounts of text in political science. Probabilistic topic models (such as LDA and the STM) assume that each document is a probability distribution over topics and each topic is a distribution over words (Blei, Ng, and Jordan, 2003; Roberts et al., 2014). These models are useful for finding clusters of words, but tell us little about their meaning. In contrast, word embeddings use the context words to predict the meaning of each individual word.

[1] The source languages include: French, Russian, Arabic, and Spanish.

Topic models can tell us which words are being used in each press statement. Word embeddings can tell us if the context of the same word–say, vulnerability or fairness–differs across each country's press statement. Crucially, most topic modelling approaches use the 'bag-of-words' approach, which analyzes the raw frequency of words in each document. Our approach analyzes words in their context, and preserves the ordering of words.

To analyze these the press statements, we use a use the doc2vec algorithim of Le and Mikolov (2014). While word2vec creates a numeric representation of a word, doc2vec creates a numeric representation of each document. This numeric representation–the document vector–is trained to predict each word within the document. The key idea of doc2vec is that the document as a whole is useful for predicting the surrounding words of any other word within the document. Similar documents will have similar predictions for words that appear in the same context. We use the `gensim` package in Python to run doc2vec.[2]. For present purposes, we treat each country's press statement for the entire time period as a single document, and only include countries that participate at five or more meetings. We then run the doc2vec algorithim to produce 169 document-vectors of length 200. The core idea is that countries that issue press statements which have similar content will also have similar document vectors.

The main analysis of these vectors consists of calculating similarity scores across each document vector. To do this, we rely on the often-used cosine similarity, defined as follows:

$$\cos(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{AB}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i}{\sqrt{\sum_{i=1}^{n} (\mathbf{A}_i)^2} \sqrt{\sum_{i=1}^{n} (\mathbf{B}_i)^2}} \tag{1}$$

Where $\mathbf{A}$ and $\mathbf{B}$ are two vectors. Calculating the cosine similarity for each pair of document vectors produces a 169 by 169 symmetrical matrix.

Robustness

[2]We use a vector size of 200 and standard tuning parameters. These include setting the number of iterations at 50 and excluding words that appear five times or less.

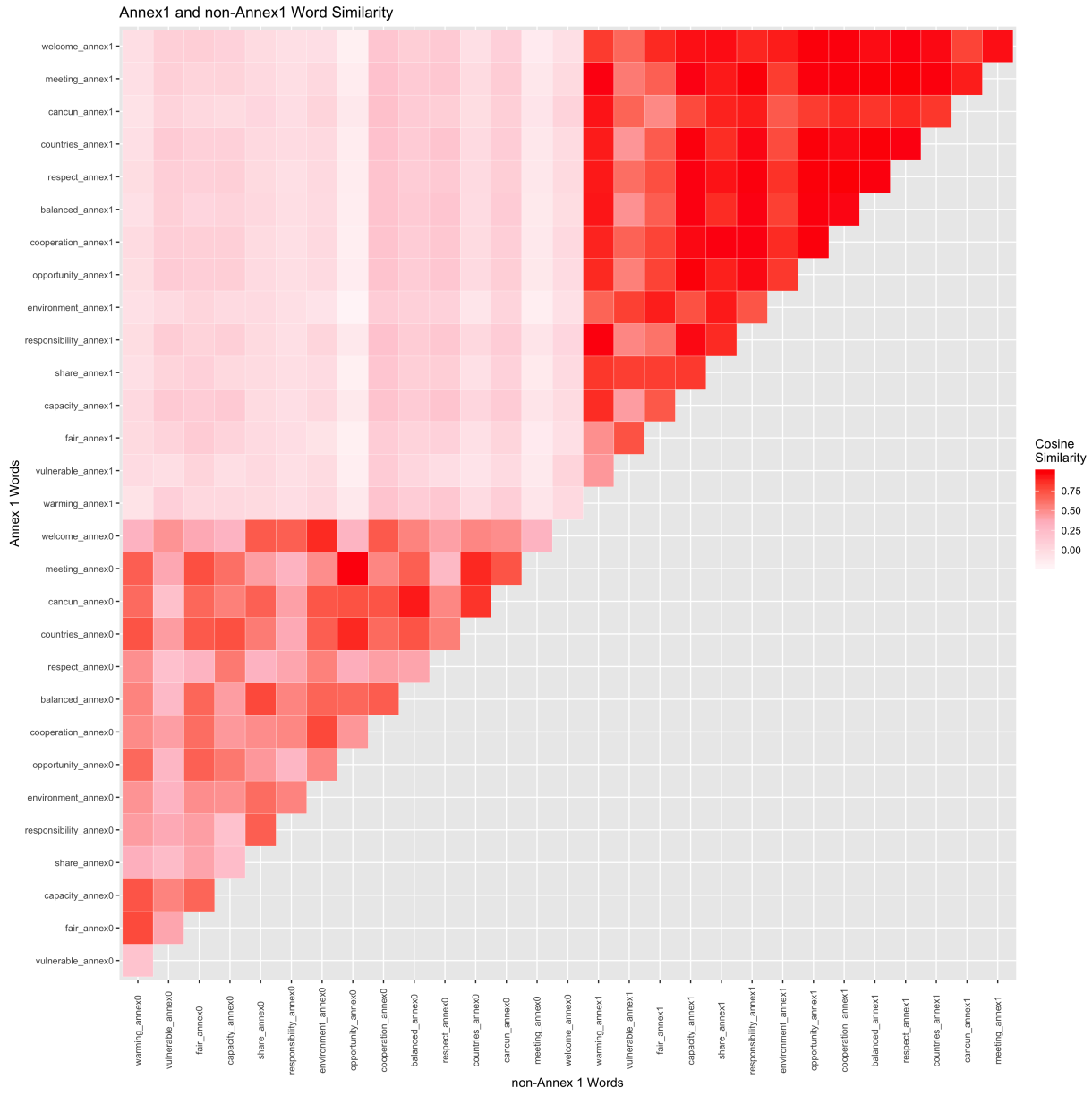

Figure A1: This plot shows the similarity scores for the same word using vectors from Annex 1 and non-Annex 1 documents. The upper right hand triangle shows the similarity scores for words that are in the Annex 1 documents, the lower left hand triangle shows the scores for words in non-Annex 1 documents, and the upper left square shows the similarity for words across the two groups.

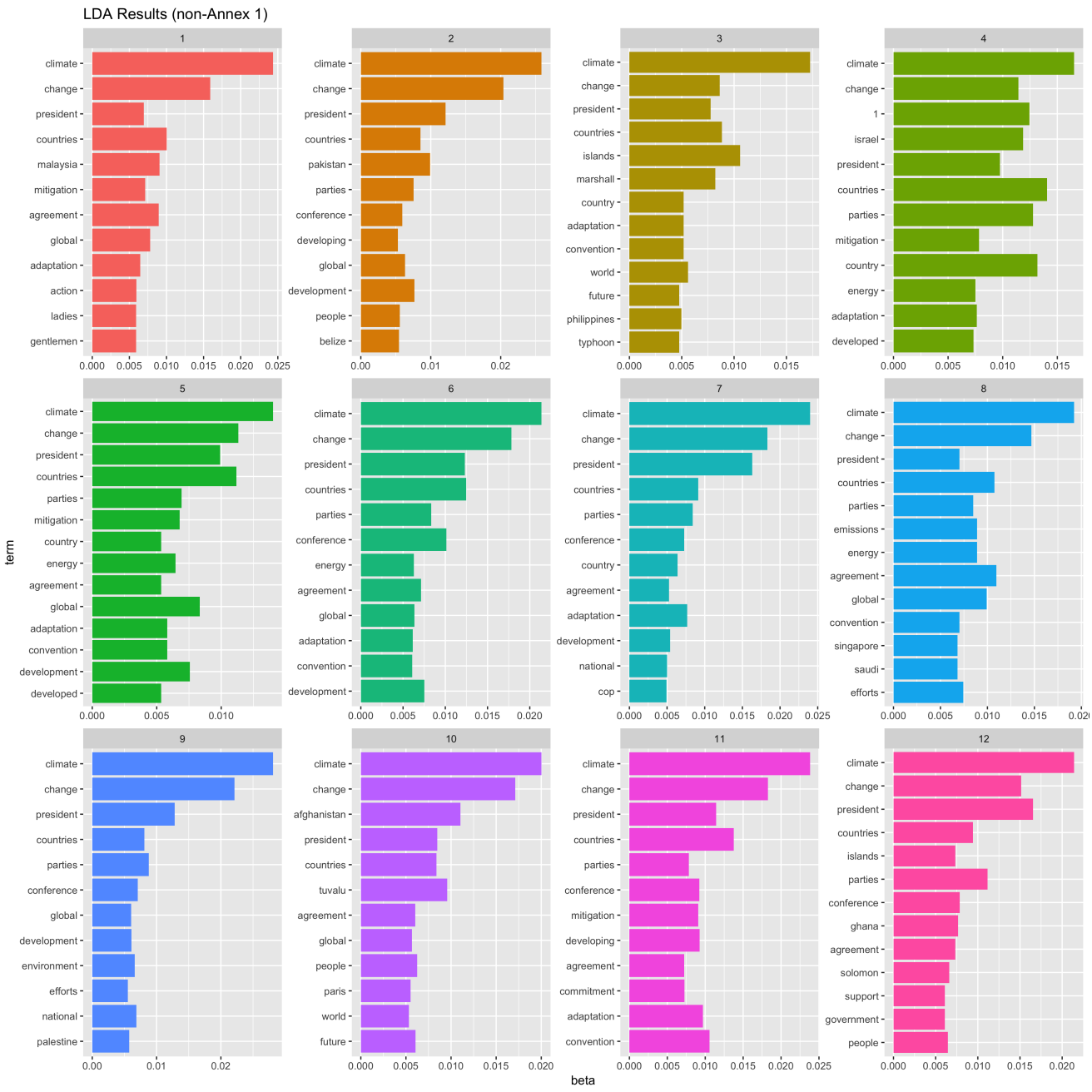

Figure A2: This plot shows the results of a Latent Dirichlet Allocation model with 12 topics for all Non-Annex 1 country statements that appear in five or more meetings. All stopwords are removed.

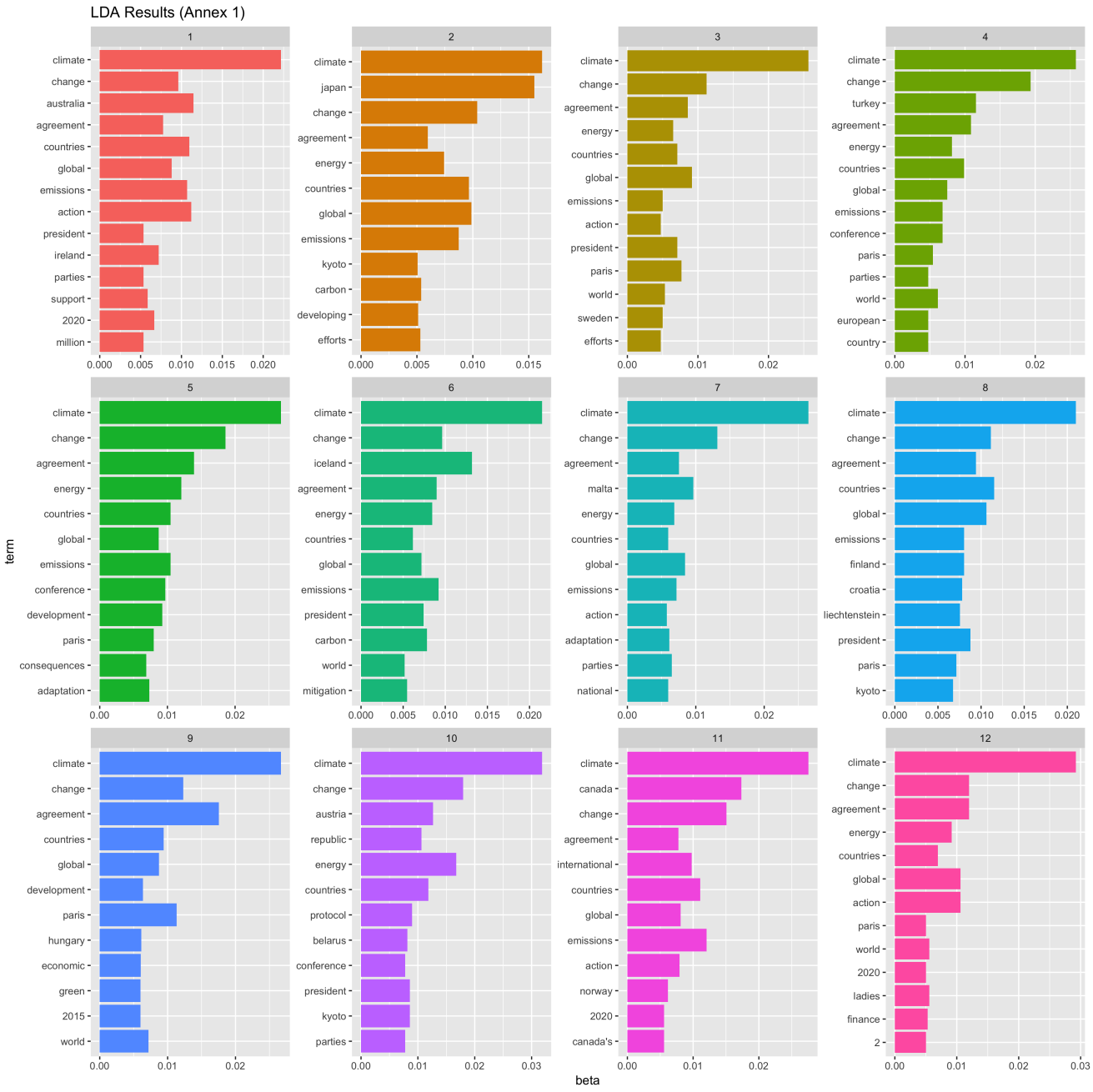

Figure A3: This plot shows the results of a Latent Dirichlet Allocation model with 12 topics for all Annex 1 country statements that appear in five or more meetings. All stopwords are removed.

	Dependent variable:	
	Similarity Scores	
	(1)	(2)
CRI Difference	0.018	−0.027
	(0.027)	(0.019)
Both non-Annex 1	0.066	0.072
	(0.056)	(0.056)
CRI Difference * Both Annex 1		0.098
		(0.076)
UN Ideal Point Difference	−0.057**	−0.054**
	(0.023)	(0.023)
Geographic Distance	−0.242***	−0.243***
	(0.036)	(0.036)
GDP Per Capita Difference	0.007	0.007
	(0.018)	(0.018)
Treaty Indicator	0.098	0.100
	(0.116)	(0.116)
Both Annex 1	0.322***	0.338***
	(0.107)	(0.106)
CRI Difference * Both non-Annex 1	−0.067**	
	(0.034)	
Observations	9,180	9,180
R^2	0.064	0.064
Adjusted R^2	0.064	0.063
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table A1: This table shows the result of OLS models where the dependent variable is the similarity score for each dyad. The CRI difference variable is interacted with variable indicating if both countries non-Annex 1 for model 1, and if both countries are in Annex 1 for model 2. All continuous variables, including the dependent variable, are standardized. Standard errors are clustered at the dyad using the method of Aronow, Samii, and Assenova (2015)

Supporting Information: References

Aronow, Peter M, Cyrus Samii, and Valentina A Assenova. 2015. "Cluster - robust variance estimation for dyadic data." *Political Analysis* 23 (4): 564–577.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *J. Mach. Learn. Res.* 3: 993–1022.
URL: *http://dl.acm.org/citation.cfm?id=944919.944937*

Denny, Matthew J., and Arthur Spirling. 2018. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26 (02): 168–189.
URL: *http://dx.doi.org/10.1017/pan.2017.44*

Le, Quoc, and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML'14 JMLR.org pp. II–1188–II–1196.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality." In *Advances in Neural Information Processing Systems 26*, ed. C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. Curran Associates, Inc. pp. 3111–3119.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.
URL: *http://dx.doi.org/10.1111/ajps.12103*